

Penerapan PCA dan Algoritma Clustering untuk Analisis Mutu Perguruan Tinggi di LLDIKTI Wilayah IV

Resa Rianti*¹, Roni Andarsyah², Rolly Maulana Awangga³

¹Program Studi D4 Teknik Informatika, Universitas Logistik Dan Bisnis Internasional, Bandung, Indonesia
E-mail: *¹Laresarianti23@gmail.com, ²roniandarsyah@ulbi.ac.id, ³awangga@ulbi.ac.id

Abstrak

Sistem Penjaminan Mutu Internal (SPMI) adalah pedoman yang digunakan oleh perguruan tinggi untuk menilai kualitas kinerja dan pelaksanaan pendidikan tinggi secara internal. SPMI sangat penting untuk diperhatikan oleh perguruan tinggi agar dapat bersaing secara positif dengan perguruan tinggi lain, baik di dalam maupun di luar negeri, serta untuk meningkatkan pengelolaan dan pelaksanaan pendidikan tinggi di institusi tersebut. Dalam penelitian ini, tiga algoritma machine learning diterapkan, yaitu K-Means, Mean Shift, dan DBSCAN, untuk melakukan klusterisasi data SPMI. Metode yang digunakan meliputi Analisis Komponen Utama (PCA) untuk mengurangi kompleksitas data tanpa kehilangan informasi penting, dan tiga algoritma clustering untuk mengelompokkan perguruan tinggi berdasarkan kesamaan indikator mutu. Algoritma K-Means mengelompokkan data berdasarkan jarak ke centroid terdekat, Mean Shift mengidentifikasi kluster berdasarkan densitas data, dan DBSCAN mengelompokkan data berdasarkan kepadatan serta mampu menangani outlier dan kluster dengan bentuk tidak beraturan. Hasil penelitian menunjukkan bahwa Mean Shift menghasilkan kluster yang paling baik dengan Silhouette Score 0.566, Davies-Bouldin Index 0.648, dan Calinski-Harabasz Index 971.07. Algoritma K-Means memberikan hasil yang cukup baik dengan Silhouette Score 0.466, Davies-Bouldin Index 0.757, dan Calinski-Harabasz Index 757.06. Sementara itu, DBSCAN memiliki performa yang lebih rendah dengan Silhouette Score 0.216, Davies-Bouldin Index 1.045, dan Calinski-Harabasz Index 105.67. Penelitian ini memberikan hasil identifikasi perguruan tinggi yang membutuhkan perhatian khusus dan membantu dalam perencanaan strategis untuk peningkatan mutu supaya dapat melakukan pembinaan secara lebih efektif serta memberikan kontribusi pada pengembangan sistem penjaminan mutu pendidikan tinggi di Indonesia.

Kata Kunci—SPMI, PCA, K-Means, Mean Shift, DBSCAN

Abstract

The Internal Quality Assurance System (SPMI) is a guideline used by universities to assess the quality of performance and implementation of higher education internally. SPMI is very important to be considered by universities in order to compete positively with other universities, both at home and abroad, as well as to improve the management and implementation of higher education in the institution. In this study, three machine learning algorithms are applied, namely K- Means, Mean Shift, and DBSCAN, to cluster SPMI data. The methods used include Principal Component Analysis (PCA) to reduce data complexity without losing important information, and three clustering algorithms to group universities based on similarity of quality indicators. The K-Means algorithm clusters data based on distance to the nearest centroid, Mean Shift identifies clusters based on data density, and DBSCAN clusters data based on density and is able to handle outliers and irregularly shaped clusters. The results show that Mean Shift produces the best cluster with Silhouette Score 0.566, Davies- Bouldin Index 0.648, and Calinski-Harabasz Index 971.07. The K-Means algorithm provides quite good results with Silhouette Score 0.466, Davies-Bouldin Index 0.757, and Calinski-Harabasz Index 757.06. Meanwhile, DBSCAN has lower performance with Silhouette Score 0.216, Davies-Bouldin Index 1.045, and Calinski-Harabasz Index 105.67. This research provides the results of identifying universities that need special attention and helps in strategic planning for quality improvement so that they can carry out guidance more effectively and contribute to the development of a quality assurance system for higher education in Indonesia.

Keywords—SPMI, PCA, K-Means, Mean Shift, DBSCAN

Diajukan: 3 Juli 2024

Disetujui: 4 Juli 2024

Dipublikasi: 20 Juli 2024

1. PENDAHULUAN

Penjaminan mutu di perguruan tinggi adalah sebuah proses berkelanjutan yang melibatkan perencanaan, pemenuhan,

pengawasan, dan peningkatan standar pendidikan tinggi secara terus-menerus. Proses ini bertujuan untuk memastikan bahwa semua pemangku kepentingan, termasuk mahasiswa, dosen, staf, masyarakat, sektor bisnis, asosiasi profesi, dan pemerintah, merasa puas dengan

kinerja dan hasil yang diberikan oleh perguruan tinggi.[1]. Proses ini menunjukkan tanggung jawab dan keterbukaan dalam pengelolaan perguruan tinggi. Berdasarkan Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi, terdapat dua komponen dalam Sistem Penjaminan Mutu Pendidikan Tinggi (SPMPT), yaitu Sistem Penjaminan Mutu Internal (SPMI) dan Sistem Penjaminan Mutu Eksternal (SPME). SPMI dirancang dan diterapkan oleh perguruan tinggi itu sendiri, sementara SPME dijalankan melalui proses akreditasi. SPMI adalah inisiatif independen setiap perguruan tinggi, yang dibuat, dioperasikan, dan diawasi oleh perguruan tinggi tanpa intervensi dari pemerintah. Pemerintah hanya menyediakan pedoman untuk pelaksanaan SPMI, yang dimaksudkan sebagai sumber inspirasi mengenai berbagai aspek yang biasanya termasuk dalam SPMI, yang penting karena tiap perguruan tinggi memiliki karakteristik unik seperti sejarah, visi, misi, budaya organisasi, ukuran, struktur, sumber daya, dan gaya kepemimpinan yang berbeda[2]. Untuk terus memenuhi kebutuhan stakeholder yang dinamis, perguruan tinggi perlu secara berkelanjutan menyesuaikan Sistem Penjaminan Mutu Internal (SPMI) dengan perubahan tersebut melalui peningkatan yang terus-menerus. LLDIKTI Wilayah IV merupakan institusi yang berfokus pada layanan pendidikan tinggi di wilayahnya, termasuk mendukung implementasi penjaminan mutu di institusi pendidikan tinggi. DIKTI menyediakan aplikasi pemetaan untuk perguruan tinggi yang memungkinkan pengguna untuk melihat grafik performa pendidikan tinggi di seluruh Indonesia. Namun, terdapat kendala dalam pelaksanaan ini karena beberapa perguruan tinggi belum mengisi atau melengkapi dokumen SPMI yang diperlukan, sehingga data yang ditampilkan belum sepenuhnya akurat menggambarkan kinerja faktual perguruan tinggi. Kesulitan umum yang dihadapi perguruan tinggi dalam melakukan penilaian mutu internal seringkali terkait dengan pengorganisasian berkas dan dokumen administrasi SPMI yang tidak teratur. Hal ini sering terjadi apabila pengelola perguruan tinggi kurang memiliki referensi yang memadai mengenai cara pengelolaan dan dokumentasi administrasi SPMI yang efektif.

Penelitian ini bertujuan untuk mengidentifikasi dan mengelompokkan perguruan tinggi di bawah naungan LLDIKTI Wilayah IV berdasarkan indikator-indikator SPMI[3]. Dalam studi ini, tiga algoritma pembelajaran mesin, yaitu K-Means, Mean Shift, dan DBSCAN, digunakan untuk mengklusterisasi data SPMI. Analisis Komponen

Utama (PCA) diaplikasikan untuk menyederhanakan data tanpa menghilangkan informasi krusial. Studi ini menyediakan pedoman yang lebih jelas dan terorganisir untuk LLDIKTI Wilayah IV dalam mendukung dan meningkatkan kualitas perguruan tinggi. Hasil klusterisasi mengidentifikasi perguruan tinggi yang memerlukan intervensi khusus dan membantu dalam perencanaan strategis untuk peningkatan mutu. Oleh karena itu, penelitian ini diharapkan menjadi dasar bagi LLDIKTI Wilayah IV untuk melakukan pembinaan lebih efektif dan berkontribusi pada pengembangan sistem penjaminan mutu pendidikan tinggi di Indonesia.

Dalam konteks peningkatan mutu pendidikan tinggi di Indonesia, Sistem Penjaminan Mutu Internal (SPMI) telah diimplementasikan sebagai bagian dari Sistem Penjaminan Mutu Pendidikan Tinggi (SPMPT), yang diwajibkan oleh Undang-Undang Nomor 12 Tahun 2012 tentang Pendidikan Tinggi. Meskipun pemerintah telah menyediakan pedoman pelaksanaan SPMI, banyak perguruan tinggi di wilayah LLDIKTI IV masih mengalami kesulitan dalam implementasinya. Kendala ini termasuk kegagalan banyak perguruan tinggi untuk mengisi dan melengkapi dokumen SPMI yang diperlukan, sehingga pemetaan mutu pendidikan tinggi tidak sepenuhnya mencerminkan performa perguruan tinggi secara faktual. Masalah dalam pengorganisasian administrasi dan dokumentasi sering terjadi karena kurangnya referensi dan pemahaman tentang standar dan prosedur yang harus diikuti. Selain itu, keragaman karakteristik setiap perguruan tinggi, seperti sejarah, visi dan misi, budaya organisasi, ukuran, struktur, sumber daya, dan gaya kepemimpinan, menyebabkan variasi dalam pelaksanaan SPMI yang kadang tidak sesuai dengan pedoman yang telah ditetapkan. Perguruan tinggi yang gagal menyesuaikan SPMI dengan kebutuhan stakeholders yang terus berkembang sering menghadapi kesulitan dalam mencapai standar mutu yang diharapkan.

Klusterisasi perguruan tinggi dapat memberikan panduan yang lebih jelas dan terstruktur bagi LLDIKTI dalam pembinaan dan peningkatan mutu pendidikan tinggi, serta membantu perguruan tinggi memahami posisi mereka dalam hal mutu pendidikan. Dengan demikian, penelitian ini muncul sebagai respons terhadap kebutuhan untuk mengatasi permasalahan di atas melalui pendekatan yang lebih sistematis dan ilmiah. Penelitian ini bertujuan untuk mengurangi kompleksitas data

SPMI tanpa kehilangan informasi penting, memungkinkan analisis yang lebih mudah dan efisien, serta mengidentifikasi kluster mutu perguruan tinggi berdasarkan indikator-indikator SPMI dengan menggunakan tiga algoritma machine learning (K-Means, Mean Shift, dan DBSCAN) dan Analisis Komponen Utama (PCA)[4].

2. METODE PENELITIAN

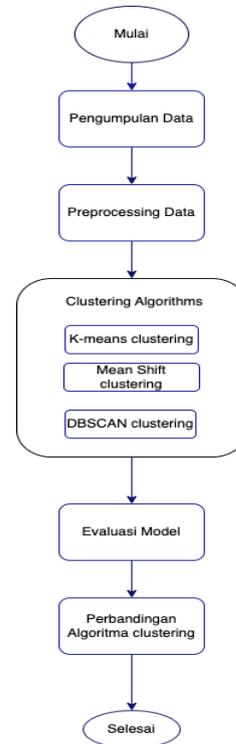
Metodologi penelitian adalah kerangka kerja sistematis yang digunakan untuk mengumpulkan, menganalisis, dan menginterpretasikan data dalam rangka menjawab pertanyaan penelitian atau menguji hipotesis. Metodologi ini mencakup berbagai langkah dan prosedur yang dirancang untuk memastikan bahwa penelitian dilakukan dengan cara yang terstruktur, valid, dan dapat diandalkan[5]. Ini melibatkan pemilihan metode penelitian yang tepat, teknik pengumpulan data, analisis statistik, dan interpretasi hasil yang diperoleh.

2.1. Kerangka penelitian

Bagan metodologi penelitian ini menjelaskan langkah-langkah yang diambil untuk menganalisis dan mengelompokkan data kinerja perguruan tinggi menggunakan tiga algoritma clustering: K-Means, Mean Shift, dan DBSCAN. Proses dimulai dengan Pengumpulan Data dari sumber yang relevan, diikuti oleh tahap Preprocessing Data yang meliputi pembersihan data dari nilai kosong (NaN), standarisasi data menggunakan StandardScaler, dan pengurangan dimensi menggunakan PCA (Principal Component Analysis) untuk mengurangi kompleksitas data. Setelah data siap, tiga algoritma clustering diterapkan untuk mengelompokkan data perguruan tinggi berdasarkan kinerja mereka. K-Means mengelompokkan data ke dalam sejumlah cluster yang telah ditentukan, Mean Shift mengidentifikasi area dengan kepadatan tinggi, dan DBSCAN mengelompokkan data berdasarkan kepadatan, memungkinkan identifikasi outliers atau noise[6].

Setelah penerapan algoritma clustering, tahap Evaluasi Model dilakukan untuk menilai kualitas hasil clustering menggunakan metrik seperti Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index. Evaluasi ini memberikan gambaran tentang seberapa baik setiap algoritma mengelompokkan data. Hasil evaluasi dari ketiga algoritma kemudian dibandingkan dalam tahap perbandingan algoritma clustering menggunakan tabel dan

visualisasi grafik batang untuk menentukan algoritma yang paling efektif[7]. Penelitian diakhiri dengan kesimpulan, di mana hasil evaluasi dan perbandingan digunakan untuk memberikan rekomendasi tentang langkah-langkah yang dapat diambil untuk meningkatkan kinerja dan pengelolaan perguruan tinggi berdasarkan pola dan cluster yang ditemukan.



Gambar 1. Kerangka Penelitian

2.2 Pengumpulan Data

Tahap awal dari penelitian ini melibatkan proses pengumpulan data dengan mengakses laman web <https://l1dikti4.kemdikbud.go.id/tag/spmi/>. Pada tahap ini, semua data dan informasi yang relevan dan diperlukan untuk penelitian ini dikumpulkan secara sistematis. Proses pengumpulan data dilakukan dengan cermat untuk memastikan kelengkapan data yang akan digunakan dalam analisis. Seluruh data yang terkumpul kemudian diimpor dan diorganisir dalam format Microsoft Excel. Setelah data terorganisir di Excel, data tersebut kemudian diekspor ke dalam format CSV (Comma-Separated Values) untuk memudahkan analisis menggunakan teknik machine learning.



Gambar 2. Website SPMI LLDIKTI IV

2.3 Data Pre-processing

Pre-processing data adalah langkah awal yang sangat penting dalam penelitian ini untuk memastikan bahwa data yang akan dianalisis bersih, terstandarisasi, dan siap digunakan. Langkah pertama dalam pre-processing adalah Data Cleaning. Pada tahap ini, kita menghapus semua baris data yang mengandung nilai kosong (NaN). Ini dilakukan untuk memastikan bahwa data yang akan dianalisis lengkap dan tidak ada missing values yang dapat mengganggu hasil analisis. Data yang bersih ini kemudian digunakan sebagai dasar untuk langkah-langkah berikutnya dalam proses pre-processing[8].

Langkah selanjutnya adalah Standarisasi Data. Standarisasi dilakukan untuk memastikan bahwa semua fitur dalam dataset memiliki skala yang sama, karena fitur dengan skala yang berbeda dapat mempengaruhi hasil analisis clustering. Standarisasi dilakukan menggunakan teknik StandardScaler dari pustaka sklearn, yang mengubah setiap fitur sehingga memiliki mean 0 dan standar deviasi 1. Setelah data distandarisasi, kita melakukan Pengurangan Dimensi (Dimensionality Reduction) menggunakan teknik Principal Component Analysis (PCA). PCA digunakan untuk mengurangi jumlah fitur menjadi tiga komponen utama, yang mencakup sebagian besar varians dalam data. Langkah ini membantu dalam mengurangi kompleksitas data dan meningkatkan interpretasi analisis tanpa menghilangkan informasi penting. Dengan data yang telah bersih, terstandarisasi, dan direduksi dimensinya, kita dapat melanjutkan ke tahap analisis clustering dengan lebih efektif.

2.4 Algoritma Clustering

2.4.1 Model K-Means

K-Means adalah salah satu algoritma clustering yang paling sederhana dan paling umum digunakan. Algoritma ini bekerja dengan membagi data ke dalam k cluster yang telah ditentukan sebelumnya. Tujuan dari K-Means

adalah meminimalkan varians dalam setiap cluster[9]. Proses ini dilakukan dengan mengelompokkan data ke dalam cluster yang memiliki centroid terdekat. Langkah-langkah dalam menggunakan algoritma K-Means:

- inisialisasi Centroid.
Menentukan kkk centroid secara acak dari dataset.
- Pengelompokan Titik Data
Setiap titik data dihubungkan ke centroid terdekat berdasarkan jarak Euclidean.
- Menghitung Ulang Centroid
Setelah semua titik data dikelompokkan, centroid baru dihitung sebagai rata-rata dari semua titik data dalam cluster.
- Iterasi.
:Langkah 2 dan 3 diulang hingga centroid tidak berubah lagi (konvergensi) atau jumlah iterasi maksimum tercapai[10].

2.4.2 Model Mean Shift

Mean Shift adalah algoritma clustering yang tidak memerlukan penentuan jumlah cluster di awal. Algoritma ini bekerja dengan mengidentifikasi area dengan kepadatan tinggi dalam data dan mengelompokkan data berdasarkan centroid yang ditemukan[11]. Langkah-langkah dalam menggunakan algoritma Mean Shift:

- Menentukan Window (Bandwidth)
Menentukan radius atau bandwidth dari window yang digunakan untuk menghitung centroid.
- Menghitung Mean Shift
Mengeser window ke arah rata-rata data poin yang berada di dalam window tersebut.
- Iterasi
Mengulangi langkah 2 hingga posisi centroid tidak berubah lagi (konvergensi)[12].

2.4.3 Metode DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah algoritma clustering yang menggunakan kepadatan dari data untuk membentuk cluster. DBSCAN memungkinkan identifikasi cluster dengan bentuk dan ukuran yang beragam, serta dapat menangani keberadaan noise dalam data[12]. Algoritma DBSCAN menggunakan

dua parameter utama: Epsilon (ϵ) dan Minimum Number of Points (MinPts). Konsep dasar DBSCAN adalah ketercapaian kepadatan dan keterhubungan kepadatan, yang bergantung pada parameter radius maksimal (epsilon) dan jumlah minimum objek dalam sebuah cluster (MinPts).

- a. Epsilon (ϵ)
Jarak yang menentukan kedekatan antara objek dengan objek di sekitarnya. Objek yang memenuhi nilai ini dianggap memiliki kedekatan dengan objek yang diamati.
- b. Minimum Number of Points (MinPts).
Jumlah minimum objek yang diperlukan untuk membentuk sebuah cluster jika saling berdekatan.
- c. Objek Border
Objek yang berdekatan dengan objek core, sedangkan objek core adalah objek yang berdekatan dengan nilai epsilon yang diberikan. Objek yang tidak termasuk core atau border disebut sebagai anomali atau pencilan.

2.5. Evaluasi Model

Evaluasi model clustering merupakan tahap penting untuk menilai seberapa baik algoritma clustering yang telah diterapkan dalam mengelompokkan data. Tiga metrik yang umum digunakan untuk mengevaluasi hasil clustering adalah Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index[13].

1. Silhouette Score
Mengukur seberapa mirip data dengan kluster mereka sendiri dibandingkan dengan kluster lain. Nilai Silhouette berkisar antara -1 dan 1, di mana nilai yang lebih tinggi menunjukkan bahwa data lebih baik dikelompokkan[13].
2. Davies-Bouldin Index
Mengevaluasi rata-rata rasio jarak antara kluster dengan jarak di dalam kluster. Nilai yang lebih rendah menunjukkan kluster yang lebih kompak dan terpisah dengan baik.
3. Calinski-Harabasz Indeks
Menilai rasio antara jumlah total dispersi antar kluster dengan jumlah total dispersi dalam kluster. Nilai yang lebih tinggi menunjukkan clustering yang lebih baik.

3. HASIL DAN PEMBAHASAN

3.1 Pre-processing Data

Pre-processing data adalah langkah penting dalam analisis data yang bertujuan untuk mempersiapkan dan mengubah data mentah menjadi bentuk yang sesuai untuk analisis lebih lanjut. Langkah ini memastikan bahwa data yang digunakan dalam proses clustering atau model lainnya memiliki kualitas yang tinggi, bersih, dan terstandarisasi[16]. Untuk data SPMI (Sistem Penjamin Mutu Internal) yang telah dianalisis sebelumnya, langkah-langkah pre-processing yang dilakukan adalah sebagai berikut:

3.1.1 Data Cleaning

Data cleaning adalah langkah pertama dalam pre-processing yang bertujuan untuk memastikan bahwa dataset bebas dari nilai-nilai yang hilang atau tidak lengkap. Pada penelitian ini, langkah data cleaning dilakukan dengan menghapus baris yang memiliki nilai NaN (Not a Number)[14]. Output dari langkah ini adalah dataset yang bersih tanpa nilai yang hilang. Ini memastikan bahwa semua baris data memiliki nilai yang lengkap, sehingga analisis selanjutnya tidak terganggu oleh missing values.

3.1.2 Standarisasi Data

Standarisasi dilakukan untuk memastikan bahwa semua fitur dalam dataset memiliki skala yang sama. Ini penting karena fitur dengan skala yang berbeda dapat mempengaruhi hasil analisis clustering[15]. Teknik StandardScaler dari pustaka sklearn digunakan untuk standarisasi data, yang mengubah setiap fitur sehingga memiliki mean 0 dan standar deviasi 1. Output dari langkah ini adalah dataset yang semua fiturnya telah distandarisasi, yang memungkinkan analisis clustering dilakukan dengan lebih akurat dan konsisten.

3.1.3 Pengurangan Dimensi (Dimensionality Reduction)

PCA digunakan untuk mengurangi jumlah fitur dalam dataset menjadi tiga komponen utama, yang mencakup sebagian besar varians dalam data. Ini membantu dalam mengurangi kompleksitas data dan mempermudah interpretasi hasil analisis clustering.

- a. PCA (Principal Component Analysis)

Principal Component Analysis (PCA) adalah salah satu metode yang paling dikenal dalam pengurangan dimensi data. Sebagai teknik statistik konvensional, PCA digunakan untuk mengurangi jumlah dimensi dalam sebuah dataset. Melalui transformasi linear, PCA efektif dalam mengurangi dimensi data, mengekstrak informasi penting dari kumpulan data yang besar, dan menganalisis struktur variabel secara menyeluruh[16]. Teknik PCA digunakan untuk mengurangi kompleksitas data dengan mengurangi jumlah dimensi menjadi tiga komponen utama. Langkah ini membantu dalam mengurangi redundansi dan kebisingan dalam data.

algorithm clustering. Dalam contoh ini, terdapat tiga cluster yang diberi label 0, 1, dan 2.

Dengan demikian, tabel ini memberikan gambaran tentang bagaimana data asli telah diproses melalui PCA dan kemudian dikelompokkan menggunakan algoritma clustering, menunjukkan efektivitas dari pengurangan dimensi dalam mempermudah analisis dan interpretasi data.

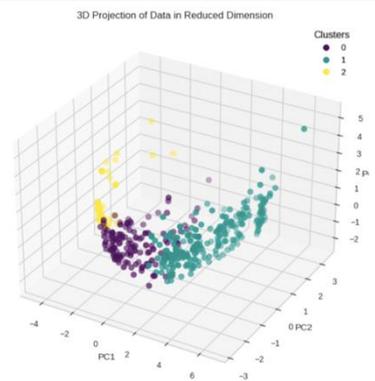
Table 1. Dimensionality Reduction with PCA

PC1	PC2	PC3	Cluster
0.495089	- 1.180862	- 1.120072	0
- 2.697291	- 0.739556	- 1.137955	0
- 2.575386	- 0.718411	- 0.378424	0
- 4.607298	0.795655	1.344903	2
1.974623	0.428249	- 1.439675	1

Tabel di atas menunjukkan hasil dari pengurangan dimensi menggunakan Principal Component Analysis (PCA) yang diikuti oleh proses clustering. Tabel ini mencakup nilai dari tiga komponen utama (PC1, PC2, dan PC3) yang dihasilkan oleh PCA, serta cluster yang dihasilkan dari algoritma clustering (dalam hal ini K-Means).

PC1, PC2, dan PC3: Ini adalah tiga komponen utama yang dihasilkan dari PCA. Setiap nilai di kolom ini mewakili koordinat dari data asli dalam ruang tiga dimensi yang baru. Komponen ini mencakup sebagian besar varians dalam data, yang berarti mereka mempertahankan sebagian besar informasi penting dari data asli meskipun jumlah dimensinya telah dikurangi.

Cluster: Kolom ini menunjukkan hasil clustering dari data yang telah direduksi dimensinya. Angka di kolom ini menunjukkan ke cluster mana data tersebut dialokasikan oleh



Gambar 3. Visualisasi Proses Reduksi Dimensi

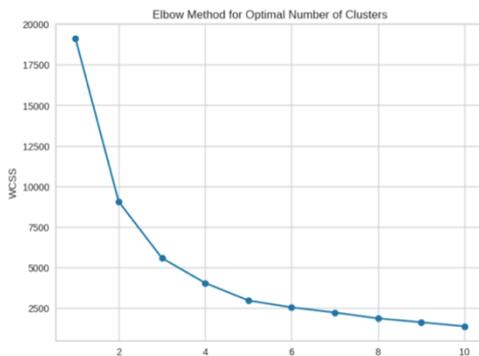
Gambar 3 di atas merupakan visualisasi tiga dimensi dari data yang telah direduksi dimensinya menggunakan Principal Component Analysis (PCA) dan kemudian dikelompokkan menggunakan algoritma KMeans. PCA digunakan untuk mengurangi data ke dalam tiga komponen utama, yaitu PC1, PC2, dan PC3, yang mencerminkan variansi terbesar dalam data. Visualisasi ini memperlihatkan bagaimana data diproyeksikan ke dalam ruang tiga dimensi yang dibentuk oleh komponen-komponen utama tersebut. Data dengan karakteristik serupa dikelompokkan dalam kluster yang berbeda: Kluster 0 (ungu), Kluster 1 (kuning), dan Kluster 2 (hijau). Kluster 0 terlihat terpisah dengan jelas dari dua kluster lainnya, menunjukkan adanya kelompok data

dengan karakteristik yang unik. Kluster 1 terdistribusi di area yang lebih tinggi pada sepanjang PC2 dan PC3, sedangkan Kluster 2 terletak di sekitar pusat distribusi data, dengan beberapa penyebaran di sepanjang PC1 dan PC2. Hasil ini menunjukkan bahwa PCA berhasil mereduksi dimensi data dan memisahkan data ke dalam kluster yang berbeda berdasarkan karakteristik utamanya, sehingga memfasilitasi pemahaman yang lebih mendalam mengenai struktur dan hubungan antar data.

3.2 Evaluasi Model

Evaluasi model clustering adalah langkah untuk menilai seberapa baik algoritma clustering bekerja dalam mengelompokkan data. Dalam penelitian ini, tiga metrik evaluasi utama digunakan: Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index. Silhouette Score mengukur seberapa mirip objek dalam cluster mereka sendiri dibandingkan dengan objek di cluster lain, dengan nilai berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan cluster yang lebih baik. Davies-Bouldin Index mengukur rata-rata jarak antara setiap cluster dengan cluster yang paling mirip dengannya, dengan nilai yang lebih rendah menunjukkan cluster yang lebih baik. Calinski-Harabasz Index mengukur rasio antara jumlah total dispersi antar cluster dengan jumlah total dispersi dalam cluster, di mana nilai yang lebih tinggi menunjukkan cluster yang lebih baik. Evaluasi ini memberikan wawasan mendalam tentang kualitas dan validitas cluster yang terbentuk dari setiap algoritma clustering yang digunakan.

Berikut adalah tabel yang menunjukkan hasil evaluasi dari tiga algoritma clustering berdasarkan Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index:



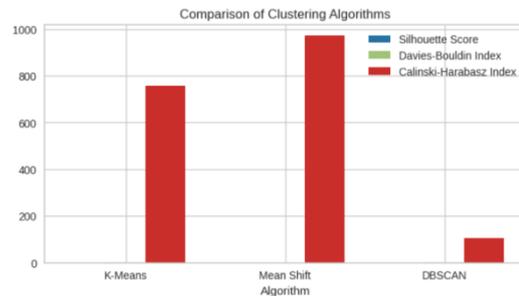
Gambar 5. Metode Elbow

Tabel 2. Hasil Evaluasi algoritma

Dari tabel ini, kita dapat melihat bahwa algoritma Mean Shift memiliki nilai Silhouette Score dan Calinski-Harabasz Index tertinggi, serta Davies-Bouldin Index terendah, menunjukkan bahwa Mean Shift memberikan hasil clustering yang paling baik dan efektif dibandingkan dengan K-Means dan DBSCAN untuk dataset ini.

3.3 Perbandingan Performa Kerja Model

Hasil evaluasi tersebut dikompilasi ke dalam bentuk tabel untuk memberikan gambaran yang jelas tentang performa masing-masing algoritma berdasarkan Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index. Selanjutnya, hasil ini divisualisasikan dalam bentuk grafik batang untuk mempermudah interpretasi dan perbandingan. Perbandingan ini bertujuan untuk mengidentifikasi algoritma mana yang paling efektif dalam mengelompokkan data kinerja perguruan tinggi[17][18]. Dari hasil evaluasi dan visualisasi ini, dapat disimpulkan algoritma terbaik yang memberikan hasil clustering yang paling valid dan bermakna, serta memberikan rekomendasi untuk langkah-langkah perbaikan atau peningkatan kinerja perguruan tinggi.



Gambar 4. Hasil Perbandingan Performa Algoritma Clustering

Berdasarkan ketiga metrik evaluasi tersebut, **Mean Shift** tampaknya menghasilkan cluster yang paling baik untuk data ini. K-Means juga memberikan hasil yang baik, sementara DBSCAN memiliki performa yang lebih rendah dibandingkan dua algoritma lainnya.

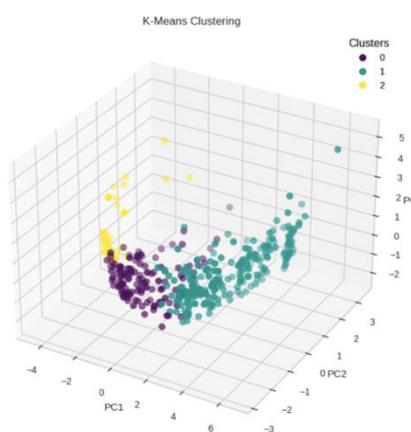
3.3.1 Performa Kerja Model K-Means

Algorith m	Silhouet te Score	Davies-Bouldin Index	Calinski-Harabasz Index
K-Means	0.4661 97	0.7572 39	757.0647 13
Mean Shift	0.56605 7	0.64808 3	971.07935 1
DBSCA N	0.21645 3	1.04495 5	105.66973 3

Metode Elbow adalah teknik yang digunakan untuk menentukan jumlah optimal cluster dalam algoritma K-Means. Gambar 5 diatas menunjukkan plot dari Within-Cluster Sum of Squares (WCSS) terhadap jumlah cluster, dan kita dapat mengamati bahwa nilai WCSS menurun tajam saat jumlah cluster meningkat dari 1 hingga sekitar 3 atau 4. Penurunan WCSS ini menunjukkan bahwa penambahan cluster hingga 3 atau 4 memberikan pengelompokan yang lebih baik. Setelah jumlah cluster mencapai 3 atau 4, penurunan WCSS mulai melambat dan grafik membentuk "siku" atau "elbow[13]". Titik ini menunjukkan jumlah cluster yang optimal, karena penambahan cluster lebih lanjut hanya memberikan pengurangan WCSS yang tidak signifikan. Oleh karena itu, memilih 3 atau 4 cluster memberikan keseimbangan yang baik antara kompleksitas model dan kualitas clustering.

Berdasarkan penelitian dan metode elbow yang telah diterapkan sebelumnya, jumlah optimal cluster untuk data ini adalah 3 atau 4. K-Means memberikan hasil clustering yang baik dengan membentuk cluster yang jelas dan terpisah.

Visualisasi dibawah ini menunjukkan hasil clustering menggunakan algoritma K-Means dengan tiga cluster optimal berdasarkan komponen utama (PC1, PC2, dan PC3) dari data kinerja perguruan tinggi. Setiap titik dalam plot 3D mewakili satu perguruan tinggi, dan warna titik menunjukkan cluster yang dihasilkan oleh K-Means. Cluster 0 ditandai dengan warna ungu, cluster 1 dengan warna hijau, dan cluster 2 dengan warna kuning.



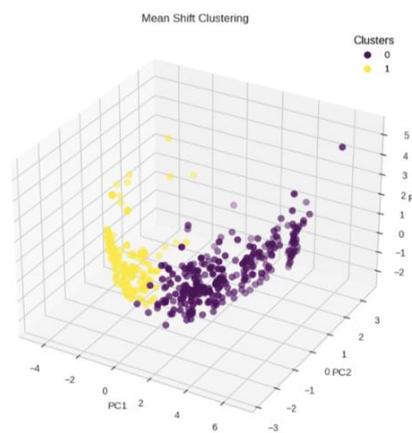
Gambar 6. Visualisasi K-Means

Plot ini menunjukkan bahwa data telah berhasil dikelompokkan ke dalam tiga cluster yang terpisah, dengan setiap cluster mengelompokkan perguruan tinggi yang memiliki karakteristik kinerja yang serupa. Cluster 1 (hijau) terlihat lebih tersebar, sementara cluster 0 (ungu) dan cluster 2 (kuning) lebih terpusat di area tertentu. Visualisasi ini membantu dalam memahami bagaimana perguruan tinggi dikelompokkan berdasarkan kinerja mereka dan memberikan wawasan tentang pola yang ada dalam data.

3.3.2 Performa Kerja Model Mean-Shift

Mean Shift adalah algoritma clustering yang tidak memerlukan penentuan jumlah cluster di awal. Algoritma ini mengidentifikasi area dengan kepadatan tinggi dan mengelompokkan data berdasarkan centroid yang ditemukan[19]. Hasil clustering dengan Mean Shift cenderung membentuk cluster yang lebih fleksibel dan adaptif terhadap distribusi data

Visualisasi 3D dari hasil clustering dengan Mean Shift menunjukkan bahwa algoritma ini mampu mengidentifikasi cluster yang mungkin tidak berbentuk bulat atau elips seperti yang dihasilkan oleh K-Means. Evaluasi performa Mean Shift menunjukkan nilai Silhouette Score yang baik, yang menunjukkan bahwa data dalam setiap cluster memiliki kemiripan yang tinggi. Davies-Bouldin Index juga menunjukkan hasil yang baik, menunjukkan bahwa cluster yang terbentuk cukup terpisah dan tidak terlalu mirip satu sama lain.

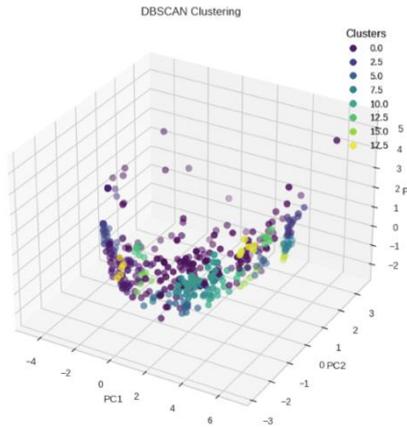


Gambar 7. Visualisasi Model Mean-Shift

3.3.3 Performa Kerja Model DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah algoritma

clustering yang bekerja dengan mengidentifikasi cluster berdasarkan kepadatan data. Algoritma ini efektif dalam mengidentifikasi cluster dengan bentuk yang tidak teratur dan dapat menangani outliers atau noise dengan baik.



Gambar 8. Visualisasi Model DBSCAN

Visualisasi 3D dari hasil clustering dengan DBSCAN menunjukkan kemampuan algoritma ini dalam mengelompokkan data yang tidak teratur dan mendeteksi outliers. DBSCAN memberikan hasil clustering yang berbeda dengan K-Means dan Mean Shift, dengan beberapa titik data yang tidak masuk dalam cluster mana pun (disebut noise). Evaluasi performa DBSCAN menunjukkan bahwa algoritma ini unggul dalam situasi di mana data memiliki noise atau outliers, dengan nilai Davies-Bouldin Index yang rendah dan Silhouette Score yang cukup baik. Namun, performa DBSCAN sangat bergantung pada parameter epsilon dan min_samples yang dipilih.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis dan mengelompokkan data kinerja perguruan tinggi menggunakan tiga algoritma clustering yang berbeda: K-Means, Mean Shift, dan DBSCAN. Melalui proses pre-processing yang meliputi data cleaning, standarisasi data, dan pengurangan dimensi menggunakan Principal Component Analysis (PCA), data kinerja perguruan tinggi diolah untuk memastikan kualitas dan keseragaman sebelum diterapkan algoritma clustering. Hasil penelitian menunjukkan bahwa Mean Shift memberikan hasil clustering yang paling baik dibandingkan dengan K-Means dan DBSCAN, ditinjau dari tiga metrik evaluasi utama: Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz

Index. Algoritma Mean Shift menunjukkan performa terbaik dengan Silhouette Score sebesar 0.566, Davies-Bouldin Index sebesar 0.648, dan Calinski-Harabasz Index sebesar 971.07. Sementara itu, K-Means menunjukkan hasil yang cukup baik dengan Silhouette Score sebesar 0.466, Davies-Bouldin Index sebesar 0.757, dan Calinski-Harabasz Index sebesar 757.06. DBSCAN memiliki performa terendah dengan Silhouette Score sebesar 0.216, Davies-Bouldin Index sebesar 1.045, dan Calinski-Harabasz Index sebesar 105.67. Dari hasil ini, dapat disimpulkan bahwa Mean Shift lebih efektif dalam mengelompokkan data perguruan tinggi berdasarkan indikator mutu, memberikan cluster yang lebih baik terpisah dan kompak. Oleh karena itu, algoritma Mean Shift direkomendasikan untuk digunakan dalam analisis lebih lanjut dan pengambilan keputusan terkait peningkatan kinerja dan pengelolaan perguruan tinggi.

5. SARAN

Untuk penelitian lebih lanjut, disarankan agar peneliti mengatasi beberapa kekurangan yang ada dalam penelitian ini. Pertama, data yang digunakan dapat ditingkatkan dengan mencakup lebih banyak variabel mutu dan sumber data yang lebih beragam untuk memastikan hasil yang lebih komprehensif. Kedua, metode preprocessing dapat diperbaiki dengan menggunakan teknik imputasi data yang lebih canggih untuk menangani nilai-nilai yang hilang, daripada hanya menghapus baris dengan nilai NaN. Ketiga, penentuan parameter optimal untuk algoritma clustering seperti epsilon dan minPts pada DBSCAN perlu diteliti lebih lanjut untuk meningkatkan kinerja algoritma tersebut. Terakhir, penelitian ini dapat memperluas evaluasi dengan menggunakan lebih banyak metrik evaluasi dan validasi silang untuk memastikan bahwa hasil klusterisasi konsisten dan dapat digeneralisasi.

REFERENSI

- [1] R. Sanjaya and R. N. Handayani, "PENGEMBANGAN SISTEM INFORMASI PENJAMINAN MUTU (SIMANTU) LLDIKTI WILAYAH IV," vol. 03, 2021.
- [2] M. Fadhli, "SISTEM PENJAMINAN MUTU INTERNAL DAN EKSTERNAL PADA LEMBAGA PENDIDIKAN TINGGI," *Jurnal Manajemen Pendidikan Islam*, vol. 04,

- no. 02, pp. 171–183, 2020, doi: 10.33650/al-tanzim.v4i2.
- [3] V. Sriviana Fatmawaty and I. Riadi, “JURNAL MEDIA INFORMATIKA BUDIDARMA Klasterisasi Perguruan Tinggi LLDIKTI V Berdasarkan Indikator Kinerja Utama dan PDDIKTI Menggunakan K-Means Clustering,” 2024, doi: 10.30865/mib.v8i2.7497.
- [4] S. Sirsikar and K. Wankhede, “Comparison of clustering algorithms to design new clustering approach,” in *Procedia Computer Science*, Elsevier B.V., 2015, pp. 147–154. doi: 10.1016/j.procs.2015.04.238.
- [5] S. Anastassia Amellia Kharis and A. Haqqi Anna Zili, “Learning Analytics dan Educational Data Mining pada Data Pendidikan,” *Jurnal Riset Pembelajaran Matematika Sekolah*, vol. 6, 2022.
- [6] S. Anastassia Amellia Kharis and A. Haqqi Anna Zili, “Learning Analytics dan Educational Data Mining pada Data Pendidikan,” *Jurnal Riset Pembelajaran Matematika Sekolah*, vol. 6, 2022.
- [7] B. Untung Saputra, W. Andriani, and A. Batik Pesisir, “PENGENALAN MOTIF BATIK PESISIR PULAU JAWA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK.” [Online]. Available: <https://journal.fkom.uniku.ac.id/ilkom>
- [8] L. Zhang, A. U. Quddus, E. Katranaras, D. Wübben, Y. Qi, and R. Tafazolli, “Performance Analysis and Optimal Cooperative Cluster Size for Randomly Distributed Small Cells Under Cloud RAN,” *IEEE Access*, vol. 4, pp. 1925–1939, 2016, doi: 10.1109/ACCESS.2016.2550758.
- [9] K. P. Sinaga and M. S. Yang, “Unsupervised K-means clustering algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [10] A. Suryadibrata and J. C. Young, “Visualisasi Algoritma sebagai Sarana Pembelajaran K-Means Clustering,” *ULTIMATICS*, vol. XII, no. 1, 2020.
- [11] C. Cariou, S. Le Moan, and K. Chehdi, “A Novel Mean-Shift Algorithm for Data Clustering,” *IEEE Access*, vol. 10, pp. 14575–14585, 2022, doi: 10.1109/ACCESS.2022.3147951.
- [12] R. G. Crețulescu, D. I. Morariu, M. Breazu, and D. Volovici, “DBSCAN Algorithm for Document Clustering,” *International Journal of Advanced Statistics and IT&C for Economics and Life Sciences*, vol. 9, no. 1, pp. 58–66, Jun. 2019, doi: 10.2478/ijasitels-2019-0007.
- [13] I. Firman Ashari, E. Dwi Nugroho, R. Baraku, I. N. Yanda, and R. Liwardana, “Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta,” 2023. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [14] A. R. Sinaga, “Variabel Non Akademik Untuk Memprediksi Prestasi Siswa Dengan Data Mining Menggunakan Metoda Naïve Bayes.” [Online]. Available: <https://journal.fkom.uniku.ac.id/ilkom>
- [15] C. Yuan and H. Yang, “Research on K-Value Selection Method of K-Means Clustering Algorithm,” *J (Basel)*, vol. 2, no. 2, pp. 226–235, Jun. 2019, doi: 10.3390/j2020016.
- [16] A. S. Ritonga and I. Muhandhis, “TEKNIK DATA MINING UNTUK MENGLASIFIKASIKAN DATA ULASAN DESTINASI WISATA MENGGUNAKAN REDUKSI DATA PRINCIPAL COMPONENT ANALYSIS (PCA).”
- [17] T. A. Cinderatama *et al.*, “Implementasi Metode K-Means, Dbscan, dan Meanshift Untuk Analisis Jenis Ancaman Jaringan Pada Intrusion Detection System,” vol. 7, no. 1, p. 2022.
- [18] J. Homepage *et al.*, “MALCOM: Indonesian Journal of Machine Learning and Computer Science Clustering Electricity Distribution Data Using Density-Based Spatial Clustering of Applications With Noise (DBSCAN) Algorithm Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Density Based Spatial Clustering of Application With Noise (DBSCAN),” vol. 4, no. 3, pp. 1024–

1033, 2024, doi:
10.57152/malcom.v4i2.1426.

- [19] M. Á. Carreira-Perpiñán, “A review of mean-shift algorithms for clustering,” Mar. 2015, [Online]. Available: <http://arxiv.org/abs/1503.00687>