

OPTIMASI ALGORITMA CONSTRAINED K-MEANS MENGUNAKAN PRINCIPAL COMPONENT ANALYSIS PADA CLUSTERING PESERTA DIDIK REKAYASA PERANGKAT LUNAK SMKN 2 KRAKSAAN

Matlubul Khairi, Mohammad Alvin Kafabihi, Ivan Lailur Romadhan

Program Studi Teknik Informatika, Universitas Nurul Jadid

Jl. PP Nurul Jadid, Kec. Paiton, Kabupaten Probolinggo,

sangrato88@gmail.com, malvinkafabihi15022004@gmail.com, ivanlailurr@gmail.com

Abstract - This study aims to optimize the Constrained K-Means algorithm with dataset dimensionality reduction using Principal Component Analysis (PCA) for clustering students in the Software Engineering program at SMKN 2 Kraksaan. The main challenges addressed are the effectiveness and efficiency of the clustering process on high-dimensional datasets and specific constraints, such as cluster size limitations or data relationships. The conventional K-Means algorithm is not designed to handle these constraints, necessitating a more effective approach. This research adopts a quantitative method with an experimental approach. The data used consist of academic ability test results for Grade X students, including attributes such as sequence, mathematics, figural, memory, and analogy scores. The research steps include data selection, data cleaning, dimensionality reduction using PCA, and clustering with the Constrained K-Means algorithm. Clustering results are evaluated using the Silhouette Score. The findings demonstrate that the combination of PCA and Constrained K-Means Clustering effectively groups students into three balanced clusters. Dimensionality reduction using PCA improves the clustering algorithm's performance in terms of both speed and accuracy. The clustering results reveal three main groups: students with high, medium, and low abilities. Evaluation with the Silhouette Score indicates the highest value at the reduction to a single PCA component, producing the most optimal clustering.

Keywords - Clustering, Constrained K-Means, Dimensionality Reduction, Principal Component Analysis, Silhouette Score.

Abstrak - Penelitian ini bertujuan untuk mengoptimalkan algoritma Constrained K-Means dengan reduksi dimensi dataset menggunakan Principal Component Analysis (PCA) pada pengelompokan siswa Rekayasa Perangkat Lunak di SMKN 2 Kraksaan. Masalah utama yang dihadapi adalah efektivitas dan efisiensi proses klusterisasi pada dataset berdimensi tinggi serta adanya keterbatasan tertentu yang harus dipatuhi, seperti batasan pada ukuran kluster atau hubungan antar data. Algoritma K-Means konvensional tidak dirancang untuk menangani kendala-kendala ini sehingga diperlukan pendekatan yang lebih efektif. Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimental. Data yang digunakan adalah hasil tes uji kemampuan akademik siswa kelas X yang mencakup atribut nilai deret, matematika, figural, mengingat, dan analogi. Tahapan penelitian meliputi seleksi data, pembersihan data, reduksi dimensi dengan PCA, dan klusterisasi menggunakan algoritma Constrained K-Means. Evaluasi hasil klusterisasi dilakukan menggunakan Silhouette Score. Hasil penelitian menunjukkan bahwa kombinasi PCA dan Constrained K-Means Clustering efektif dalam mengelompokkan siswa ke dalam tiga kluster yang seimbang. Reduksi dimensi menggunakan PCA meningkatkan kinerja algoritma klusterisasi baik dari segi kecepatan maupun akurasi. Hasil klusterisasi menunjukkan adanya tiga kelompok utama yaitu siswa dengan kemampuan tinggi, sedang, dan rendah. Evaluasi dengan Silhouette Score menunjukkan nilai tertinggi pada pengurangan dimensi hingga satu komponen PCA, yang menghasilkan klusterisasi paling optimal.

Kata Kunci - Clustering, Constrained K-Means, Principal Component Analysis, Reduksi Dimensi, Silhouette Score.

I. PENDAHULUAN

Sekolah Menengah Kejuruan (SMK) berperan sebagai lembaga pendidikan vokasi yang dirancang untuk menghasilkan lulusan yang terampil, kompeten, dan sesuai dengan kebutuhan dunia industri[1], [2], [3]. Salah satu upaya strategis yang diinisiasi oleh pemerintah untuk mendukung tujuan tersebut adalah implementasi program kelas industri[4]. Program ini bertujuan untuk menjembatani kesenjangan antara

dunia pendidikan dan dunia kerja melalui kerja sama dengan dunia usaha dan dunia industri (DUDI). Kelas industri tidak hanya berfokus pada peningkatan kompetensi teknis peserta didik sesuai standar industri, tetapi juga bertujuan untuk meningkatkan kualitas pendidikan vokasional secara menyeluruh serta memperkuat daya saing peserta didik di pasar kerja[5]. Dengan adanya program ini, diharapkan lulusan SMK dapat lebih siap menghadapi tantangan global dan

memberikan kontribusi nyata terhadap pengembangan sektor industri di Indonesia.

Pengelompokan peserta didik dalam kelas industri memiliki peran penting dalam merancang strategi pembelajaran yang efektif. Dengan mengelompokkan peserta didik berdasarkan minat, bakat, dan tingkat kemampuan, pendidik dapat menyusun metode pengajaran yang lebih terarah dan sesuai dengan kebutuhan individu maupun kelompok[6]. Hal ini memungkinkan peserta didik untuk menyerap pengetahuan dan keterampilan dengan lebih optimal, karena materi yang diberikan dapat disesuaikan dengan kemampuan dan tujuan karier mereka di masa depan. Selain itu, pengelompokan yang tepat juga mendorong kolaborasi antar peserta didik dengan tingkat pemahaman yang serupa, menciptakan lingkungan belajar yang produktif dan mendukung[7]. Strategi ini memastikan bahwa setiap peserta didik mendapatkan pengalaman belajar yang relevan dengan standar industri, sehingga lebih siap untuk bersaing di dunia kerja[8].

Pengelompokan peserta didik sering kali dilakukan menggunakan algoritma K-Means yang memungkinkan pengelompokan berdasarkan kemiripan (*similarity*) data[9]. Algoritma K-Means adalah salah satu metode klasterisasi yang paling populer dan banyak digunakan karena kesederhanaannya dan efisiensinya[9]. Algoritma ini efektif dalam mengelompokkan data yang memiliki kesamaan karakteristik, sehingga menghasilkan kelompok-kelompok yang homogen[10]. Namun, algoritma K-Means memiliki beberapa keterbatasan, terutama ketika diterapkan pada dataset berdimensi tinggi dan ketika terdapat keterbatasan tertentu (*constrains*) yang harus dipatuhi[11].

Masalah serupa juga ditemukan dalam bidang pendidikan, khususnya dalam pengelompokan peserta didik di kelas industri. Kelas industri seringkali membutuhkan pengelompokan peserta didik berdasarkan kemampuan atau minat mereka untuk tujuan pembelajaran yang lebih efektif[12]. Namun, salah satu tantangan utama adalah memastikan bahwa setiap klaster atau kelompok memiliki jumlah peserta didik yang seimbang (*balance*). Ketidakseimbangan dalam klaster dapat menyebabkan ketidakadilan dalam distribusi sumber daya pendidikan dan dapat mempengaruhi kualitas pembelajaran peserta didik[1].

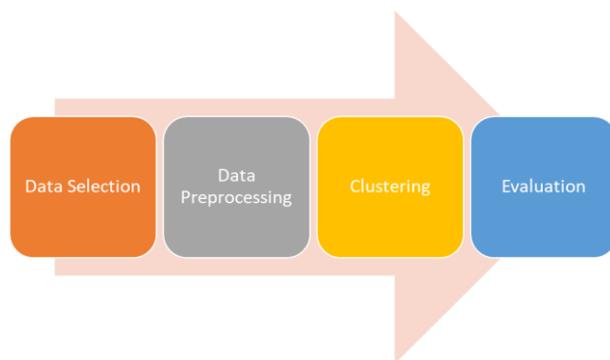
Banyak penelitian telah dilakukan untuk meningkatkan kinerja algoritma K-Means, salah satu algoritma yang digunakan untuk mengatasi persebaran data yang tidak merata menggunakan *Constrained K-means*[13]. Metode *Constrained K-Means* bekerja berdasarkan nilai K sebagai pusat cluster yang paling dekat dengan nilai N (anggota *cluster*) dengan pendekatan *Linear Programming Algorithm* (LPA) sehingga setiap cluster memiliki anggota N yang seimbang[14]. Namun, hasil evaluasi *Silhouette Score* menunjukkan hasil *clustering* algoritma ini lebih rendah dibandingkan dengan K-Means tradisional[15].

Salah satu cara untuk meningkatkan kualitas clustering adalah dengan mengurangi dimensi dataset. Reduksi dimensi bertujuan untuk menyederhanakan dataset dengan mempertahankan informasi yang paling signifikan, sehingga pola data menjadi lebih jelas dan mudah dikenali oleh algoritma clustering[16], [17]. Dengan mengurangi dimensi dataset sebelum melakukan klasterisasi dapat meningkatkan kinerja algoritma klasterisasi baik dari segi kecepatan maupun akurasi[18]. Salah satu metode yang sering digunakan untuk reduksi dimensi adalah *Principal Component Analysis* (PCA). PCA bekerja dengan mengidentifikasi arah variansi terbesar dalam data dan merepresentasikan data sepanjang komponen utama, yang merupakan kombinasi linear dari fitur asli[19]. Dengan menggunakan PCA, dataset yang kompleks dapat disederhanakan tanpa kehilangan informasi penting, sehingga memudahkan algoritma klasterisasi dalam mengenali pola, meningkatkan efisiensi komputasi, dan menghasilkan cluster yang lebih akurat dan terpisah dengan baik[20].

Penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengembangkan metode yang mengoptimalkan proses klasterisasi menggunakan algoritma *Constrained K-Means* yang dikombinasikan dengan reduksi dimensi menggunakan PCA. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam bidang analisis data, khususnya dalam konteks pengolahan dataset berdimensi tinggi dan klasterisasi dengan kendala. *Constrained K-Means* memastikan bahwa setiap cluster memiliki komposisi yang seimbang dan sesuai dengan kebutuhan industri. Dengan demikian, penelitian ini tidak hanya mengoptimalkan pendistribusian siswa dalam kelas industri, tetapi juga memberikan solusi yang lebih adil dan efisien untuk pemilihan siswa yang akan mengikuti program kelas industri, khususnya pada bidang kompetensi keahlian Rekayasa Perangkat Lunak di SMKN 2 Kraksaan.

II. METODE PENELITIAN

Metode penelitian ini dilakukan dengan langkah-langkah eksperimen yang terstruktur untuk mengoptimalkan algoritma *Constrained K-Means* menggunakan PCA. Berikut adalah bagan tahapan penelitian yang menggambarkan langkah-langkah eksperimen ini:



Gambar 1. Tahapan Penelitian

Seleksi data merupakan proses memilih subset data yang relevan dan berkualitas dari dataset yang lebih besar untuk digunakan dalam analisis atau pemodelan, seperti klustering[21]. Proses ini sangat penting untuk memastikan bahwa data yang digunakan dalam penelitian benar-benar representatif dan relevan dengan tujuan penelitian. Dataset yang digunakan merupakan hasil tes uji kemampuan akademik peserta didik SMKN 2 Kraksaan kelas X tahun ajaran 2024/2025 Rekayasa Perangkat Lunak berupa nilai deret, nilai matematika, nilai figural, nilai mengingat dan nilai analogi.

Tahapan preprocessing merupakan tahapan untuk menyiapkan data yang akan dilakukan *clustering* dengan metode:

- Pembersihan Data: Menghapus atau menangani data yang hilang, data outlier, dan inkonsistensi dalam data.
- Reduksi Dimensi: Jika dataset memiliki banyak fitur, teknik seperti *Principal Component Analysis* (PCA) dapat digunakan untuk mereduksi dimensi data tanpa kehilangan informasi penting.

Tahapan clustering menggunakan algoritma *constrained K-means* dengan menggunakan 3 cluster dengan batasan ukuran minimum cluster 16 dan maksimal 60. *Constrained K-Means Clustering* adalah varian dari algoritma K-Means yang memperhitungkan sejumlah batasan atau "constraints" dalam proses klustering[22]. Algoritma ini digunakan ketika ada kebutuhan untuk memastikan bahwa hasil klustering memenuhi beberapa aturan tertentu yang tidak diperhatikan dalam K-Means standar. *Constrained K-Means Clustering* menambahkan batasan-batasan pada K-Means untuk memastikan hasil klustering sesuai dengan aturan tertentu.

Evaluasi *Clustering* menggunakan *Silhouette Score* adalah metode yang sering digunakan untuk menilai seberapa baik hasil klustering dalam memisahkan data ke dalam kluster yang berbeda[23]. *Silhouette Score* memberikan gambaran tentang seberapa dekat setiap data poin dengan kluster yang seharusnya dibandingkan dengan kluster lainnya.

III. HASIL DAN PEMBAHASAN

Penelitian ini berfokus pada penerapan algoritma *Constrained K-Means Clustering* dengan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) untuk mengelompokkan siswa kelas X kompetensi keahlian Rekayasa Perangkat Lunak di SMKN 2 Kraksaan. Data yang digunakan dalam penelitian ini adalah hasil tes uji kemampuan akademik siswa yang mencakup beberapa atribut, yaitu nilai deret, nilai matematika, nilai figural, nilai mengingat, dan nilai analogi. Dalam sub-bab ini, akan disajikan hasil dari proses klustering yang dilakukan pada dataset yang telah direduksi dimensinya menggunakan PCA, diikuti dengan evaluasi hasil klustering menggunakan *Silhouette Score*.

A. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini berasal dari hasil Tes Kemampuan Dasar Akademik (TKDA) siswa kelas X kompetensi keahlian di bidang Rekayasa Perangkat Lunak di SMKN 2 Kraksaan sebanyak 96 siswa. TKDA mencakup berbagai kemampuan kognitif, termasuk keterampilan analogi, keterampilan urutan, kemampuan figural, keterampilan matematika, dan kemampuan mengingat.

NO	NAME	CLASS	FIGURAL	SEQUENCE	REMEMBERING	MATH	ANALOGY	
0	1	ADE RAKA MUHAIMIN	X-RPL-1	NaN	66.67	86.67	20.0	53.33
1	2	ADE RAKA MUHAIMIN	X-RPL-1	30.0	66.67	93.33	NaN	NaN
2	3	AFANDI SAHRONI	X-RPL-3	30.0	66.67	86.67	20.0	26.67
3	4	AFIF WAHYU FAIR SAFITRI	X-RPL-2	50.0	80.00	80.00	40.0	46.67
4	5	AHMAD DANIVAL HAKIM	X-RPL-3	60.0	66.67	66.67	30.0	13.33
--	--	--	--	--	--	--	--	--
91	92	ULIN NIKMAH FARAH KHOIRUN NISA	X-RPL-1	60.0	80.00	100.00	40.0	13.33
92	93	VIQI INDRAMA MAULANA	X-RPL-1	70.0	80.00	66.67	50.0	40.00
93	94	WIKIK	X-RPL-1	50.0	60.00	86.67	10.0	33.33
94	95	YUANDA FAJAR APRIANDA	X-RPL-3	70.0	46.67	93.33	NaN	NaN
95	96	YUNITA PUSPA NINGTIAS	X-RPL-3	30.0	66.67	73.33	40.0	NaN

96 rows x 8 columns

Gambar 2. Dataset Hasil TKDA

B. Preprocessing Data

Pada tahapan ini melakukan persiapan data dengan melakukan proses menghilangkan nilai *null*, menghilangkan duplikasi data serta menerapkan metode PCA untuk mereduksi dimensi dataset.

```

#Cek data kosong
data_null = data.isna().sum()
print(data_null)

#menghapus data kosong
data.dropna(inplace=True)

#Menghilangkan Duplikat Data
dataset = data.drop_duplicates(subset = ['NAME'], keep = 'last').reset_index(drop = True)
data_null = data.isna().sum()
print(data_null)
dataset

```

✓ 0.0s

Gambar 3. Proses Data Cleansing

Setelah melakukan *cleansing*, dataset menjadi 93 data. Sebelum melakukan reduksi dimensi menggunakan PCA, terlebih dahulu memilih fitur atau atribut yang akan dilakukan reduksi dimensi. Proses reduksi dimensi dapat dilakukan seperti pada gambar 4.

```

# Memilih fitur yang relevan untuk clustering
features = data[['FIGURAL', 'SEQUENCE', 'REMEMBERING', 'MATH', 'ANALOGY']]
# Visualisasi 2D menggunakan PCA
pca = PCA(n_components=2)
pca_features = pca.fit_transform(features)

```

✓ 0.0s

Gambar 4. Proses Reduksi Dimensi PCA

Dengan menggunakan PCA, data yang awalnya berada dalam lima dimensi dikurangi menjadi dua dimensi, yang memudahkan visualisasi dan analisis. Proses ini penting karena clustering dalam dimensi yang lebih rendah memungkinkan interpretasi yang lebih mudah dan dapat membantu dalam mengidentifikasi pola atau kelompok dalam data. Hasil transformasi PCA ini kemudian dapat digunakan untuk langkah selanjutnya dalam analisis data atau visualisasi clustering.

C. Clustering

Setelah reduksi dimensi, proses clustering dilakukan menggunakan algoritma *Constrained K-Means*. Algoritma ini diterapkan dengan mempertimbangkan *constraints* yang telah ditentukan, seperti memastikan setiap kluster memiliki distribusi jumlah siswa yang seimbang.

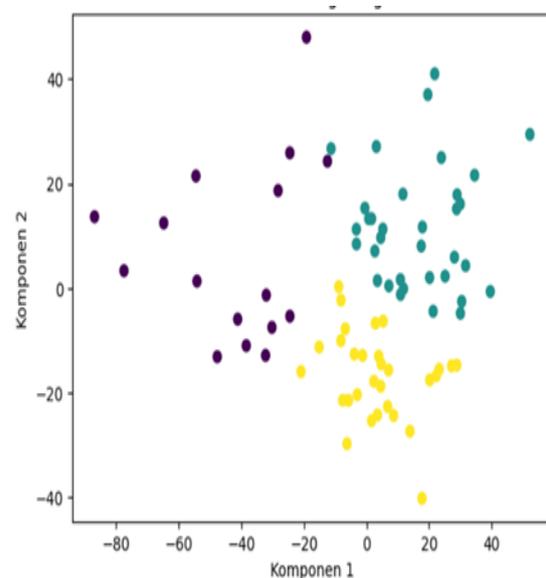
```

# Menjalankan Constrained K-Means dengan 3 cluster
ckmeans = KMeansConstrained(n_clusters=3, size_min=16, size_max=60, random_state=0)
ckmeans.fit(pca_features)
labels = ckmeans.labels_

```

Gambar 5. Model Constrained K-Means Clustering

Berdasarkan proses clustering yang dilakukan, data siswa berhasil dikelompokkan ke dalam tiga kluster yang seimbang seperti divisualisasikan pada gambar 6.



Gambar 6. Hasil Clustering Constrained K-Means

Dari hasil clustering di atas, Hasil clustering pada visualisasi menunjukkan bahwa data telah dikelompokkan menjadi tiga cluster dengan pemisahan yang cukup jelas. Ketiga cluster tampak memiliki jarak yang signifikan satu sama lain, yang mengindikasikan bahwa algoritma berhasil mengidentifikasi pola dan memisahkan data sesuai dengan karakteristiknya. Secara keseluruhan, pemisahan cluster sudah cukup baik dan seimbang antar cluster. Kemudian diambil rata-rata nilai pada masing-masing atribut berdasarkan pada tabel 1.

Tabel 1. Rata-rata Nilai Atribut Tiap Cluster

Figural	Sequene	Remem bering	Math	Analogy	Cluster
41	47	40	24	44	0
58	71	84	48	34	1
50	68	86	19	42	2

Tabel 1 di atas menunjukkan rata-rata skor untuk setiap atribut uji kemampuan akademik (Figural, Sequence, Remembering, Math, dan Analogy) berdasarkan hasil klasterisasi siswa ke dalam tiga klaster yang berbeda. Klaster 0 memiliki rata-rata skor yang relatif lebih rendah pada sebagian besar atribut, terutama pada Math dengan skor 24, menunjukkan siswa dalam klaster ini memiliki kemampuan yang lebih rendah dalam matematika. Klaster 1, dengan skor tinggi pada Remembering (84) dan nilai rata-rata yang lebih tinggi pada atribut lainnya, mengindikasikan kelompok siswa dengan kemampuan akademik yang lebih baik secara keseluruhan, terutama dalam daya ingat. Sementara itu, Klaster 2 memiliki skor tertinggi pada Remembering (86) tetapi skor rendah pada Math (19), menunjukkan kelompok siswa dengan kemampuan mengingat yang sangat baik namun lemah dalam matematika. Berdasarkan tabel tersebut kelas industry Kompetensi Keahlian Perangkat Lunak diambil dari siswa klaster 1.

D. Evaluasi

Hasil evaluasi *clustering* dengan PCA (*Principal Component Analysis*) menggunakan *Silhouette Score* menunjukkan bahwa jumlah komponen utama yang digunakan dalam pengurangan dimensi dataset sangat mempengaruhi kualitas *clustering*. Semakin tinggi nilai *Silhouette Score*, semakin baik pemisahan kluster yang tercapai seperti pada tabel 2.

Tabel 2. Evaluasi Clustering

No	Jumlah Komponen	Silhouette Score
1	4	0,2310949314925484
2	3	0,27708767011176183
3	2	0,37177190214955075

Pendekatan yang digunakan dalam penelitian ini merupakan gabungan metode *Constrained K-Means Clustering* dengan *Principal Component Analysis* (PCA) untuk mengelompokkan siswa berdasarkan hasil tes kemampuan akademik di SMKN 2 Kraksaan.

Hasil penelitian menunjukkan bahwa pengurangan dimensi data menggunakan PCA diikuti dengan penerapan *Constrained K-Means* menghasilkan kluster yang lebih jelas dan terdefinisi dibandingkan dengan data asli tanpa reduksi dimensi. Hasil ini selaras dengan teori dasar yang menyatakan bahwa reduksi dimensi menggunakan PCA dapat membantu meningkatkan kinerja algoritma *clustering* dengan menghilangkan noise dan menonjolkan pola-pola utama dalam data[24].

Jika dibandingkan dengan penelitian terdahulu yang menunjukkan bahwa penggunaan PCA dalam kombinasi dengan algoritma *clustering* meningkatkan kejelasan dan akurasi kluster[25], penelitian ini mendukung temuan tersebut. Reduksi dimensi hanya sampai dua komponen PCA memberikan hasil *clustering* yang paling optimal. Temuan ini sedikit berbeda dari studi yang dilakukan oleh Alkhayrat[26] yang menemukan bahwa menggunakan beberapa komponen PCA dapat lebih optimal dalam konteks yang berbeda seperti analisis pola pembelian konsumen.

IV. KESIMPULAN

Hasil klasterisasi menunjukkan bahwa siswa dapat dikelompokkan menjadi tiga kluster utama berdasarkan kemampuan akademik mereka, yaitu kelompok dengan kemampuan tinggi digolongkan pada kelas industry, sedang, dan rendah. Evaluasi menggunakan *Silhouette Score* menunjukkan bahwa pengurangan dimensi hingga dua komponen PCA memberikan hasil klasterisasi yang paling optimal, dengan pemisahan kluster yang jelas.

Penelitian ini berhasil mengembangkan metode pengelompokan siswa dengan mengintegrasikan algoritma *Constrained K-Means* dan *Principal Component Analysis* (PCA) untuk menangani dataset berdimensi tinggi serta memenuhi batasan tertentu dalam klasterisasi. Penggunaan PCA terbukti mampu mereduksi dimensi dataset tanpa kehilangan informasi penting, sehingga meningkatkan kinerja algoritma baik dari segi kecepatan maupun akurasi.

DAFTAR PUSTAKA

- [1] S. S. Danutirta, "Pengelolaan Kelas Industri Di Smk N 2 Klaten," *Hanata Widya*, vol. 7, no. 6, Art. no. 6, Oct. 2018.
- [2] S. N. Halizah, "Kesenjangan Kurikulum Smk Dengan Kebutuhan Industri," *ADIBA J. Educ.*, vol. 4, no. 2, pp. 227–233, 2024.
- [3] S. Sarifudin and A. Suhri, "Revitalisasi Sekolah Berbasis Teaching Factory (Tefa) Sebagai Rujukan Sekolah Menengah Kejuruan Swasta Di Pamekasan (Studi Kasus Di Smks Mambaul Ulum Bata-Bata Panaan Palengan Pamekasan),"

- Stud. Religia J. Pemikir. Dan Pendidik. Islam*, vol. 8, no. 2, pp. 317–334, 2024.
- [4] A. W. Bramantiya and D. Nurhadi, “Eksplorasi Daya Saing Lulusan SMK Kompetensi Keahlian Teknik Pengelasan Kelas Industri PT. INKA Melalui Sertifikasi Welder (Studi Kasus di SMK Negeri 1 Bendo),” *Didakt. J. Kependidikan*, vol. 13, no. 3, pp. 4193–4202, 2024.
- [5] N. Kurniasih, “Kurikulum Kolaborasi Komunitas sebagai Jawaban Terhadap Kebutuhan Skill Tenaga Kerja di SMK Al Muallim Kesugihan,” *Chatra J. Pendidik. Dan Pengajaran*, vol. 2, no. 1, pp. 28–37, 2024.
- [6] “Evaluasi Pelaksanaan Kelas Industri di SMK Muhammadiyah 1 Surakarta dan SMK Pancasila Surakarta | Jurnal Pendidikan Tambusai.” Accessed: Jul. 06, 2024. [Online]. Available: <https://jptam.org/index.php/jptam/article/view/10209>
- [7] M. V. Roesminingsih, T. R. Hariastuti, and F. Agustina, “Perencanaan Peningkatan Mutu Sekolah di SMKN Purwosari Bojonegoro,” *J. Pendidik. Tambusai*, vol. 6, no. 1, pp. 1892–1906, 2022.
- [8] N. Thantasia, “Manajemen Kerjasama Sekolah Dengan Dunia Usaha Dan Dunia Industri Dalam Meningkatkan Kompetensi Lulusan Pada Smk N 7 Bandar Lampung,” PhD Thesis, UIN RADEN INTAN LAMPUNG, 2023. Accessed: Dec. 22, 2024. [Online]. Available: <http://repository.radenintan.ac.id/29825/>
- [9] E. M. Fitri, R. R. Suryono, and A. Wantoro, “Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz,” *J. Komputasi*, vol. 11, no. 2, Art. no. 2, Nov. 2023, doi: 10.23960/komputasi.v11i2.12582.
- [10] A. H. Lutfiannisa, M. Maimunah, and P. Sukmasetya, “Clustering Data Pasien Berdasarkan Usia di Puskesmas Menerapkan Metode K-Means,” *J. Inf. Syst. Res. JOSH*, vol. 5, no. 2, Art. no. 2, Jan. 2024, doi: 10.47065/josh.v5i2.4755.
- [11] M. Capó, A. Pérez, and J. A. Lozano, “An efficient K-means clustering algorithm for tall data,” *Data Min. Knowl. Discov.*, vol. 34, no. 3, pp. 776–811, May 2020, doi: 10.1007/s10618-020-00678-9.
- [12] E. P. Utami, D. A. K. Handayani, and T. L. Prihandoko, “Implementasi Pelayanan Bimbingan Karir Dalam Program Kelas Industri Di Smk Nu Ma’arif Kudus,” *Emphaty Cons-J. Guid. Couns.*, vol. 5, no. 2, pp. 1–15, 2024.
- [13] Q. Qian, Y. Xu, J. Hu, H. Li, and R. Jin, “Unsupervised visual representation learning by online constrained k-means,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16640–16649. Accessed: Dec. 22, 2024. [Online]. Available: http://openaccess.thecvf.com/content/CVPR2022/html/Qian_Unsupervised_Visual_Representation_Learning_by_Online_Constrained_K-Means_CVPR_2022_paper.html
- [14] P. Huang, P. Yao, Z. Hao, H. Peng, and L. Guo, “Improved Constrained k-Means Algorithm for Clustering with Domain Knowledge,” *Mathematics*, vol. 9, no. 19, Art. no. 19, Jan. 2021, doi: 10.3390/math9192390.
- [15] M. Khairi, M. Syafiih, and A. Khairi, “Industry Class Clustering of Software Expertise Competency at SMKN 2 Kraksaan Using Constrained K-Means Clustering Algorithm,” *J. Comput. Netw. Archit. High Perform. Comput.*, vol. 6, no. 3, pp. 1128–1133, 2024.
- [16] P. Ray, S. S. Reddy, and T. Banerjee, “Various dimension reduction techniques for high dimensional data analysis: a review,” *Artif. Intell. Rev.*, vol. 54, no. 5, pp. 3473–3515, Jun. 2021, doi: 10.1007/s10462-020-09928-0.
- [17] P. Chhikara, N. Jain, R. Tekchandani, and N. Kumar, “Data dimensionality reduction techniques for Industry 4.0: Research results, challenges, and future research directions,” *Softw. Pract. Exp.*, vol. 52, no. 3, pp. 658–688, Mar. 2022, doi: 10.1002/spe.2876.
- [18] I. Maulana, A. M. Siregar, R. Rahmat, and A. Fauzi, “Optimization Of Machine Learning Model Accuracy For Brain Tumor Classification With Principal Component Analysis,” *J. Tek. Inform. Jutif*, vol. 5, no. 3, Art. no. 3, Jun. 2024, doi: 10.52436/1.jutif.2024.5.3.2058.
- [19] F. L. Gewers *et al.*, “Principal Component Analysis: A Natural Approach to Data Exploration,” *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–34, May 2022, doi: 10.1145/3447755.
- [20] J. Singh and D. Singh, “A comprehensive review of clustering techniques in artificial intelligence for knowledge discovery: Taxonomy, challenges, applications and future prospects,” *Adv. Eng. Inform.*, vol. 62, p. 102799, 2024.
- [21] S. Alelyani, J. Tang, and H. Liu, “Feature selection for clustering: A review. data clustering: Algorithms and applications, editor: Charu aggarwal and chandan reddy.” CRC Press, 2013.
- [22] T. F. Covoos, E. R. Hruschka, and J. Ghosh, “A study of k-means-based algorithms for constrained clustering,” *Intell. Data Anal.*, vol. 17, no. 3, pp. 485–505, 2013.
- [23] A. M. Bagirov, R. M. Aliguliyev, and N. Sultanova, “Finding compact and well-separated clusters: Clustering using silhouette coefficients,” *Pattern Recognit.*, vol. 135, p. 109144, 2023.
- [24] W. Jia, M. Sun, J. Lian, and S. Hou, “Feature dimensionality reduction: a review,” *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2663–2693, Jun. 2022, doi: 10.1007/s40747-021-00637-x.
- [25] D. Hedyati and I. M. Suartana, “Penerapan Principal Component Analysis (PCA) Untuk

Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro,” *JIEET J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 49–54, Dec. 2021, doi: 10.26740/jieet.v5n2.p49-54.

- [26] M. Alkhayrat, M. Aljnidi, and K. Aljoumaa, “A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA,” *J. Big Data*, vol. 7, no. 1, p. 9, Dec. 2020, doi: 10.1186/s40537-020-0286-0.