

Pemetaan Lintasan Karir Alumni Berdasarkan Analisis Cluster: Kombinasi K-Means dan Reduksi Dimensi Autoencoder

Daru Prasetyawan^{1,*}, Agus Mulyanto¹, Rahmadhan Gatra¹

¹ Program Studi Informatika, UIN Sunan Kalijaga Yogyakarta, Indonesia

* Correspondence: daru.prasetyawan@uin-suka.ac.id

Copyright: © 2025 by the authors

Received: 5 Maret 2025 | Revised: 13 Maret 2025 | Accepted: 27 Maret 2025 | Published: 15 April 2025

Abstrak

Pemetaan karir alumni merupakan salah satu aspek penting dalam evaluasi dan pengembangan program pendidikan tinggi. Analisis *cluster*, khususnya kombinasi metode *k-means* dan *autoencoder*, menjadi solusi efektif dalam mengelompokkan data karir alumni yang kompleks dan multi-dimensional. Penelitian ini bertujuan untuk menerapkan dan mengevaluasi kombinasi algoritma *k-means* dan *autoencoder* dalam pemetaan karir alumni berdasarkan IPK, masa studi, waktu tunggu, jenis pekerjaan, gaji, tingkat pekerjaan, dan kesesuaian bidang ilmu. *Autoencoder* digunakan untuk mereduksi dimensi, sementara *k-means* bertugas mengelompokkan alumni ke dalam kluster berdasarkan kemiripan profil karir mereka. Data yang digunakan dalam analisis diperoleh dari sistem *tracer* studi. Pra pemrosesan data *tracer* study dilakukan melalui beberapa tahapan, seperti *cleaning*, *encoding*, dan *normalization*. Hasil evaluasi menunjukkan bahwa kombinasi *k-means* dan *autoencoder* menghasilkan skor *silhouette* dan DBI yang lebih baik. Skor *silhouette* dengan *autoencoder* mencapai 0,6112, sedangkan tanpa *autoencoder* hanya 0,3956. Nilai DBI dengan *autoencoder* adalah 0,566, sementara tanpa *autoencoder* mencapai 1,022. Analisis *cluster* ini berhasil mengelompokkan data *tracer* study ke dalam enam kluster berdasarkan kesamaan profil karir. Hasil *clustering* menunjukkan kluster yang terbentuk lebih dipengaruhi oleh jenis pekerjaan alumni dan waktu tempuh studi.

Kata kunci: alumni; clustering; k-means; autoencoder

Abstract

Alumni career mapping is a crucial aspect of evaluating and developing higher education programs. Cluster analysis, particularly the integration of k-means and autoencoder methods, has emerged as an effective solution for grouping complex and multi-dimensional alumni career data. This study aims to implement and assess the combination of k-means and autoencoder algorithms in alumni career mapping based on GPA, study duration, waiting time, job type, salary, job level, and field of study suitability. The autoencoder is employed to reduce dimensions, while k-means clusters alumni into groups based on the similarity of their career profiles. The data used in the cluster analysis is sourced from the tracer study. Pre-processing of the tracer study data is conducted through several stages, including cleaning, encoding, and normalization. The evaluation results indicate that the combination of k-means and autoencoder yields superior Silhouette and DBI scores. The Silhouette score with the autoencoder achieved 0.6112, while without it, the score was only 0.3956. The DBI value with the autoencoder is 0.566, whereas without it, the DBI reached 1.022. This cluster analysis effectively grouped the tracer study data into six clusters based on similarities in career profiles. The clustering results suggest that the formed clusters are more influenced by the alumni's job type and duration of study.

Keywords: alumni; autoencoder; clustering; k-means



PENDAHULUAN

Alumni menjadi aset berharga bagi suatu perguruan tinggi dan masyarakat secara keseluruhan. Alumni memiliki peran strategis dalam peningkatan mutu pendidikan (Wathoni, 2021). Alumni dapat menjadi sumber informasi dan jaringan yang membantu mahasiswa dan sesama alumni dalam pengembangan karir. Umpan balik dari alumni memainkan peran penting dalam membentuk kualitas program dan reputasi institusional (Hadijaya et al., 2024). Selain itu, alumni juga berperan dalam pendidikan kewirausahaan di perguruan tinggi (Baroncelli et al., 2022). Pemetaan karir alumni merujuk pada proses menggambarkan posisi kerja dan karier alumni dari perguruan tinggi mereka. Pemetaan karir alumni bertujuan untuk mengetahui profil pekerjaan alumni, mengetahui kesesuaian profil lulusan dengan bidang pekerjaan, dan mengetahui lamanya masa tunggu sejak lulus sampai memperoleh pekerjaan pertama (Achmad et al., 2023).

Pemanfaatan teknologi kecerdasan buatan, termasuk di dalamnya teknologi pembelajaran mesin di berbagai bidang sudah banyak dilakukan. Salah satu metode pembelajaran mesin yang dapat digunakan dalam memetakan alumni adalah dengan analisis *cluster*. *Clustering* merupakan suatu metode analisis data yang bertujuan untuk mengelompokkan objek ke dalam kelompok atau cluster berdasarkan kesamaan karakteristik tertentu (Surya et al., 2023). Metode tersebut membagi data ke dalam kluster-kluster sehingga data dengan karakteristik serupa ditempatkan dalam satu kluster yang sama, sementara data dengan karakteristik berbeda dikelompokkan ke dalam kluster lainnya (Ashari et al., 2022). Beberapa jenis algoritma *clustering* yang populer antara lain *K-means* (Miraftabzadeh et al., 2023), *DBSCAN* (Deng, 2020), *K-medoids* (Martins et al., 2024), dan *Bayesian Clustering* (Zhu et al., 2023).

K-means adalah algoritma pengelompokan fundamental yang banyak digunakan dalam aplikasi akademis dan industri (Liu et al., 2023). Algoritma *k-means* bertujuan untuk meminimalkan jumlah jarak antara objek-objek di dalam kluster dengan pusat kluster masing-masing, dan mengelompokkannya secara iteratif (Cui, 2020). Dimensi data yang besar menyebabkan kompleksitas komputasi menjadi lebih tinggi di dalam *k-means clustering*. Setiap iterasi *k-means* memerlukan perhitungan jarak antara setiap titik data dengan setiap *centroid* untuk menentukan kluster yang tepat. Algoritma *k-means* merupakan salah satu algoritma *clustering* yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan jarak dari *centroid* (Wardhana & Winarno, 2024). Reduksi dimensi menjadi salah satu langkah penting dalam proses *k-means clustering*, terutama ketika bekerja dengan data yang memiliki dimensi besar.

Reduksi dimensi bertujuan untuk menyederhanakan data yang kompleks menjadi bentuk yang lebih mudah dikelola. Salah satu teknik yang digunakan dalam mereduksi dimensi data adalah *autoencoder*. *Autoencoder* dirancang untuk menghasilkan representasi data yang lebih ringkas dan informasi yang lebih penting dengan cara mengompresi dan kemudian melakukan dekomposisi data masukan. Reduksi dimensi tidak hanya mengurangi dimensi data asli tetapi juga membuat fitur tersembunyi lebih mudah dimengerti dalam *clustering*. Oleh karena itu, penggunaan teknik reduksi dimensi yang tepat dapat secara substansial meningkatkan kinerja *clustering* (Berahmand et al., 2024). *Autoencoder* merupakan salah satu bentuk jaringan syaraf tiruan yang terdiri dari dua bagian yaitu *encoder* dan *decoder* (Cortés et al., 2024). *Autoencoder* mengatur, mengompresi, dan mengekstraksi fitur, yang memungkinkan pembelajaran tanpa pengawasan dan ekstraksi fitur *non-linier* (Chen & Guo, 2023).

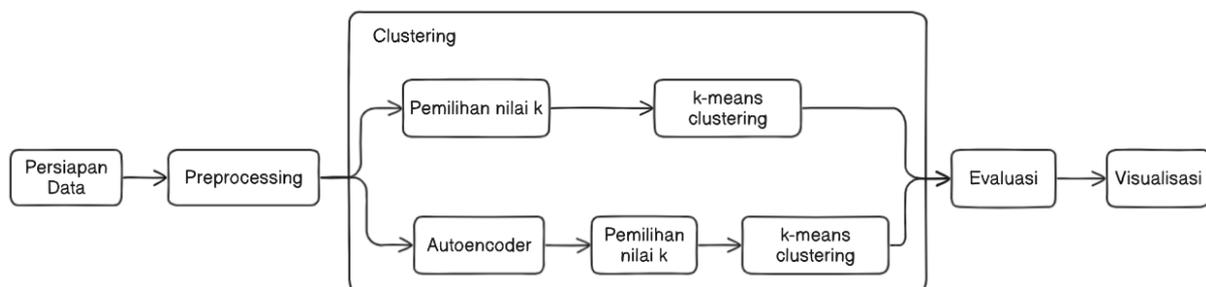
Penelitian sebelumnya telah banyak membahas tentang analisis data alumni. Analisis data alumni dapat digunakan untuk pemetaan alumni dan reorientasi profil lulusan pada sebuah program studi (Achmad et al., 2023). Alumni pada sebuah perguruan tinggi dipetakan berdasarkan IPK, masa studi, masa tunggu, jenis pekerjaan, kesesuaian bidang, dan lingkup pekerjaan (Sitorus et al., 2023). Penelusuran alumni yang lebih mendalam dilakukan dengan

menganalisis berbagai variabel, seperti biaya kuliah, metode pembelajaran, kepuasan alumni terhadap layanan kampus, aktivitas alumni, kesesuaian bidang kerja, masa tunggu, level pendidikan yang dibutuhkan pada pekerjaan sekarang, dan kompetensi alumni (Warni et al., 2023). Tetapi sebagian besar pengelompokan dan pemetaan alumni dilakukan secara tradisional, yaitu memisahkan data menjadi kelompok-kelompok yang berbeda berdasarkan variabel-variabel tertentu. Oleh karena itu, analisis terhadap variabel-variabel yang lebih kompleks sangat diperlukan untuk memperoleh hasil yang lebih komprehensif. Penelitian ini melakukan analisis data alumni menggunakan pendekatan pembelajaran mesin (*machine learning*), khususnya *clustering*.

Tujuan penelitian ini adalah menganalisis dan memetakan lintasa karir alumni dengan pendekatan pembelajaran mesin menggunakan kombinasi *k-means* dan reduksi dimensi *autoencoder*. *Autoencoder* digunakan untuk mereduksi dimensi, sementara *k-means* bertugas mengelompokkan alumni ke dalam kluster berdasarkan kemiripan profil karir mereka. Melalui teknologi pembelajaran mesin, diharapkan dapat menghasilkan pemetaan alumni secara komprehensif.

METODE

Penelitian ini dimulai dengan persiapan data, termasuk di dalamnya pengumpulan data dan analisis data. Kemudian dilanjutkan dengan pra-pemrosesan data dengan melakukan pembersihan data dan normalisasi atau penskalaan data. Proses *clustering* dilakukan dengan menggunakan 2 metode, yaitu *k-means* dan *k-means + autoencoder*. Hasil dari proses *clustering* kemudian dibandingkan dan dievaluasi. Tahapan dalam analisis *cluster* menggunakan *k-means* dan *autoencoder* dapat dilihat pada gambar 1.



Gambar 1. Tahapan dalam analisis *cluster*

Pengumpulan data menjadi langkah awal dalam penelitian, terutama untuk penelitian yang terkait dengan analisis data. Dalam penelitian ini, data yang digunakan adalah data *tracer study* (devalumni.uin-suka.ac.id) alumni yang lulus pada tahun 2021 sampai 2023 sebanyak 1062 data, yang setelah dibersihkan menjadi 1008 data. Pra-pemrosesan data mempersiapkan data mentah untuk diproses lebih lanjut. Pra-pemrosesan mencakup beberapa teknik seperti pembersihan data seperti menghapus duplikasi data dan nilai *null*, *encoding*, dan normalisasi atau standarisasi.

Data *tracer study* memiliki 3 jenis data, yaitu data numerik, kategorik nominal, dan kategorik ordinal. Untuk menangani jenis data kategorik nominal dan kategorik ordinal, penelitian ini menggunakan teknik *one-hot encoding* dan *ordinal encoding*. Di dalam *one-hot encoding*, setiap kategori dipetakan menjadi variabel baru yang berisi nilai 0 atau 1 (Dahouda & Joe, 2021). Di dalam *ordinal encoding*, setiap kategori diwakili sebuah bilangan bulat, dengan catatan bahwa urutan kategori telah diketahui. Selanjutnya data dinormalisasi untuk memperoleh skala yang sama pada setiap atribut data. Teknik yang digunakan untuk normalisasi data adalah *Z-score Normalization*. Tujuan utama dari *Z-score Normalization*

adalah mengubah skala data sehingga memiliki rata-rata (*mean*) bernilai nol dan simpangan baku (*standard deviation*) satu.

Selanjutnya *autoencoder* digunakan untuk mereduksi dimensi data sebelum proses *clustering* dilakukan. Jumlah kluster yang optimal ditentukan dengan menggunakan metode *elbow*. Hasil *clustering* selanjutnya dievaluasi menggunakan metode *silhouette coefficient* dan *Davis-Bouldin Index* (DBI) mengukur kinerja model *clustering*. *Silhouette coefficient* menggambarkan seberapa dekat suatu data dengan data lainnya di dalam kluster yang sama dibandingkan dengan data pada kluster lainnya. *Silhouette coefficient* memiliki rentang nilai antara -1 sampai 1. Nilai *silhouette* yang mendekati 1 menunjukkan adanya hubungan yang erat antara objek dengan kluster di mana objek tersebut berada. DBI adalah suatu metrik evaluasi untuk mengukur kualitas model *clustering*. Jumlah kluster yang memiliki DBI yang minimal dianggap sebagai jumlah kluster yang optimal (Baser & Saini, 2015).

HASIL DAN PEMBAHASAN

Hasil

Data *tracer study* alumni memiliki 7 atribut atau fitur yang digunakan dalam pemetaan pekerjaan alumni, yaitu ipk, masa studi, masa tunggu, kriteria pekerjaan, tingkat pekerjaan, gaji, dan kesesuaian bidang kerja. Data *tracer study* memiliki tiga jenis data, yaitu data numerik, kategorik nominal, dan kategorik ordinal. Pra-pemrosesan data dilakukan untuk memastikan data siap digunakan dalam pemodelan. Pengolahan data pada tahap ini antara lain pembersihan data dan standarisasi data. Tujuan pembersihan data di sini antara lain untuk menangani data yang hilang (*missing value*), memastikan tidak ada duplikasi data, dan menangani pencilan (*outliers*). Sedangkan standarisasi bertujuan agar semua variabel data memiliki skala nilai yang sama. Deskripsi data *tracer study* dapat dilihat pada tabel 1.

Tabel 1. Deskripsi data *tracer study*

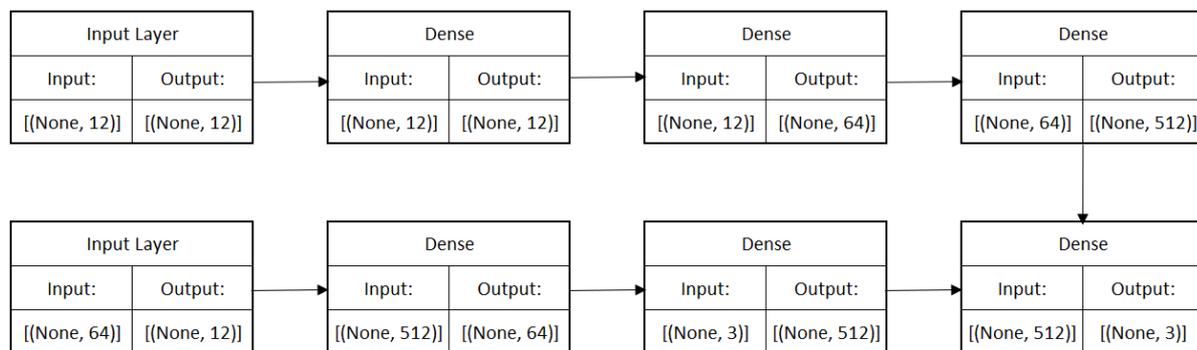
Atribut	Tipe data	Keterangan
IPK	Numerik	Min: Rp. 2.000.000; max: Rp. 9.000.000; rata-rata: Rp. 3.371.241
Masa studi	Kategorik Ordinal	Distribusi data: Tepat waktu (76%), Ovverstudi (24%)
Masa tunggu	Numerik	Min: 1 bulan, max: 12 bulan rata-rata: 3 bulan
Kriteria pekerjaan	Kategorik Nominal	Distribusi data: Perusahaan swasta (46,8%), BUMN/BUMD (11,7%), Lembaga pemerintahan (12,3%), Organisasi multilateral (13,9%), Wirausaha (15,3%)
Tingkat pekerjaan	Kategorik Ordinal	Distribusi data: Lokal (37,1%), Nasional (55,7%), Internasional (7,2%)
gaji	Numerik	Min: Rp. 2.000.000; max: Rp. 9.000.000; rata-rata: Rp. 3.371.241
Kesesuaian bidang	Kategorik Nominal	Distribusi data: Rendah (10,8%), Sedang (47,4%), Tinggi (41,8%)

Dataset *tracer study* memiliki 3 jenis data, sehingga diperlukan perlakuan yang berbeda terhadap ketiga jenis data tersebut. Untuk jenis data kategorik ordinal, data tersebut perlu diubah menjadi data numerik sebelum diubah menjadi skala yang ditentukan. Variabel tingkat pekerjaan memiliki nilai kategoris “Lokal”, “Nasional”, dan “Internasional”. Ketiga nilai kategorik tersebut memiliki nilai yang dapat diurutkan, yaitu “Lokal” lebih kecil dari “Nasional”, dan “Nasional” lebih kecil dari “Internasional”. Oleh karena itu nilai-nilai tersebut

dapat digantikan dengan nilai 0 untuk kategori “Lokal”, 1 untuk kategori “Nasional”, dan 2 untuk kategori “Internasional”. Sama halnya dengan variabel tingkat pekerjaan, variabel “kesesuaian bidang kerja” juga memiliki nilai kategorik yang dapat diurutkan, yaitu “Rendah”, “Sedang”, dan “Tinggi”. Oleh karena itu nilai pada variabel “kesesuaian bidang kerja” dapat diubah menjadi nilai 0 untuk mewakili kategori “Rendah”, 1 untuk kategori “Sedang”, dan 2 untuk kategori “Tinggi”. Proses encoding tersebut menjadikan data tracer studi memiliki 12 atribut, yaitu ipk, masa studi tepat waktu, masa studi overstudi, kriteria pekerjaan BUMD/BUMN, kriteria pekerjaan lembaga pemerintahan, kriteria pekerjaan organisasi multilateral, kriteria pekerjaan perusahaan swasta, kriteria pekerjaan wirausaha.

Selanjutnya standarisasi data dilakukan dengan metode *Z-Score Normalization* atau *standardization*. *Z-score* dipilih karena *z-score* mengukur seberapa jauh sebuah nilai berada dari rata-rata dalam satuan standar deviasi, sehingga *z-score* hanya mengubah skala dan lokasi data, tetapi tidak merubah bentuk distribusi data. Pustaka *sklearn* menyediakan fungsi *StandardScaler()* untuk normalisasi dataset. Data hasil standarisasi selanjutnya dapat digunakan dalam proses *clustering*.

Autoencoder dapat digunakan untuk kompresi dan reduksi dimensi. Pada penelitian ini, dataset akan direduksi menjadi 3 fitur. Pemodelan *autoencoder* dilakukan menggunakan keras dari *Tensorflow* yang dapat mempermudah dalam mendefinisikan model jaringan syaraf tiruan dan melatih model tersebut di dalam bahasa pemrograman *Python*. Arsitektur model *autoencoder* yang digunakan pada penelitian ini dapat dilihat pada gambar 2.

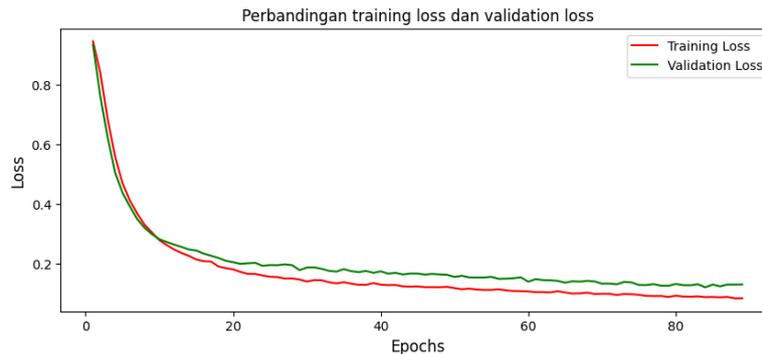


Gambar 2. Arsitektur model *autoencoder*

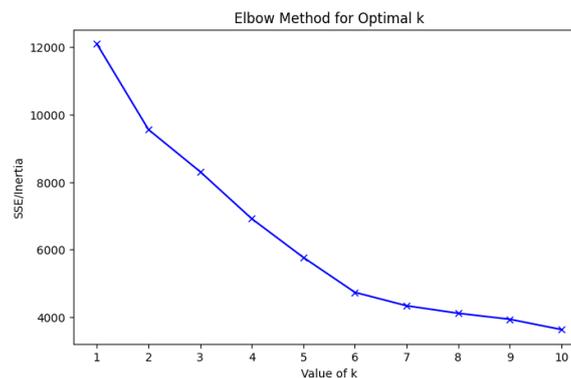
Proses pelatihan model *autoencoder* melibatkan pengoptimalan parameter dan struktur model sehingga model tersebut mampu memahami pola di dalam data. Parameter yang digunakan dalam pelatihan model tersebut adalah *optimizer* dan *loss*. *Optimizer* yang digunakan adalah *adam* dan pengukuran *loss* yang digunakan adalah *Mean Squared Error (MSE)*. Model dilatih menggunakan dataset yang sudah distandarisasi sebanyak 100 *epoch* dengan *batch size* sebanyak 32 data. Model *autoencoder* menerima input sejumlah fitur data yang digunakan untuk pelatihan. Selama proses pelatihan diperoleh *loss* sebesar 0,9338 pada *epoch* pertama dan terus menurun menjadi 0,1203 pada *epoch* ke-84. Kemudian *validation loss* tidak mengalami perbaikan setelah 5 *epoch* berikutnya, sehingga proses pelatihan dihentikan pada *epoch* ke 89. Perbandingan training *loss* dan *validation loss* yang dihasilkan selama proses pelatihan dapat dilihat pada gambar 3.

Penentuan jumlah kluster optimal menggunakan metode *Elbow*. Proses ini dimulai dengan menguji berbagai nilai *k* (jumlah kluster), lalu menghitung nilai *Sum of Squared Errors (SSE)* untuk setiap *k*. *SSE* mengukur total kesalahan kuadrat antara titik data dan *centroid* kluster masing-masing, yang mencerminkan seberapa rapat data dikelompokkan. Pada titik tertentu, penurunan *SSE* mulai melambat, yang terlihat pada grafik sebagai titik "*elbow*". Titik ini menunjukkan jumlah kluster yang optimal. Pada penelitian ini, hasil grafik *SSE* untuk setiap

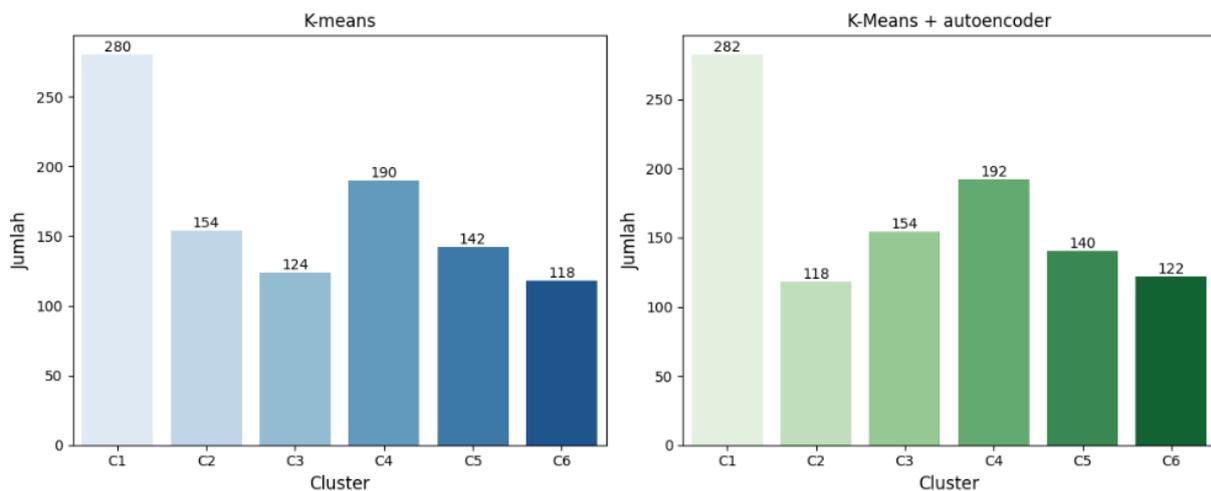
k menunjukkan bahwa pada $k=6$ terlihat membentuk siku. Hal ini menunjukkan bahwa penambahan jumlah kluster hanya menghasilkan sedikit perbaikan dalam pemisahan data, sehingga jumlah kluster yang optimal adalah 6, dan hasil SSE ditunjukkan pada gambar 4.



Gambar 3. Perbandingan *training loss* dan *validation loss*



Gambar 4. Hasil SSE untuk setiap k



Gambar 5. Distribusi kluster

Proses *clustering* dilakukan menggunakan bahasa pemrograman *Python* dengan memanfaatkan fungsi *Kmeans* yang disediakan oleh pustaka *scikit-learn*. *Scikit-learn* merupakan sebuah pustaka (*library*) dalam bahasa pemrograman *Python* untuk pembelajaran mesin (*machine learning*) dan tugas-tugas analisis data. *Clustering* pertama diterapkan pada data tanpa *autoencoder*. Proses *clustering* tersebut menghasilkan 6 kluster dengan jumlah anggota pada kluster pertama (C1) sebanyak 280 anggota, kluster kedua (C2) sebanyak 154 anggota, kluster ketiga (C3) sebanyak 124 anggota, kluster keempat (C4) sebanyak 190

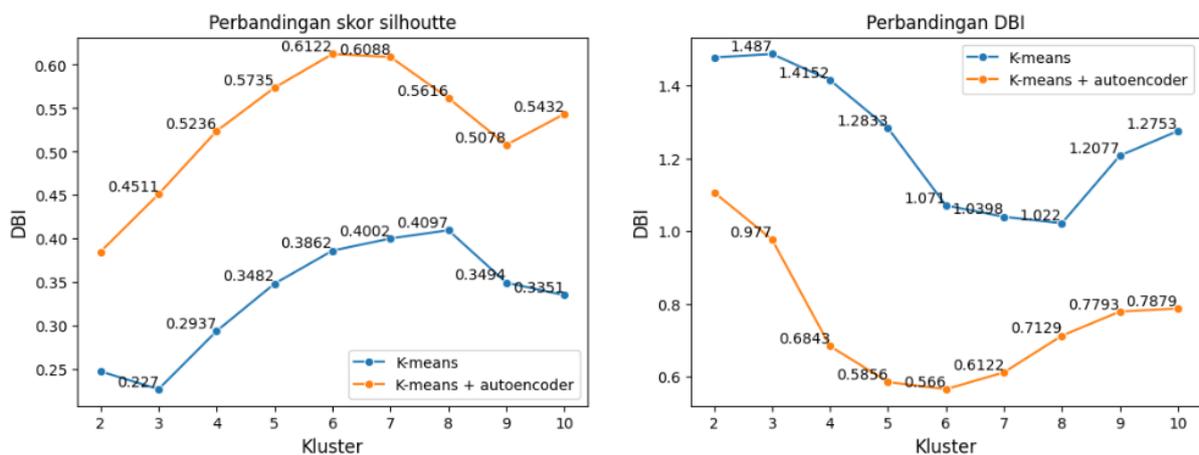
anggota, kluster kelima (C5) sebanyak 142 anggota, dan kluster keenam (C6) sebanyak 118 anggota. *Clustering* kedua dilakukan pada data hasil *autoencoder*. Dari proses tersebut diperoleh anggota pada kluster C1 sebanyak 282 anggota, kluster C2 sebanyak 118 anggota, kluster C3 sebanyak 154 anggota, kluster C4 sebanyak 192 anggota, kluster C5 sebanyak 140 anggota, dan kluster C6 sebanyak 122. Distribusi anggota kluster dari proses *clustering* menggunakan *k-means* dan kombinasi *k-means + autoencoder* disajikan pada gambar 5.

Evaluasi model *clustering* dilakukan dengan *silhouette coefficient* dan DBI. Skor *Silhouette* pada $k=6$ yang diperoleh model *clustering* dengan kombinasi *k-means* dan *autoencoder* lebih besar dibandingkan hanya dengan *k-means* saja, yaitu 0,6122 berbanding 0,3862. Pada *clustering* tanpa *autoencoder*, nilai *silhouette* terbaik terjadi pada nilai $k=7$ yaitu sebesar 0,3956. Kriteria kualitas kluster berdasarkan skor *silhouette* tabel 2.

Tabel 2. Kriteria skor *silhouette*

Kriteria	Interpretasi
0.71 – 1.00	Struktur kluster yang terbentuk sangat kuat.
0.51 – 0.70	Struktur yang terbentuk cukup baik dan bisa diterima.
0.26 – 0.50	Struktur kluster lemah, masih ada objek yang salah masuk kluster tertentu.
≤ 0.25	Tidak ditemukan struktur kluster yang substansial.

Evaluasi model menggunakan DBI menunjukkan bahwa model *clustering* kombinasi *k-means* dan *autoencoder* lebih kecil dibandingkan hanya menggunakan *k-means* saja. Pada nilai $k=6$, nilai DBI yang dihasilkan sebesar 0,566, sedangkan *clustering* tanpa *autoencoder* menghasilkan nilai DBI sebesar 1.071. Jika semua kandidat jumlah kluster $k=2$ sampai $k=10$ divalidasi menggunakan DBI, model *clustering* dengan kombinasi *k-means* dan *autoencoder* juga lebih baik dibandingkan tanpa *autoencoder*. Perbandingan skor *silhouette* dan nilai DBI pada model *clustering* menggunakan *k-means* tanpa *autoencoder* dan dengan *autoencoder* dapat dilihat pada gambar 6.

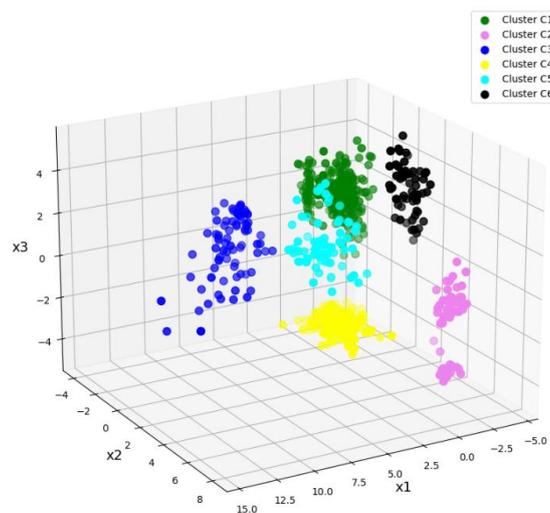


Gambar 6. Perbandingan DBI *k-means* dan *k-means + autoencoder*

Setiap kluster yang terbentuk dari proses *clustering* memiliki karakteristik masing-masing. Kluster C1 berisi alumni yang bekerja di perusahaan swasta dan menempuh studi dengan tepat waktu. Rata-rata IPK yang dimiliki oleh alumni yang berada pada kluster C1 adalah 3,69. Rata-rata gaji alumni yang berada pada kluster C1 adalah Rp. 3.259.995. Kluster C2 hanya berisi alumni yang bekerja di BUMN/BUMD. Sebanyak 80 alumni lulus tepat waktu dan 38 alumni lulus overstudi. Rata-rata IPK yang dimiliki oleh alumni yang berada pada

kluster C2 adalah 3,66. Rata-rata gaji alumni yang berada pada kluster C2 adalah Rp. 3.908.677. Kluster C3 hanya berisi alumni yang bekerja sebagai wirausaha dan menempuh studi dengan tepat waktu. Rata-rata IPK yang dimiliki oleh alumni yang berada pada kluster C3 adalah 3,71. Rata-rata gaji alumni yang berada pada kluster C3 adalah Rp. 4.293.129. Kluster C4 berisi sebagian besar alumni yang bekerja di perusahaan swasta. Semua alumni pada kluster ini menempuh studi overstudi. Rata-rata IPK yang dimiliki oleh alumni yang berada pada kluster C4 adalah 3,55. Rata-rata gaji alumni yang berada pada kluster C4 adalah Rp. 2.941.046. Kluster C5 berisi alumni yang bekerja pada organisasi multilateral dan menempuh studi dengan tepat waktu. Rata-rata IPK yang dimiliki oleh alumni yang berada pada kluster C5 adalah 3,42. Rata-rata gaji alumni yang berada pada kluster C5 adalah Rp. 2.918.642. Kluster C6 hanya berisi alumni yang bekerja pada lembaga pemerintahan. Semua alumni pada kluster ini menempuh studi dengan tepat waktu. Rata-rata IPK alumni yang berada pada kluster C6 adalah 3,72. Rata-rata gaji alumni yang berada pada kluster C6 adalah Rp. 3.141.278.

Reduksi dimensi menggunakan *autoencoder* juga dapat mempermudah dalam visualisasi sebaran data. *Autoencoder* mereduksi dimensi data menjadi 3 dimensi, sehingga data tersebut dapat divisualisasikan dalam bentuk 3D. Gambar 7 menunjukkan sebaran kluster dalam bentuk 3 dimensi. Setiap kluster terlihat terpisah meskipun terdapat kluster yang terlihat beririsan dan terdiri dari 6 *cluster*.



Gambar 7. Visualisasi hasil *clustering*

Pembahasan

Kombinasi *k-means* dan *autoencoder* secara efektif digunakan dalam pengelompokan alumni. Hasil evaluasi menunjukkan bahwa penggunaan kombinasi *k-means* dan *autoencoder* menjadikan *clustering* lebih optimal dan efisien. Skor *silhouette* yang dihasilkan dari kombinasi *k-means* dan *autoencoder* mencapai 0,6112 (dibandingkan 0,3956 tanpa *autoencoder*). Sedangkan nilai DBI yang dihasilkan dari kombinasi tersebut sebesar 0,566, (dibandingkan dengan 1,022 tanpa *autoencoder*). Hal ini menunjukkan kluster yang terbentuk menjadi lebih jelas dan terpisah dengan baik. Melalui analisis *cluster* ini, alumni berhasil dikelompokkan ke dalam 6 kluster berdasarkan kesamaan profil karir. Setiap kluster memiliki karakteristik yang berbeda, antara lain alumni yang bekerja pada perusahaan swasta dengan IPK tinggi dan masa studi tepat waktu, alumni yang bekerja di BUMN/BUMD dengan kombinasi masa studi tepat waktu dan overstudi, alumni yang berwirausaha dengan IPK tertinggi di antara kluster lainnya, alumni overstudi yang bekerja di perusahaan swasta dengan gaji yang lebih rendah, alumni yang bekerja di organisasi multilateral dengan masa studi tepat waktu, serta alumni yang bekerja di lembaga pemerintahan dengan IPK tinggi dan masa studi tepat waktu.

Penelitian yang dilakukan oleh Achmad et al. (2023) berhasil mengelompokkan alumni berdasarkan bidang pekerjaan (pendidikan, industri, wirausaha) dan waktu tunggu untuk memperoleh pekerjaan. Selanjutnya, penelitian yang dilakukan oleh Sitorus et al. (2023) berhasil mengidentifikasi alumni berdasarkan variabel-variabel tertentu seperti, IPK, masa studi, masa tunggu, jenis pekerjaan, kesesuaian bidang, dan lingkup pekerjaan. Sementara penelitian yang dilakukan oleh Warni et al. (2023) berhasil mengidentifikasi mengetahui status, kualitas, dan prestasi lulusan/alumni dengan melakukan analisis berdasarkan berbagai variabel, seperti biaya kuliah, metode pembelajaran, kepuasan alumni terhadap layanan kampus, aktivitas alumni, kesesuaian bidang kerja, masa tunggu, level pendidikan yang dibutuhkan pada pekerjaan sekarang, dan kompetensi alumni.

Beberapa hasil temuan sebelumnya, hanya mengelompokkan berdasarkan satu variabel tertentu. Pendekatan berbasis *machine learning* ini lebih komprehensif dalam analisis dan pengelompokan alumni. Penelitian-penelitian sebelumnya cenderung membagi data alumni dalam kategori tidak mempertimbangkan interaksi antar variabel. Sebaliknya, metode *clustering* berbasis *machine learning* mampu mengidentifikasi pola tersembunyi dalam data sehingga menghasilkan kluster yang lebih representatif.

SIMPULAN

Berdasarkan analisis data *tracer study* yang mencakup variabel seperti IPK, masa studi, waktu tunggu, jenis pekerjaan, gaji, tingkat pekerjaan, dan kesesuaian bidang ilmu, penelitian ini menghasilkan beberapa temuan penting. Kombinasi *k-means* dan *autoencoder* terbukti lebih efektif dibandingkan penggunaan *k-means* saja. *autoencoder* berhasil mereduksi dimensi data, meningkatkan kualitas *clustering* dengan nilai *silhouette score* 0,6112 (dibandingkan 0,3956 tanpa *autoencoder*) dan nilai DBI 0,566 (dibandingkan 1,022 tanpa *autoencoder*). Enam kluster karir alumni berhasil diidentifikasi, dengan karakteristik unik masing-masing, seperti alumni yang bekerja di perusahaan swasta, BUMN/BUMD, wirausaha, organisasi multilateral, dan lembaga pemerintahan. Pendekatan berbasis *machine learning* ini lebih komprehensif dibandingkan metode tradisional, karena mampu menangani kompleksitas data multi-dimensi dan mengungkap pola tersembunyi yang tidak terlihat dengan analisis konvensional. Implikasi praktis dari penelitian ini adalah memberikan dasar bagi perguruan tinggi untuk mengevaluasi dan mengembangkan program studi berdasarkan profil karir alumni, serta memfasilitasi kebijakan yang lebih terarah dalam mendukung kesiapan karir lulusan.

REFERENSI

- Achmad, F., Syariffuddin, Z., & Endryanyah, E. (2023). Application of Tracer Study for Alumni Mapping and Reorientation of Graduate Profiles. *Proceedings of the International Joint Conference on Arts and Humanities 2022 (IJCAH 2022)*, 891–897. Surabaya, Indonesia. https://doi.org/10.2991/978-2-38476-008-4_94
- Ashari, I. A., Negara, I. S. M., & Sumantri, R. B. B. (2022). Evaluasi Pembayaran Keuangan Siswa berdasarkan Penghasilan Wali Siswa menggunakan Metode Clustering K-Means. *Edumatic: Jurnal Pendidikan Informatika*, 6(2), 324–333. <https://doi.org/10.29408/edumatic.v6i2.6395>
- Baroncelli, A., Bolzani, D., & Landoni, M. (2022). Mapping the engagement of alumni organisations in entrepreneurship education and support at UK universities. *The International Journal of Management Education*, 20(2), 100648. <https://doi.org/10.1016/j.ijme.2022.100648>
- Baser, P., & Saini, J. R. (2015). Agent based Stock Clustering for Efficient Portfolio Management. *International Journal of Computer Applications*, 116(3), 35–41. <https://doi.org/10.5120/20317-2381>

- Berahmand, K., Daneshfar, F., Salehi, E. S., Li, Y., & Xu, Y. (2024). Autoencoders and their applications in machine learning: a survey. *Artificial Intelligence Review*, 57(2), 1-52. <https://doi.org/10.1007/s10462-023-10662-6>
- Chen, S., & Guo, W. (2023). Auto-Encoders in Deep Learning—A Review with New Perspectives. *Mathematics*, 11(8), 1–54. <https://doi.org/10.3390/math11081777>
- Cortés, D. G., Onieva, E., López, I. P., Trinchera, L., & Wu, J. (2024). Autoencoder-Enhanced Clustering: A Dimensionality Reduction Approach to Financial Time Series. *IEEE Access*, 12, 16999–17009. <https://doi.org/10.1109/ACCESS.2024.3359413>
- Cui, M. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. *Accounting, Auditing and Finance Clausius Scientific Press*, 1(1), 5-8.
- Dahouda, M. K., & Joe, I. (2021). A Deep-Learned Embedding Technique for Categorical Features Encoding. *IEEE Access*, 9, 114381–114391. <https://doi.org/10.1109/ACCESS.2021.3104357>
- Deng, D. (2020). DBSCAN Clustering Algorithm Based on Density. *7th International Forum on Electrical Engineering and Automation (IFEEA)*, 949–953. Hefei, China: IEEE. <https://doi.org/10.1109/IFEEA51475.2020.00199>
- Hadijaya, Y., Said, H. B., Siregar, M. F. Z., & Wirtati, I. (2024). A Comparative Study on Mapping Alumni Satisfaction Levels Regarding Doctoral Program Management in Indonesia and Malaysia. *Edutec: Journal of Education and Technology*, 8(2), 429-440. <https://doi.org/10.29062/edu.v8i2.1042>
- Liu, H., Chen, J., Dy, J., & Fu, Y. (2023). Transforming Complex Problems Into K-Means Solutions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7), 9149–9168. <https://doi.org/10.1109/TPAMI.2023.3237667>
- Martins, A. S. C., Araujo, L. R., & Penido, D. R. B. (2024). K-Medoids clustering applications for high-dimensionality multiphase probabilistic power flow. *International Journal of Electrical Power & Energy Systems*, 157, 109861. <https://doi.org/10.1016/j.ijepes.2024.109861>
- Miraftabzadeh, S. M., Colombo, C. G., Longo, M., & Foiadelli, F. (2023). K-Means and Alternative Clustering Methods in Modern Power Systems. *IEEE Access*, 11, 119596–119633. <https://doi.org/10.1109/ACCESS.2023.3327640>
- Sitorus, R. A., Arya, D., Dasopang, B. S., & Zufria, I. (2023). Analisis Tracer Study Alumni Program Studi S1 Ilmu Komputer UIN Sumatera Utara. *Jurnal Kridatama Sains Dan Teknologi*, 05(2), 411–420. <https://doi.org/10.53863/kst.v5i02.967>
- Suraya, S., Sholeh, M., & Lestari, U. (2023). Evaluation of Data Clustering Accuracy using K-Means Algorithm. *International Journal of Multidisciplinary Approach Research and Science*, 2(01), 385–396. <https://doi.org/10.59653/ijmars.v2i01.504>
- Wardhana, F. P., & Winarno, S. (2024). Analisis Pemain Terbaik Sepak Bola dengan menggunakan Algoritma K-Means. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 409–417. <https://doi.org/10.29408/edumatic.v8i2.27105>
- Warni, S., Komara, C., & Kaniadewi, N. (2023). The Tracer Study: An In-Depth Search of The English Education Program Graduates. *Ellter Journal*, 4(1), 93–106. <https://doi.org/10.22236/ellter.v4i1.11137>
- Wathoni, K. (2021). Alumni Menurut Perspektif Total Quality Management (TQM). *MA'ALIM: Jurnal Pendidikan Islam*, 2(01), 34-48. <https://doi.org/10.21154/maalim.v2i01.3036>
- Zhu, Z., Xu, M., Ke, J., Yang, H., & Chen, X. (Michael). (2023). A Bayesian clustering ensemble Gaussian process model for network-wide traffic flow clustering and prediction. *Transportation Research Part C: Emerging Technologies*, 148, 104032. <https://doi.org/10.1016/j.trc.2023.104032>