

Analisis Karier Alumni Perguruan Tinggi menggunakan *Web Scraping LinkedIn* dan *K-Means Clustering*

Analysis of Higher Education Alumni Careers using LinkedIn Web Scraping and K-Means Clustering

¹Qurrotul Aini*, ²Eri Rustamaji, ³Denina Nastiti Putri Amani, ⁴Elvi Fetrina

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, UIN Syarif Hidayatullah

^{1,2,3,4}Jl. Ir. H. Juanda No. 95 Ciputat Tangerang Selatan, Banten, Indonesia

*e-mail: qurrotul.aini@uinjkt.ac.id

(received: 15 May 2026, revised: 16 June 2026, accepted: 23 June 2026)

Abstrak

Penggunaan data alumni sebagai dasar evaluasi kurikulum masih menghadapi kendala akibat keterbatasan metode pengumpulan data konvensional, seperti survei manual yang sering menghasilkan tingkat respons rendah. Di sisi lain, LinkedIn menyediakan data karier alumni yang relatif lengkap dan mutakhir, namun pemanfaatannya dalam *tracer study* masih terbatas. Penelitian ini bertujuan menganalisis pola karier alumni perguruan tinggi menggunakan data LinkedIn yang diperoleh melalui teknik *web scraping* dan dianalisis menggunakan algoritma *K-Means clustering* dalam kerangka *Knowledge Discovery in Databases* (KDD). Pendekatan ini menerapkan kerangka KDD untuk menghasilkan pemetaan pola karier alumni berbasis data sebagai pelengkap *tracer study* konvensional. Data yang dianalisis terdiri atas 133 profil alumni melalui tahapan seleksi, prapemrosesan, transformasi, *clustering*, dan evaluasi. Hasil penelitian menunjukkan bahwa mayoritas alumni bekerja pada sektor teknologi dan berada pada tingkat karier menengah (*mid-level/specialist*). *Clustering* menghasilkan tiga kelompok pola karier yang merepresentasikan kecenderungan pada bidang proses bisnis dan operasional, pengembangan sistem dan teknologi, serta pemanfaatan data dan jaminan kualitas sistem. Temuan ini mengindikasikan sebaran kompetensi alumni pada aspek bisnis, data, dan teknologi. Namun, nilai *Silhouette Score* sebesar 0,0321 dan *Davies-Bouldin Index* sebesar 3,0487 menunjukkan kualitas pemisahan antar cluster masih rendah, sehingga hasil pengelompokan lebih tepat dipandang sebagai pemetaan awal pola karier alumni. Penelitian ini menunjukkan potensi pemanfaatan data media sosial profesional untuk mendukung analisis karier alumni berbasis data.

Kata kunci: pola karier alumni, *career analytics*, *K-Means clustering*, linkedin, *web scraping*.

Abstract

The use of alumni data to support curriculum evaluation continues to face challenges due to the limitations of conventional data collection methods, such as manual surveys, which often result in low response rates. Meanwhile, LinkedIn provides relatively comprehensive and up-to-date alumni career information; however, its potential for supporting tracer studies remains underutilized. This study aims to analyze the career patterns of higher education alumni using LinkedIn data collected through web scraping and analyzed with the K-Means clustering algorithm within the Knowledge Discovery in Databases (KDD) framework. The proposed approach applies the KDD process to generate a data-driven mapping of alumni career patterns as a complement to conventional tracer studies. The dataset consisted of 133 alumni profiles, which were processed through the stages of data selection, preprocessing, transformation, clustering, and evaluation. The results indicate that the majority of alumni are employed in the technology sector and occupy mid-level or specialist positions. The K-Means algorithm identified three distinct career clusters, representing career tendencies in business process and operations, systems and technology development, and data utilization and software quality assurance. These findings reveal the distribution of alumni competencies across business, data, and technology domains. However, the clustering quality was relatively low, as indicated by a Silhouette Score of 0.0321 and a Davies-Bouldin Index of 3.0487, suggesting limited separation among the identified clusters. Therefore, the clustering results should be interpreted as an initial

<http://sistemasi.ftik.unisi.ac.id>

mapping of alumni career patterns rather than definitive classifications. Overall, this study demonstrates the potential of professional social media data as a valuable resource for supporting data-driven alumni career analysis and complementing traditional tracer study practices.

Keywords: *careers pattern, data alumni, K-means clustering, linkedin, web scraping.*

1 Pendahuluan

Perguruan tinggi bertanggung jawab secara strategis untuk memastikan bahwa lulusan memiliki kemampuan yang sesuai dengan kebutuhan dunia kerja. Salah satu alat penting untuk menilai pencapaian ini adalah *tracer study* alumni, yang bertujuan untuk melacak lulusan dari pendidikan ke dunia kerja dan menilai kesesuaian kompetensi akademik dengan kebutuhan industri. *Tracer study* telah diakui secara luas sebagai alat untuk menjamin kualitas pendidikan tinggi karena mampu memberikan umpan balik langsung pada perubahan kurikulum, perumusan kebijakan institusional, dan peningkatan layanan karier [1], [2]. Namun, dalam realisasinya, banyak perguruan tinggi masih menghadapi banyak tantangan untuk menerapkan riset *tracer*. Metode berbasis kuesioner *online* sering menghasilkan pengisian yang subjektif, data yang tidak akurat, dan jumlah respons yang rendah. Data alumni tidak mencerminkan keadaan sebenarnya di lapangan, sehingga studi *tracer* strategis tidak lagi berguna untuk membuat keputusan [3].

Dengan munculnya platform media sosial profesional, ada kesempatan baru untuk mengatasi keterbatasan ini. Penelitian terbaru menunjukkan bahwa LinkedIn dapat digunakan secara efektif untuk memetakan karier alumni dan menganalisis keterkaitan pendidikan dengan dunia kerja, terutama di bidang teknologi dan bisnis digital. LinkedIn telah berkembang menjadi repositori data karier publik yang mencerminkan kondisi profesional pengguna secara relatif aktual, termasuk informasi tentang jabatan, sektor industri, mobilitas kerja, dan tingkat karier. Oleh karena itu, LinkedIn menjadi sumber penelitian data alternatif yang semakin relevan [4], [5].

Pendekatan *web scraping* banyak digunakan dalam penelitian pendidikan dan sistem informasi untuk mendapatkan data secara sistematis. *Scraping* web memungkinkan pengambilan data publik secara otomatis dan berskala besar, meningkatkan efisiensi pengumpulan data alumni dibandingkan dengan teknik manual. Selain itu, penelitian terbaru menunjukkan bahwa, jika dilakukan pada data publik dan sesuai dengan etika penelitian, *scraping* media sosial profesional dapat meningkatkan kelengkapan dan keakuratan penelitian lacak data [6], [7].

Tracer study alumni masih bersifat deskriptif (seperti persentase industri dan jabatan, tanpa mengelompokkan karier berdasarkan kesamaan karakteristik pekerjaan, tingkat karier, dan sektor industri) dan berfokus pada pengembangan sistem atau pelaporan distribusi pekerjaan alumni tanpa melakukan analisis pola karier berbasis *data mining*. Sebuah penelitian [8] menemukan bahwa penelitian *tracer* media sosial bertujuan untuk mengumpulkan informasi penting seperti tempat kerja dan profesi. Namun, penelitian tersebut tidak menyelidiki pemetaan pola karier yang lebih lanjut. Selain itu, penelitian yang memanfaatkan data LinkedIn untuk analisis alumni pada tingkat program studi di Indonesia masih relatif terbatas, padahal analisis pada tingkat ini penting untuk mengevaluasi kurikulum, profil lulusan, dan merumuskan strategi pengembangan kompetensi. Selain itu, teknik clustering masih belum banyak digunakan dalam studi *tracer* digital. Algoritma *K-Means* berhasil dalam studi pengelompokan *tracer*. Studi [9] menerapkan *K-Means* untuk mengelompokkan data pelacakan alumni berdasarkan IPK dan waktu menunggu pekerjaan. Hasil lain juga menunjukkan keunggulan *K-Means*, di mana [10] membuktikan bahwa *K-Means* mengungguli *K-Medoids* dan *X-Means* (DBI: -0,377 vs -0,930 dan -0,497); dan [11] menyatakan bahwa *K-Means* melampaui DBSCAN (*Silhouette Score*: 0,570 vs 0,258). Studi-studi ini secara kolektif menunjukkan efisiensi komputasi dan kualitas *cluster K-Means*.

Namun demikian, sebagian besar penelitian tersebut masih berfokus pada distribusi pekerjaan alumni atau pengelompokan berdasarkan variabel tertentu, sementara pemanfaatan data karier digital dari LinkedIn untuk memetakan pola karier alumni secara komprehensif masih terbatas. Selain itu, pelaporan metrik evaluasi *clustering* masih jarang dilakukan, sehingga kualitas dan validitas pola karier yang dihasilkan belum dapat dinilai secara objektif. Keterbatasan ini menunjukkan perlunya pendekatan yang tidak hanya mampu mengelompokkan data alumni, tetapi juga menyajikan evaluasi kualitas *cluster* secara terukur.

Berdasarkan latar belakang dan *research gap* tersebut, penelitian ini bertujuan untuk memetakan pola karier alumni Program Studi Sistem Informasi UIN Syarif Hidayatullah Jakarta menggunakan pendekatan *web scraping* data publik LinkedIn dan *clustering K-Means*. Adapun kontribusi penelitian ini adalah:

- Usulan pendekatan *tracer study* berbasis *web scraping* LinkedIn sebagai alternatif pelengkap *tracer study* yang sudah ada.
- Pemetaan pola karier alumni program studi, sehingga lebih kontekstual dan aplikatif.
- Penerapan *K-Means clustering* sebagai pendekatan eksploratif untuk menghasilkan pemetaan awal pola karier alumni berdasarkan karakteristik pekerjaan dan sektor industri.

2 Tinjauan Literatur

Sebagai hasil dari penelusuran literatur yang menggunakan kerangka PICO (populasi, intervensi, perbandingan, hasil), enam penelitian utama ditemukan terkait langsung dengan analisis karier alumni menggunakan data LinkedIn. Dengan menggabungkan metode *scraping web* dan *clustering K-Means* untuk mendeskripsikan profil LinkedIn pengguna umum, [12] membuat kemajuan dalam bidang ini. Sebaliknya, [13] berkonsentrasi pada analisis karier alumni teknologi informasi melalui pengambilan dan klasifikasi posisi LinkedIn. Selain itu, [14] untuk melihat jalur karier dan kemajuan profesional secara deskriptif, kami mengumpulkan data profil alumni LinkedIn dari berbagai institusi. Bahkan, sistem yang menggunakan teknik *scraping web* untuk mengekstrak dan memperbarui data alumni telah dikembangkan [15]. Seperti [16] memberikan rekomendasi pekerjaan yang dipersonalisasi dengan menggunakan AI untuk menganalisis *resume* dan *scraping* LinkedIn. Selain itu, perbedaan gender dalam jalur karier di industri konstruksi dipelajari [17] dengan menggunakan data LinkedIn dan pembelajaran mesin. Keenam penelitian ini menjadi referensi utama karena memiliki topik yang sama (analisis karier atau profil profesional), metode yang sama (menggunakan data LinkedIn), atau teknik yang sama (*scraping web* dan/atau *clustering*). Tabel 1 memberikan ringkasan keenam penelitian sebelumnya berdasarkan populasi, metode utama, dan temuan utama.

Tabel 1 Ringkasan penelitian terdahulu yang relevan

Peneliti	Populasi	Metode	Temuan Utama
[12]	Pengguna LinkedIn umum	<i>Web scraping</i> + <i>K-Means clustering</i>	Profil LinkedIn terkelompok ke kategori TI dan manajemen sebagai yang dominan
[13]	Alumni teknologi informasi	<i>Web scraping</i> + klasifikasi jabatan	Jabatan dominan: <i>software engineer</i> dan <i>project manager</i>
[14]	Alumni berbagai institusi	<i>Web scraping</i> + analisis deskriptif	Alumni teknologi cenderung di perusahaan besar; lainnya di pemerintah/pendidikan
[15]	Alumni dari berbagai profil web	<i>Web scraping</i> untuk ekstraksi data	Teknik <i>scraping</i> efektif untuk memperoleh dan memperbarui data alumni
[16]	Pencari kerja aktif	<i>Scraping</i> + AI untuk rekomendasi pekerjaan	AI meningkatkan akurasi rekomendasi pekerjaan di sektor teknologi
[17]	Profesional industri konstruksi	<i>Machine learning</i> + analisis gender	Perempuan masih terkonsentrasi di posisi lebih rendah dibanding laki-laki

Berikut ini adalah rangkuman beberapa hal penting dan keterbatasan penelitian sebelumnya, berdasarkan Tabel 1. Pertama, dari segi populasi, sebagian besar penelitian menggunakan populasi umum atau lintas institusi. Ini terlihat dalam penelitian [12] yang memeriksa pengguna LinkedIn umum, [14] yang memeriksa alumni dari berbagai institusi tanpa terbatas pada program studi mereka, dan [17] bahkan membatasi penelitian pada industri tertentu (konstruksi). Akibatnya, rekomendasi yang dibuat biasanya bersifat umum dan sulit untuk diterapkan secara langsung pada program studi tertentu. Kedua, dari sudut pandang metode, studi [12] telah menggunakan *clustering K-Means*, tetapi tidak memberikan penilaian objektif terhadap kualitas *clustering* yang dihasilkan. Studi tersebut juga

tidak mengukur *Silhouette Score* atau *Davies-Bouldin Index* (DBI). Studi lain, seperti [13] dan [14], hanya menggunakan analisis deskriptif atau klasifikasi sederhana daripada menggunakan *clustering* sama sekali. Ketiga, dari sudut pandang kedalaman analisis, [15] hanya melakukan ekstraksi data tanpa melakukan analisis lebih lanjut, sementara [16] lebih berfokus pada rekomendasi pekerjaan individu daripada pemetaan pola karier secara keseluruhan yang bermanfaat bagi institusi pendidikan. Secara umum, belum ada penelitian yang secara sistematis menggunakan kerangka *Knowledge Discovery in Databases* (KDD) lengkap, dari *scraping* hingga evaluasi *clustering*, pada data alumni program studi tertentu.

Penelitian saat ini memiliki beberapa perbedaan dibandingkan enam studi sebelumnya. Penelitian ini secara khusus menargetkan alumni Program Studi Sistem Informasi UIN Syarif Hidayatullah Jakarta. Berbeda dengan penelitian [12] yang berfokus pada profesional di industri tertentu, penelitian ini menghasilkan konteks analisis yang lebih spesifik pada tingkat program studi. Dari aspek metodologi, penelitian ini menerapkan kerangka KDD secara sistematis dan lengkap, mulai dari pengumpulan data sebanyak 133 profil alumni, praproses data melalui normalisasi jabatan, pemetaan industri, dan pembersihan teks, transformasi data menggunakan label *encoding* dan standarisasi, hingga proses *clustering* menggunakan *K-Means* dengan metode *Elbow* untuk menentukan jumlah *cluster* yang digunakan dalam analisis serta evaluasi hasil menggunakan *Silhouette Score* sebesar 0,0321 dan *Davies-Bouldin Index* sebesar 3,0487. Melalui pendekatan tersebut, penelitian ini memberikan gambaran awal mengenai kecenderungan pola karier alumni berdasarkan karakteristik pekerjaan dan industri. Dengan demikian, penelitian ini melengkapi penelitian sebelumnya yang umumnya masih berfokus pada analisis deskriptif atau ekstraksi data alumni, serta menunjukkan potensi pemanfaatan data LinkedIn dan kerangka KDD sebagai pendekatan pelengkap dalam *tracer study* berbasis data.

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan kerangka kerja *Knowledge Discovery in Databases* (KDD). KDD dipilih karena menyediakan alur sistematis untuk menggali pola tersembunyi dari data mentah, mulai dari seleksi, pra-pemrosesan, transformasi, hingga interpretasi. Fokus utama penelitian adalah mengelompokkan pola karier alumni berdasarkan kemiripan karakteristik pekerjaan menggunakan algoritma *K-Means clustering*, sehingga desain ini bersifat *unsupervised learning*.

Populasi dan Sampel

Populasi penelitian adalah seluruh alumni Program Studi Sistem Informasi UIN Syarif Hidayatullah Jakarta yang memiliki profil aktif dan publik di platform LinkedIn. Sampel ditentukan melalui *purposive sampling* dengan kriteria inklusi: (1) alumni tercantum sebagai lulusan program studi tersebut, (2) profil menyediakan informasi pekerjaan terkini (jabatan, perusahaan, industri), dan (3) data dapat diakses secara publik tanpa memerlukan koneksi tingkat lanjut. Jumlah sampel yang berhasil dikumpulkan adalah 133 profil alumni. Jumlah tersebut merepresentasikan alumni yang memenuhi kriteria inklusi dan memiliki profil LinkedIn yang aktif serta dapat diakses secara publik pada saat pengumpulan data dilakukan. Meskipun jumlah sampel relatif terbatas, data tersebut dipandang cukup untuk mendukung analisis eksploratif awal terhadap pola karier alumni yang memiliki profil LinkedIn aktif dan dapat diakses secara publik.

Instrumen dan Perangkat Penelitian

Penelitian ini menggunakan perangkat keras Apple M2, RAM 8 GB DDR4, SSD 512 GB, dan perangkat lunak *Python 3.12.0*, *Apify 1.2.0*, *Google Colab*, *MacOS Ventura*. Pustaka *Python* yang digunakan meliputi:

- *Pandas & NumPy*: manipulasi dan pemrosesan data tabular;
- *Scikit-learn*: implementasi *K-Means clustering*, *StandardScaler*, *LabelEncoder*, *TF-IDF*, serta metrik evaluasi (*silhouette score*, *Davies-Bouldin index*);
- *Yellowbrick*: visualisasi *silhouette plot*;
- *Matplotlib & Seaborn*: visualisasi data dan hasil *clustering*.

Prosedur Pengumpulan Data

Pengumpulan data dilakukan menggunakan *actor* LinkedIn pada platform *Apify* untuk memperoleh informasi profil yang tersedia secara publik. Autentikasi sesi digunakan hanya untuk mendukung proses akses dan stabilitas pengambilan data.

Adapun prosedur pengumpulan data dengan mengidentifikasi daftar URL profil LinkedIn alumni berdasarkan afiliasi institusi; mengkonfigurasi *actor Apify* dengan fitur manajemen request yang tersedia untuk menjaga stabilitas proses pengumpulan data; mengekstraksi atribut profil yang relevan, meliputi: *job title* (jabatan), *company name* (nama perusahaan), *industry name* (sektor industri), dan *geo location* (lokasi); dan menyimpan data dalam format CSV untuk tahap analisis selanjutnya. Pemilihan *Apify* didasarkan pada keunggulannya dalam menangani *scraping* skala besar dengan fitur manajemen *request* dan *proxy* yang terintegrasi, yang mengatasi keterbatasan metode konvensional seperti *Selenium* yang rentan terblokir. Pengumpulan data dilakukan hanya pada informasi yang tersedia secara publik pada profil LinkedIn dan tidak mencakup data pribadi yang bersifat sensitif. Data yang digunakan dibatasi pada atribut yang relevan dengan tujuan penelitian, yaitu jabatan, perusahaan, industri, dan lokasi kerja. Hasil analisis disajikan dalam bentuk agregat sehingga identitas individu tidak dapat diidentifikasi secara langsung. Namun demikian, penggunaan data LinkedIn berpotensi menimbulkan bias representasi karena hanya mencakup alumni yang memiliki profil aktif dan terbuka untuk publik.

Tahapan Analisis Data

Analisis data dilaksanakan mengacu pada kerangka KDD yang terdiri atas lima tahapan utama, seperti disajikan pada Gambar 1. Tahap pertama adalah seleksi data (data selection), di mana dipilih empat atribut yang relevan untuk pemetaan pola karier alumni, yaitu *job title*, *company*, *industry*, dan *location*. Keempat atribut tersebut digunakan sebagai dasar proses prapemrosesan dan analisis deskriptif. Adapun proses clustering dilakukan menggunakan fitur yang merepresentasikan karakteristik pekerjaan, yaitu *job title*, *job level*, dan *industry category* yang telah ditransformasikan ke dalam bentuk numerik. Pemilihan atribut tersebut didasarkan pada pertimbangan bahwa *job title* merepresentasikan fungsi pekerjaan, *company* menggambarkan konteks organisasi tempat alumni bekerja, *industry* menunjukkan sektor karier yang ditekuni, sedangkan *location* merefleksikan distribusi geografis pekerjaan. Keempat atribut tersebut dipilih karena tersedia secara konsisten pada sebagian besar profil LinkedIn dan dinilai mampu memberikan representasi yang memadai terhadap pola karier alumni. Tahap kedua adalah prapemrosesan data (preprocessing) yang bertujuan membersihkan dan menstandarkan data sebelum dilakukan transformasi. Proses yang dilakukan meliputi:

a) Normalisasi kategori pekerjaan (*job title normalization*)

Job title pada LinkedIn seringkali bervariasi meskipun menggambarkan fungsi pekerjaan yang sama. Normalisasi dilakukan menggunakan *keyword mapping* berdasarkan pencocokan kata kunci. Contoh: "backend developer", "frontend developer", "fullstack engineer" dikelompokkan ke dalam kategori *software engineer/developer*.

b) Ekstraksi tingkat karier (*career level extraction*)

Tingkat karier alumni diekstraksi dari *job title* berdasarkan keberadaan kata kunci tertentu. Kriteria yang digunakan:

- Junior/entry level: mengandung kata junior, jr, intern, entry, assistant, trainee;
- Senior/manager/executive: mengandung senior, lead, principal, head, manager, director, chief, executive;
- Mid-level/specialist: untuk semua jabatan yang tidak memenuhi kedua kriteria sebelumnya.

c) Normalisasi kategori industri (*industry mapping*)

Nama perusahaan dipetakan ke kategori industri berdasarkan pencocokan kata kunci. Kategori yang digunakan meliputi: 'technology', 'finance/banking', 'government/BUMN', 'consulting', dan 'other'.

d) Penentuan ukuran perusahaan

Ukuran perusahaan dikategorikan menjadi 'small', 'medium', dan 'large' berdasarkan informasi yang tersedia pada profil LinkedIn. Variabel ini digunakan pada tahap analisis deskriptif untuk menggambarkan distribusi alumni berdasarkan ukuran perusahaan tempat bekerja.

Tahap ketiga adalah transformasi data (*data transformation*), di mana data hasil preprocessing diubah ke dalam bentuk numerik yang dapat diproses oleh algoritma *K-Means*:

a) *Label encoding*

Variabel kategori (*job level, industry category*) dikonversi menjadi nilai numerik menggunakan *LabelEncoder* dari *scikit-learn*.

b) TF-IDF Vektorisasi pada Job Title

Teks dari kolom *job title* diubah menjadi representasi numerik menggunakan *Term frequency-inverse document frequency* (TF-IDF).

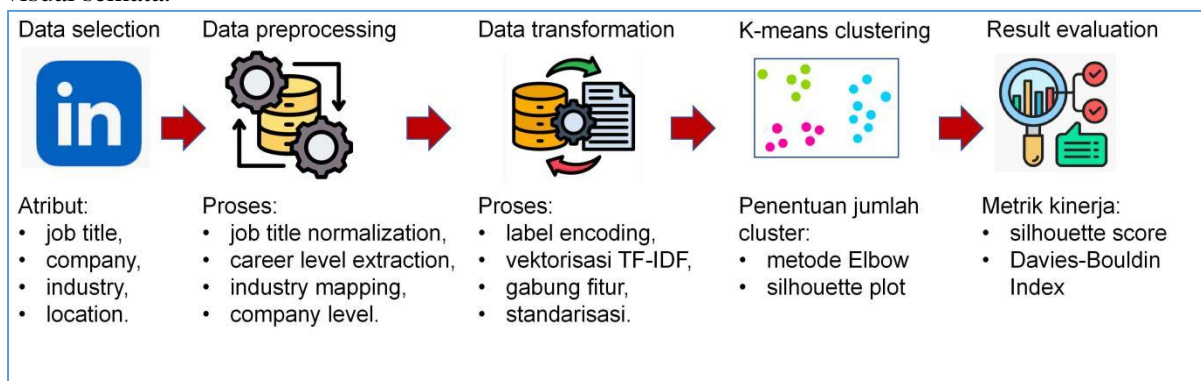
c) Penggabungan fitur

Matriks hasil TF-IDF ($X_{\text{job_tfidf}}$) digabungkan dengan fitur numerik hasil *label encoding* (X_{numeric}) menggunakan *hstack* dari *scipy.sparse*.

d) Standardisasi

Karena *K-Means* sensitif terhadap skala fitur, seluruh fitur distandardisasi menggunakan *StandardScaler* dengan parameter `with_mean=False` karena matriks fitur bersifat sparse.

Tahap keempat adalah *K-Means clustering*. Algoritma *K-Means* dipilih karena memiliki efisiensi komputasi yang baik, mampu menangani data berdimensi tinggi hasil transformasi TF-IDF, serta telah banyak digunakan dalam penelitian *clustering* data alumni dan profil profesional. Selain itu, beberapa penelitian terdahulu menunjukkan bahwa *K-Means* mampu menghasilkan performa yang kompetitif dibandingkan metode *clustering* lainnya pada kasus yang serupa [9]–[11]. Penentuan jumlah cluster dilakukan menggunakan metode *Elbow* sebagai dasar pemilihan jumlah cluster yang digunakan dalam analisis. Metode *Elbow* digunakan untuk memplot nilai inerti (*sum of squared distances*) pada rentang $k = 2$ hingga $k = 7$. Jumlah cluster dipilih pada titik ketika penurunan inerti mulai melandai atau membentuk *elbow point*. Sementara itu, *silhouette plot* digunakan untuk melihat kualitas pemisahan cluster. Nilai yang semakin mendekati 1 menunjukkan pemisahan *cluster* yang semakin baik. Algoritma *K-Means* diimplementasikan menggunakan *KMeans* dari *scikit-learn* dengan parameter `n_clusters` (jumlah cluster yang diperoleh berdasarkan metode *Elbow*); `random_state = 42` (untuk reproduksibilitas); dan `n_init = 'auto'` untuk inialisasi *centroid* otomatis. Proses *clustering* menghasilkan label *cluster* untuk setiap alumni, yang kemudian ditambahkan sebagai kolom baru dalam *dataset*. Selanjutnya tahap kelima adalah evaluasi hasil *clustering*, di mana kualitas hasil *clustering* dievaluasi menggunakan dua metrik internal, (1) *Silhouette score* (seberapa padat dan terpisahnya *cluster* yang terbentuk), dimana nilai berkisar antara -1 hingga 1, dengan nilai $> 0,5$ mengindikasikan struktur *cluster* yang baik; (2) DBI, yaitu mengukur rasio antara sebaran dalam dan antar *cluster*. Nilai yang lebih rendah (mendekati 0) menunjukkan pemisahan *cluster* yang lebih baik [18]. Evaluasi menggunakan kedua metrik tersebut dilakukan untuk memberikan penilaian yang lebih objektif terhadap kualitas cluster yang terbentuk dan mengurangi ketergantungan pada interpretasi visual semata.



Gambar 1 Tahapan analisis data

4 Hasil dan Pembahasan

4.1 Data Preprocessing

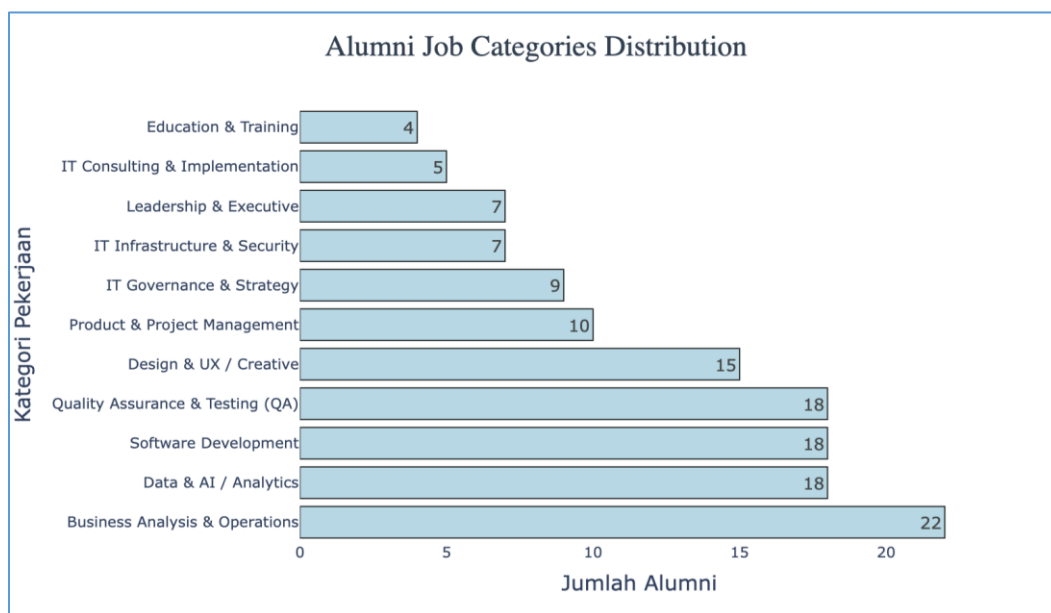
Data hasil *scraping* masih memiliki variasi penulisan dan *noise*, sehingga perlu dilakukan pra-proses.

- Normalisasi kategori pekerjaan (*job title normalization*), *job title* yang beragam diseragamkan menggunakan pencocokan kata kunci. Contoh: Backend developer, frontend developer, fullstack engineer → Software engineer/developer. Data scientist, machine learning engineer → Data & AI.
- Normalisasi kategori industri (*industry mapping*), di mana industri tempat alumni bekerja dikelompokkan menjadi beberapa kategori:
 - Technology (perusahaan teknologi seperti Tokopedia, Gojek, Google)
 - Finance/banking (bank, lembaga keuangan)
 - Government/BUMN (kementerian, BUMN, universitas)
 - Consulting (konsultan seperti Deloitte, Accenture)
 - Other

Eksplorasi dan Visualisasi Data

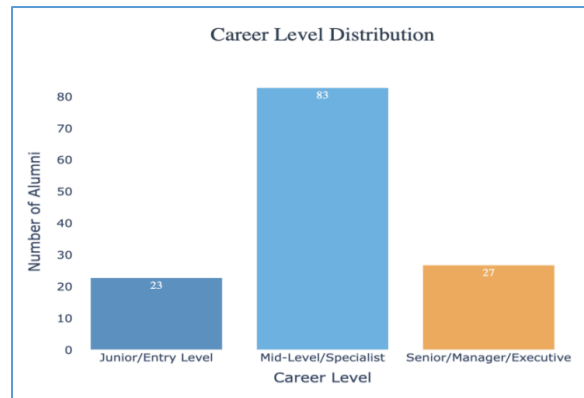
Eksplorasi dilakukan untuk mengetahui bagaimana alumni tersebar dalam kategori pekerjaan, tingkat karier, jenis industri, serta perusahaan tempat bekerja.

- Distribusi kategori pekerjaan alumni: Gambar 2 menunjukkan bahwa kategori pekerjaan business analysis & operations mendominasi, diikuti oleh data & AI/analytics, software development, dan quality assurance & testing (QA). Hal ini menunjukkan bahwa banyak alumni berkarier pada bidang yang berkaitan dengan analisis data dan pengembangan teknologi, yang mencerminkan kecenderungan distribusi karier alumni pada sektor tersebut. Sementara itu, kategori design & UX/creative juga cukup populer, namun bidang seperti IT governance & strategy dan leadership & executive memiliki jumlah alumni yang lebih sedikit. Temuan ini menunjukkan bahwa alumni cenderung bekerja pada bidang yang bersifat teknis dan analitis dibandingkan bidang tata kelola TI atau posisi kepemimpinan.



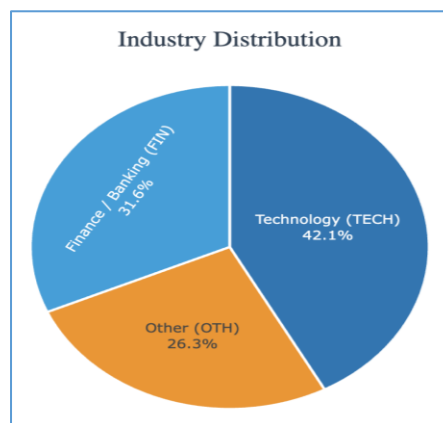
Gambar 2 Distribusi kategori pekerjaan alumni

- Distribusi tingkat karier alumni: Mayoritas alumni berada pada level mid-level/specialist dengan 83 orang (Gambar 3), yang menunjukkan bahwa sebagian besar alumni telah menempati posisi profesional pada tahap karier menengah. Alumni pada level junior/entry level berjumlah 23 orang, sementara level senior/manager/executive berjumlah 27 orang. Temuan ini menunjukkan bahwa alumni tersebar pada berbagai jenjang karier, meskipun proporsi terbesar masih berada pada level menengah.



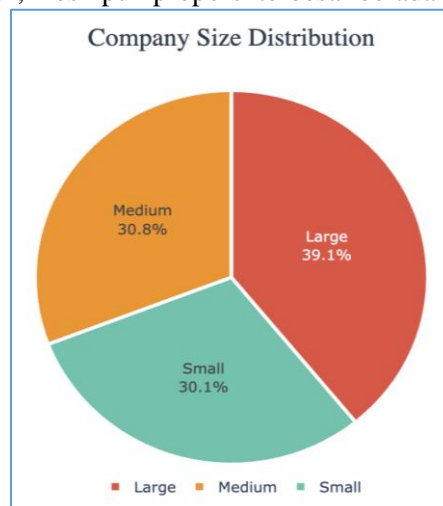
Gambar 3 Distribusi kategori tingkat karier alumni

- Distribusi industri: Gambar 4 menunjukkan bahwa industri technology (TECH) menjadi sektor terbesar dengan 42,1%, diikuti oleh sektor finance/banking (FIN) sebesar 31,6% dan Other (OTH) sebesar 26,3%. Temuan ini menunjukkan bahwa mayoritas alumni bekerja pada sektor teknologi dan keuangan, yang mencerminkan kecenderungan distribusi karier alumni pada kedua sektor tersebut.



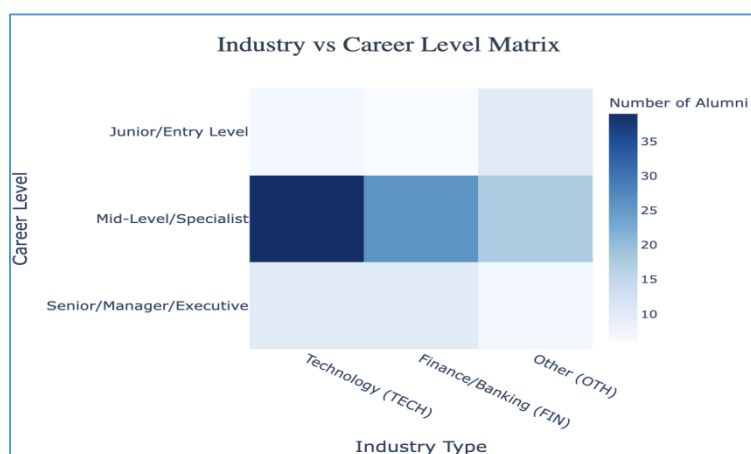
Gambar 4 Distribusi industri karier alumni

- Distribusi ukuran perusahaan (Gambar 5) menunjukkan bahwa 39,1% alumni bekerja di perusahaan besar (large), diikuti oleh perusahaan menengah (medium) sebesar 30,8% dan perusahaan kecil (small) sebesar 30,1%. Temuan ini menunjukkan bahwa alumni bekerja pada berbagai ukuran perusahaan, meskipun proporsi terbesar berada pada perusahaan besar.



Gambar 5 Distribusi ukuran perusahaan

Dari data para alumni juga diperoleh informasi, perusahaan dengan jumlah alumni terbanyak adalah Nusantara Beta Studio (5 alumni), lalu MotionPay, CIMB Niaga, BCA, Craxion Studio, Bank Indonesia, Pemprov DKI Jakarta, Huawei, Permata Bank, dan Shopee. Adapun keterkaitan industri dan karier di berbagai industri ditunjukkan Gambar 6. Dari *heatmap* ini, terlihat bahwa sebagian besar alumni berada pada level *mid-level/specialist* di sektor *technology* (TECH), dengan jumlah yang relatif besar, diikuti oleh sektor *finance/banking* (FIN). Sementara itu, untuk level *senior/manager/executive*, jumlah alumni lebih merata, tetapi tetap dominan di sektor *technology* (TECH). *Other* (OTH) memiliki kontribusi yang lebih sedikit di semua tingkatan karier. Ini menunjukkan bahwa sektor teknologi mendominasi untuk alumni dengan tingkat karier menengah dan senior.



Gambar 6 Keterkaitan industri dan tingkat karier

4.2 Data Transformation

Proses mengubah data yang sudah melewati tahap data *cleaning* menjadi bentuk representasi fitur yang dapat digunakan oleh algoritma *data mining*. Transformasi dilakukan terutama untuk mengubah atribut berbasis teks dan kategori menjadi bentuk numerik. Tahapan transformasi meliputi:

Label encoding untuk atribut kategorikal

Proses *label encoding* untuk mengonversi data kategorikal ke dalam format numerik yang dapat digunakan dalam model pembelajaran mesin. Kolom *Job_Level* dan *Industry_Type* berisi informasi kategorikal, seperti level jabatan (misalnya, '*Junior/Entry Level*', '*Mid-Level/Specialist*') dan jenis industri (misalnya, '*Technology (TECH)*', '*Finance/Banking (FIN)*'). Dengan menggunakan *LabelEncoder* dari *sklearn.preprocessing*, kategori-kategori tersebut diubah menjadi angka. Kolom *Job_Level_Encoded* berisi nilai numerik yang mewakili level jabatan, sedangkan kolom *Industry_Encoded* berisi nilai numerik yang mewakili jenis industri. Hasil transformasi menghasilkan representasi numerik dari atribut kategorikal yang dapat digunakan dalam proses clustering menggunakan algoritma K-Means.

Transformasi teks pekerjaan dengan TF-IDF

Kolom Pekerjaan Saat Ini yang berbentuk teks ditransformasikan ke dalam representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini digunakan untuk mengukur tingkat kepentingan suatu kata dalam kumpulan data job title alumni berdasarkan frekuensi kemunculannya. Melalui TF-IDF, setiap job title direpresentasikan sebagai vektor numerik yang mampu menangkap karakteristik pekerjaan secara lebih informatif dibandingkan representasi kategorikal sederhana. Hasil transformasi menghasilkan matriks fitur TF-IDF yang kemudian digabungkan dengan fitur numerik lainnya untuk proses clustering.

Penggabungan seluruh fitur menjadi dataset final

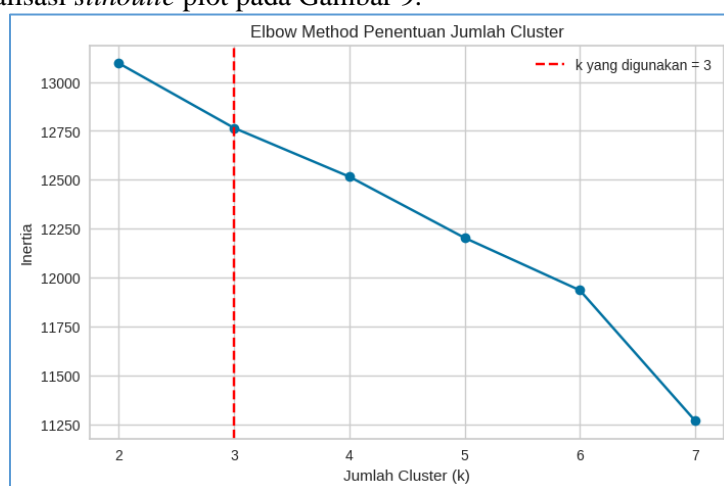
Data dibaca dari file CSV dan kolom *industry* serta *Pekerjaan Saat Ini* diproses untuk mengklasifikasikan industri dan level pekerjaan. Kolom *Job_Level* dan *Industry* yang bersifat kategorikal kemudian diubah ke format numerik menggunakan *LabelEncoder*. Selanjutnya, TF-IDF

<http://sistemasi.ftik.unisi.ac.id>

digunakan untuk mengonversi teks pada kolom Pekerjaan Saat Ini menjadi fitur numerik. Hasil encoding dan transformasi TF-IDF kemudian digabungkan menjadi matriks fitur akhir (X_{final}) yang digunakan sebagai input pada proses clustering menggunakan algoritma K-Means.

4.3 Data Mining: K-Means Clustering untuk Pola Karier Alumni

Tujuan tahap ini adalah untuk mengetahui kelompok pola karier alumni berdasarkan kesamaan beberapa faktor, yaitu level jabatan, industri tempat bekerja, dan ukuran perusahaan. Metode yang digunakan dalam analisis ini adalah *K-Means clustering*, yaitu teknik yang membagi data ke dalam sejumlah k kelompok (*cluster*) berdasarkan kedekatan setiap data terhadap pusat kelompok atau *centroid* terdekat. Dengan cara ini, pola karier alumni dapat dieksplorasi berdasarkan kemiripan karakteristik pekerjaan, industri, dan tingkat karier yang dimiliki. Hasil visualisasi penentuan jumlah *cluster* ditunjukkan pada Gambar 7. Visualisasi Elbow Method menunjukkan penurunan nilai inerti yang relatif konsisten pada rentang $k=2$ hingga $k=7$. Dengan mempertimbangkan titik perubahan laju penurunan inerti, kemudahan interpretasi hasil, serta konsistensi pembahasan, nilai $k=3$ dipilih sebagai jumlah cluster yang digunakan dalam analisis. Kemudian, hasil implementasi ditunjukkan Gambar 8 dan visualisasi *silhouette* plot pada Gambar 9.



Gambar 7 Visualisasi *elbow method* penentuan jumlah *cluster*

```

--- 2. Hasil Clustering Awal (k=3) ---
Pekerjaan Saat Ini                                Nama Perusahaan \
0 Junior Cloud Support Engineer                    Cloud Ace Integra
1 IT Trainer and consultant                        Executrain of Jakarta
2 Implementation consultant                       Data On (PT Indodev Niaga Internet)
3 Chief Technology Officer                         Craxion Studio
4 Software Engineer                               Artajasa Pembayaran Elektronik

Cluster_Karier
0      0
1      0
2      0
3      0
4      0

Jumlah data per cluster:
Cluster_Karier
0      110
1       2
2      21
Name: count, dtype: int64

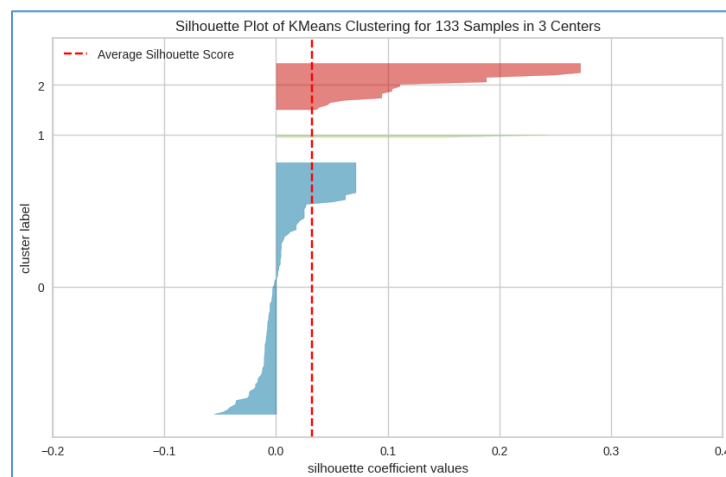
Silhouette Score (Model Awal): 0.0321
Davies-Bouldin Index (Model Awal): 3.0487

=====

--- 3. Evaluasi dengan One-Hot Encoding ---
Silhouette Score (One-Hot Encoding): -0.0214
Davies-Bouldin Index (One-Hot Encoding): 3.2625
    
```

Gambar 8 Output syntax implementasi

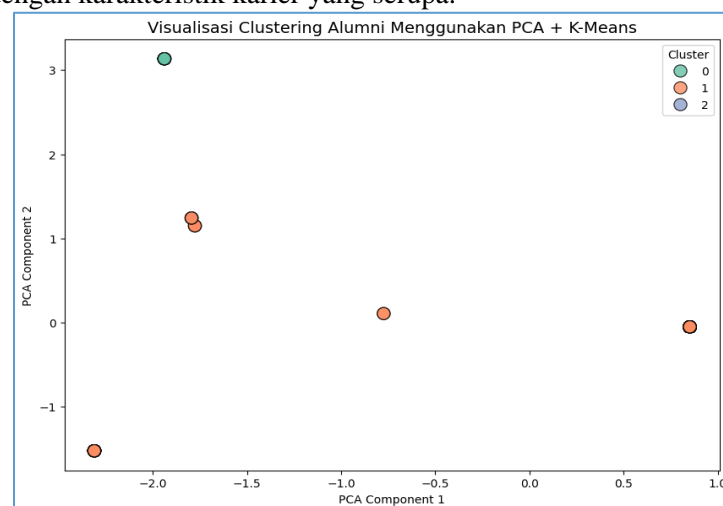
Berdasarkan hasil Elbow Method yang menunjukkan penurunan inertia secara bertahap, serta dengan mempertimbangkan kemudahan interpretasi dan konsistensi pembahasan hasil, penelitian ini menetapkan 3 cluster sebagai jumlah cluster yang digunakan dalam analisis. Hasil *clustering* awal dengan $k=3$ menunjukkan distribusi data yang tidak merata, yaitu *Cluster 0* sebanyak 110 data, *Cluster 1* sebanyak 2 data, dan *cluster 2* sebanyak 21 data. Distribusi anggota cluster yang tidak seimbang mengindikasikan bahwa sebagian besar alumni memiliki karakteristik karier yang relatif serupa, sementara hanya sebagian kecil alumni menunjukkan karakteristik yang lebih spesifik. Nilai Silhouette Score yang mendekati nol menunjukkan bahwa batas antar cluster belum terbentuk secara jelas, sedangkan nilai Davies-Bouldin Index yang relatif tinggi mengindikasikan masih terdapat kemiripan antar cluster. Kondisi ini diduga dipengaruhi oleh karakteristik data alumni yang relatif homogen karena berasal dari program studi yang sama, ukuran sampel yang terbatas, serta representasi fitur yang belum sepenuhnya mampu menangkap kompleksitas jalur karier alumni. Oleh karena itu, hasil clustering dalam penelitian ini lebih tepat dipandang sebagai pemetaan awal pola karier alumni daripada struktur cluster yang terpisah secara kuat. Pada evaluasi menggunakan One-Hot Encoding diperoleh Silhouette Score sebesar -0,0214 dan Davies-Bouldin Index sebesar 3,2625. Hasil tersebut menunjukkan performa yang lebih rendah dibandingkan model awal, sehingga pendekatan transformasi fitur yang digunakan pada model awal dipertahankan untuk analisis selanjutnya.



Gambar 9 Visualisasi *silhoutte plot*

4.4 Result Evaluation

Visualisasi hasil pengelompokan menggunakan algoritma K-Means setelah reduksi dimensi dengan PCA ditunjukkan pada Gambar 10, di mana masing-masing warna merepresentasikan satu kelompok alumni dengan karakteristik karier yang serupa.



Gambar 10 Visualisasi *clustering* alumni menggunakan PCA+K-Means

Setiap titik merepresentasikan alumni, sedangkan warna menunjukkan kelompok cluster masing-masing berdasarkan kemiripan karakteristik pekerjaan. Visualisasi ini digunakan untuk membantu interpretasi hasil clustering dalam ruang dua dimensi. Interpretasi karakteristik tiap cluster disajikan pada Tabel 2.

Tabel 2 Kelompok cluster

<i>Cluster</i>	Karakteristik Umum	Kategori Pekerjaan Dominan	Ciri Pola Karier
<i>Cluster 0</i>	Cenderung terkait proses bisnis dan koordinasi	<i>Business analyst & operations</i>	Berperan dalam komunikasi lintas divisi, perencanaan, dan pengelolaan proses bisnis.
<i>Cluster 1</i>	Cenderung terkait pengembangan sistem dan teknologi	<i>Software engineer/ developer</i>	Fokus pada implementasi teknis dan pengembangan aplikasi. Biasanya berada di perusahaan teknologi atau startup.
<i>Cluster 2</i>	Cenderung terkait pemanfaatan data & jaminan kualitas sistem	Data & AI dan QA <i>testing</i>	Terkait analisis data, pemodelan, validasi sistem, serta pengujian kualitas perangkat lunak.

Pola karier alumni Program Studi Sistem Informasi menunjukkan kecenderungan pada tiga kelompok utama, yaitu proses bisnis dan operasional, pengembangan sistem dan teknologi, serta pemanfaatan data dan jaminan kualitas sistem. Temuan ini mengindikasikan bahwa alumni berkarier pada bidang yang mencakup aspek bisnis, data, dan teknologi.

4.5 Diskusi Penelitian

Penelitian [12] menerapkan scraping dan K-Means clustering untuk karakterisasi profil LinkedIn secara umum, menemukan bahwa profil dapat dikelompokkan ke dalam kategori profesional di bidang teknologi informasi dan manajemen. Penelitian ini menunjukkan kecenderungan yang sejalan dengan temuan tersebut, namun pada konteks yang lebih spesifik, yaitu alumni Program Studi Sistem Informasi UIN Syarif Hidayatullah Jakarta, serta menghasilkan tiga kelompok pola karier yang merepresentasikan aspek bisnis, teknologi, dan data. Selain itu, [13] dalam menentukan dan mengklasifikasikan jabatan pekerjaan LinkedIn untuk analisis karier alumni menemukan bahwa sebagian besar alumni bekerja di bidang teknologi informasi dan manajemen, dengan jabatan software engineer dan project manager. Penelitian ini juga menunjukkan kecenderungan yang konsisten dengan temuan tersebut, serta melengkapinya dengan pemetaan distribusi level karier (Junior/Entry: 17,3%, Mid-Level: 62,4%, Senior/Manager: 20,3%) yang tidak tersedia dalam penelitian [13]. Hasil penelitian [14] tentang ekstraksi data profil alumni LinkedIn untuk menganalisis jalur karier menemukan bahwa alumni teknologi lebih sering bekerja di perusahaan besar, sedangkan alumni bidang lain terkonsentrasi di sektor pemerintahan atau pendidikan. Penelitian ini menambahkan normalisasi industri berbasis keyword matching untuk mengidentifikasi alumni pada sektor Government/BUMN yang tidak dilakukan secara eksplisit oleh [14]. Secara lengkap, perbandingan penelitian ini dengan penelitian terdahulu diringkas pada Tabel 3.

Tabel 3 Ringkasan perbandingan dan keunggulan penelitian

Aspek	Penelitian Terdahulu	Penelitian Saat ini	Keunggulan
Populasi	Umum/multi-disiplin	Spesifik: alumni Sistem Informasi UIN Jakarta	Analisis institusional spesifik
Metode clustering	<i>K-Means</i> umum	Method serta evaluasi Silhouette Score dan Davies-Bouldin Index	Evaluasi kualitas clustering secara kuantitatif
Industri	TI dan Manajemen	Technology (42,1%),	Distribusi kuantitatif empiris

		Finance (31,6%), Other (26.3%)	
Level karier	Tidak dianalisis	Junior 17,3%, Mid 62,4%, Senior 20,3%	Pemetaan jenjang karier
Ukuran perusahaan	Perusahaan besar vs kecil	Large 39,1%, Medium 30,8%, Small 30,1%	Pemetaan distribusi alumni berdasarkan ukuran perusahaan
Perusahaan utama	Tidak dilaporkan	Nusantara Beta Studio (5), BCA (3), Shopee (2)	Identifikasi perusahaan yang banyak menyerap alumni

5 Kesimpulan

Penelitian ini menerapkan teknik web scraping pada data LinkedIn dan algoritma K-Means dalam kerangka Knowledge Discovery in Databases (KDD) untuk memetakan pola karier alumni Program Studi Sistem Informasi UIN Syarif Hidayatullah Jakarta. Hasil penelitian menunjukkan tiga kecenderungan pola karier alumni yang berkaitan dengan aspek proses bisnis dan operasional, pengembangan sistem dan teknologi, serta pemanfaatan data dan jaminan kualitas sistem. Selain itu, mayoritas alumni bekerja pada sektor teknologi dan keuangan dengan dominasi level karier menengah (mid-level/specialist). Temuan ini menunjukkan potensi pemanfaatan data media sosial profesional sebagai sumber informasi pelengkap dalam analisis karier alumni dan tracer study berbasis data.

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Jumlah sampel yang terbatas dan cakupan yang hanya pada satu program studi mengurangi generalisasi temuan untuk populasi yang lebih luas. Nilai Silhouette Score yang rendah (0,0321) dan *Davies-Bouldin Index* yang tinggi (3,0487) menunjukkan bahwa kualitas cluster yang terbentuk masih rendah, sehingga hasil pengelompokan lebih tepat dipandang sebagai pemetaan awal pola karier alumni daripada klasifikasi yang bersifat definitif. Kondisi ini mengindikasikan bahwa fitur yang digunakan untuk clustering (jabatan, industri, dan ukuran perusahaan) belum optimal dalam memisahkan kelompok karier alumni secara jelas. Selain itu, normalisasi jabatan dan industri berbasis keyword matching masih memiliki keterbatasan dalam menangkap variasi penulisan yang kompleks. Untuk penelitian mendatang, direkomendasikan penambahan fitur seperti riwayat pekerjaan, durasi karier, dan keterampilan (*skills*) untuk meningkatkan kualitas clustering; penerapan algoritma clustering alternatif seperti Hierarchical Clustering atau DBSCAN untuk perbandingan kinerja; serta integrasi data temporal untuk menganalisis transisi dan perkembangan karier alumni dari waktu ke waktu.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih atas dukungan dana untuk penelitian dan publikasi artikel ini dari Pusat Penelitian dan Penerbitan (Puslitpen) Lembaga Penelitian dan Pengabdian kepada Masyarakat (LP2M) UIN Syarif Hidayatullah Jakarta.

Referensi

- [1] V. Y. Saki, R. Ambarsarie, D. Soemantri, H. R. Ashan, and R. Anggraini, "Profil Alumni dan Pengguna Lulusan: Analisis *Tracer Study*," *Jurnal Kesehatan Tambusai*, Vol. 4, No. 4, pp. 6829–6836, Dec. 2023, DOI: 10.31004/jkt.v4i4.22451.
- [2] R. P. Elburdah, "*Tracer Study Analysis of Accounting Program Graduates' Employment Outcomes at Universitas XXX*," *International Journal of Scientific Multidisciplinary Research*, Vol. 3, No. 7, pp. 1147–1156, Aug. 2025, DOI: 10.55927/ijsmr.v3i7.488.
- [3] M. Z. A. Islam and A. Sudrajat, "Sistem Informasi *Tracer Study* pada Politeknik TEDC Bandung berbasis *Web*," *Jurnal Informatika dan Teknik Elektro Terapan*, Vol. 13, No. 3S1, pp. 1409–1416, Oct. 2025, DOI: 10.23960/jitet.v13i3s1.7637.
- [4] H. Zou, "*From School to Work: How Alumni Networks on LinkedIn Shape PhD Journeys*," *Economic Modelling*, Vol. 155, Art. No. 107419, Dec. 2025, DOI: 10.1016/j.econmod.2025.107419.
- [5] L. Pena, M. Oliveira, and C. Curado, "*Understanding LinkedIn use and Its Relationship with Career Performance Expectations: Are there Gender Differences?*," *Journal of Organizational Effectiveness People and Performance*, pp. 1–19, Sep. 2025, DOI: 10.1108/joep-01-2025-0004.

- [6] Y. A. Mahmood and B. Mahmood, "A Web Scraper for Data Mining Purposes," *SISTEMASI*, Vol. 13, No. 3, p. 1243–1252, May 2024, DOI: 10.32520/stmsi.v13i3.4107.
- [7] J. Liu, Y. C. Ng, K. L. Wood, and K. H. Lim, "IPOD: An Industrial and Professional Occupations Dataset and Its Applications to Occupational Data Mining and Analysis," *arXiv (Cornell University)*, Oct. 2019, DOI: 10.48550/arxiv.1910.10495.
- [8] R. Ariyanto, C. Rachmad, and A. R. Syulistyo, "Automatically Collect Alumni Data on Social Media," *IOP Conference Series Materials Science and Engineering*, Vol. 732, No. 1, Art. No. 012071, Jan. 2020, DOI: 10.1088/1757-899x/732/1/012071.
- [9] F. A. Fernaldy, A. A. Arifiyanti, and D. S. Y. Kartika, "Klasterisasi Tracer Study Alumni Universitas XYZ menggunakan Algoritma K-Means," *Jurnal Informatika dan Teknik Elektro Terapan*, Vol. 13, No. 1, pp. 270–279, Jan. 2025, DOI: 10.23960/jitet.v13i1.5581.
- [10] G. B. Kaligis and S. Yulianto, "Analisa Perbandingan Algoritma K-Means, K-Medoids, dan X-Means untuk Pengelompokan Kinerja Pegawai," *IT-Explore Jurnal Penerapan Teknologi Informasi dan Komunikasi*, Vol. 1, No. 3, pp. 179–193, Oct. 2022, DOI: 10.24246/itexplore.v1i3.2022.pp179-193.
- [11] M. S. Hasibuan, A. H. Lubis, and M. N. Sari, "Perbandingan Algoritma Clustering DBSCAN dan K-Means dalam Pengelompokan Siswa Terbaik," *INFOTECH Jurnal Informatika & Teknologi*, Vol. 5, No. 2, pp. 301–309, Dec. 2024, DOI: 10.37373/infotech.v5i2.1457.
- [12] K. Dai, C. G. Nespereira, A. F. Vilas, and R. P. D. Redondo, "Scraping and Clustering Techniques for the Characterization of LinkedIn Profiles," *arXiv (Cornell University)*, May 2015, DOI: 10.5121/csit.2015.50101.
- [13] L. Li, S. Peltsverger, J., Zheng, L. Le, & M. Handlin, "Retrieving and Classifying LinkedIn Job Titles for Alumni Career Analysis," in *Proceedings of the 22st Annual Conference on Information Technology Education*, pp. 85–90, 2021, DOI: 10.1145/3450329.3476858.
- [14] S. Lade, A. Billade, A. Chandrapatle, S. Chenna, & G. Chinchalpalle, "LinkedIn Alumni Profile Data Extraction," in *Proceedings of 4th International Conference on Pervasive Computing and Social Networking (ICPCSN)*, 2024, pp. 174–178, DOI: 10.1109/ICPCSN62568.2024.00037.
- [15] P. Chaparala, A. Jukuntla, V. S. Reddy, V. V. Vinayak, & T. P. Sudha, "Extraction and Updation of Alumni Information from Web Profiles using Web Scraping," in *Proceedings of International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CHESS)*, 2023, pp. 1–7, DOI: 10.1109/iQ-CHESS56596.2023.10391404.
- [16] K. S. Kumar, P. Srihari, & C. J. Raman, "AI for Career Growth: Advanced Resume Analysis and LinkedIn Scraping for Personalized Job Recommendations," in *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 2024, pp. 1287–1293, DOI: 10.1109/ICSSAS64001.2024.10760972.
- [17] P. J. Hickey, A. Erfani, and Q. Cui, "Use of LinkedIn Data and Machine Learning to Analyze Gender Differences in Construction Career Paths," *Journal of Management in Engineering*, Vol. 38, No. 6, Aug. 2022, DOI: 10.1061/(asce)me.1943-5479.0001087.
- [18] H. Subaekti, L. Hakim, H. Khaulasari, and D. Yuliati, "An Integrated K-Means++–Davies–Bouldin Index Approach for Educational Resource-based District Clustering: A Case Study of Districts in Surabaya," *Jambura Journal of Mathematics*, Vol. 8, No. 1, pp. 111–120, Feb. 2026, DOI: 10.37905/jjom.v8i1.35412.