

Pengelompokan Perusahaan IT di Vietnam Menggunakan Algoritma K-Means dan K-Medoids

Diterima:

25 Juni 2026

Revisi:

20 Juli 2026

Terbit:

21 Juli 2026

Khalif Davianno Ramadhan

Universitas Brawijaya

Abstrak— Latar Belakang: Pesatnya perkembangan industri teknologi informasi yang didorong oleh investasi global dan kebijakan pemerintah menyebabkan meningkatnya jumlah perusahaan IT dengan karakteristik yang beragam di Negara Vietnam. Hal ini akhirnya menyebabkan pemerintah, pelaku industri dan investor kesulitan dalam memahami segmentasi industri IT secara menyeluruh. **Tujuan:** Penelitian ini bertujuan untuk mengelompokkan perusahaan IT di Vietnam menggunakan algoritma K-Means dan K-Medoids berdasarkan karakteristik perusahaan. **Metode:** Dataset yang digunakan berjumlah 2042 data meliputi jumlah karyawan, lokasi operasional dan jenis layanan. Proses penelitian meliputi tahap preprocessing data, pemodelan Clustering serta evaluasi menggunakan silhouette score. **Hasil:** Hasil penelitian menunjukkan bahwa algoritma K-Means menghasilkan performa terbaik dengan nilai silhouette score tertinggi sebesar 0,5162 pada lima Cluster lebih unggul dibandingkan K-Medoids yang memperoleh nilai 0,4879. **Kesimpulan:** Hasil pengelompokan menunjukkan dominasi perusahaan berbasis produk serta Ho Chi Minh City sebagai pusat utama industri IT di Vietnam.

Kata Kunci— Data Mining; Vietnam; K-means; Clustering; Silhouette Score

Abstract— Background: Rapid development in the information technology industry, driven by global investment and government policies, has led to an increase in the number of IT companies with diverse characteristics in Vietnam. This has ultimately caused difficulties for the government, industry players, and investors in understanding the overall segmentation of the IT industry. **Objective:** This study aims to group IT companies in Vietnam using the K-Means and K-Medoids algorithms based on company characteristics. **Methods:** The dataset used consists of 2042 data points covering the number of employees, operational location, and type of service. The research process includes data preprocessing, Clustering modeling, and evaluation using the silhouette score. **Results:** The results show that the K-Means algorithm produced the best performance with the highest silhouette score of 0.5162 for five Clusters, superior to K-Medoids, which obtained a score of 0.4879. **Conclusion:** The Clustering results show the dominance of product-based companies and Ho Chi Minh City as the main center of the IT industry in Vietnam.

Keywords— Data Mining; Vietnam; K-means; Clustering; Silhouette Score

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Khalif Davianno Ramadhan,
Sistem Informasi,
Universitas Brawijaya,
Email: khalifdavi@gmail.com

I. PENDAHULUAN

Perkembangan teknologi digital saat ini telah meningkatkan pertumbuhan ekonomi dan kemajuan teknologi di berbagai macam negara seperti di negara Vietnam. Adapun sektor teknologi informasi telah menjadi salah satu dasar dalam mendukung proses transformasi digital tersebut. Banyaknya investasi dari perusahaan teknologi global seperti Qualcomm, Renesas, Intel, Samsung, Apple sampai Amkor telah mendorong Vietnam berubah dari basis manufaktur sederhana menjadi pusat produksi layanan digital berteknologi tinggi. Hal tersebut juga oleh pemerintah Vietnam melalui kebijakan pro-investasi, pembangunan kawasan industri berteknologi tinggi seperti Saigon Hi-Tech Park dan Hoa Lac Hi-Tech Park serta adanya peningkatan kualitas pendidikan dan pelatihan tenaga kerja sehingga semakin memperkuat posisi Vietnam sebagai destinasi strategis bagi investor teknologi internasional (Cholanda, 2024).

Dengan banyaknya investasi teknologi tersebut akhirnya membuat jumlah perusahaan khususnya Perusahaan IT di Vietnam berkembang sangat cepat. Perusahaan-perusahaan ini memiliki perbedaan yang besar seperti ukuran Perusahaan, jumlah karyawan, lokasi, dan jenis layanan yang diberikan. Adapun dengan perbedaan tersebut membuat Vietnam kesulitan untuk melihat gambaran industri IT secara jelas. Hal ini dikarenakan setiap perusahaan mempunyai ciri yang berbeda sehingga sulit mengetahui mana yang memiliki kesamaan kelompok yang dimana jika tidak dikelompokkan dengan baik maka pemerintah akan kesulitan membuat kebijakan yang tepat, pelaku industri kesulitan menemukan peluang kerja sama dan investor dapat salah mengambil keputusan. Maka dari hal tersebut dibutuhkan solusi untuk mengelompokkan perusahaan IT di Vietnam untuk memberikan gambaran mengenai segmentasi perusahaan IT yang dapat digunakan sebagai dasar dalam pengambilan keputusan oleh pemerintah, pelaku industri maupun investor.

Salah satu solusi yang dapat dipergunakan adalah dengan menerapkan data mining menggunakan teknik Clustering yaitu teknik yang dipergunakan untuk mengelompokkan data ke dalam Cluster berdasarkan kemiripan (Rafi Nahjan et al., 2023). Adapun metode Clustering yang telah banyak digunakan adalah metode K-means dan K-medoids. K-means adalah algoritma Clustering yang mengelompokkan data ke dalam Cluster berdasarkan kemiripan jarak dan dikenal mampu menangani data berskala besar dengan cepat (Fitri et al., 2023). Sedangkan K-medoids yang adalah algoritma yang mengelompokkan data dengan memilih titik data nyata sebagai pusat Cluster (medoid) dan meminimalkan jarak total ke medoid yang dihitung menggunakan Euclidian Distance (Afifah et al., 2023). Dengan menerapkan metode K-means dan K-medoids maka data seperti jumlah karyawan, lokasi, dan jenis layanan dapat dikelompokkan ke dalam beberapa Cluster berdasarkan tingkat kemiripan kelompok. Pemilihan kedua metode ini juga didasarkan dengan kelebihan masing-masing algoritma disbanding algoritma lain seperti k-means yang

memiliki keunggulan yaitu proses komputasi yang cepat dan efisien dalam mengolah data berukuran besar serta hasil Clusterisasi yang mudah dipahami (Dhewayani et al., 2022). Sedangkan algoritma K-medoids memiliki keunggulan yaitu lebih tahan terhadap data outlier sehingga hasil Clusterisasi lebih stabil dan akurat dibandingkan K-Means serta dapat menangani jenis data dengan lebih fleksibel (Hermadi & Alexander, 2023)

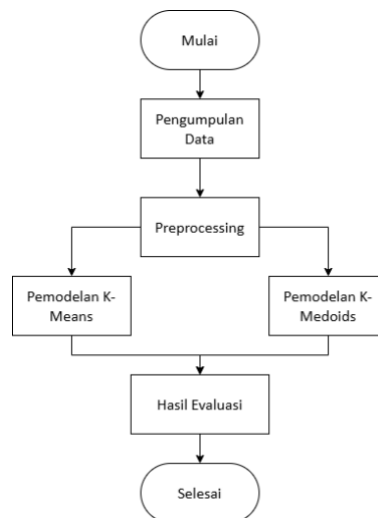
Beberapa penelitian sebelumnya telah menerapkan algoritma Clustering pada berbagai bidang seperti oleh (Effendi et al., 2021) yang menggunakan metode K-Means untuk mengelompokkan lahan kelapa sawit produktif sehingga memudahkan pemantauan kondisi lahan. Kemudian (Wisna et al., 2023) yang menerapkan K-Means untuk mengelompokkan perusahaan berdasarkan rasio likuiditas dan menghasilkan lima kelompok perusahaan dengan karakteristik likuiditas yang berbeda. Selanjutnya (Sugiono et al., 2022) yang menggunakan metode K-Medoids untuk mengelompokkan perusahaan berdasarkan data pengeluaran, jumlah personel litbang dan jumlah pekerja sehingga diperoleh empat kelompok perusahaan sebagai dasar penyusunan kebijakan. Namun penelitian-penelitian tersebut belum berfokus pada pengelompokan perusahaan teknologi informasi (IT) di Vietnam serta belum membandingkan performa algoritma K-Means dan K-Medoids pada objek penelitian tersebut. Perbedaan penelitian ini dengan penelitian sebelumnya yaitu penelitian ini berfokus pada pengelompokan perusahaan teknologi informasi (IT) di Vietnam menggunakan dua algoritma yaitu K-Means dan K-Medoids sekaligus membandingkan hasil pengelompokan untuk memperoleh metode yang menghasilkan kualitas Cluster terbaik.

Maka berdasarkan latar belakang diatas penelitian ini akan mengelompokkan perusahaan IT di Vietnam menggunakan algoritma K-means dan K-medoids. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai karakteristik kelompok perusahaan IT di Vietnam yang dapat dijadikan dasar pengambilan Keputusan oleh pemerintah, pelaku industri, dan investor.

II. METODE

Penelitian ini dilakukan melalui tahapan KDD yaitu proses untuk menggali informasi berharga dari kumpulan data besar dengan menggunakan teknik statistik pembelajaran mesin dan analisis data. Adapun tahapan dalam KDD meliputi pengumpulan data, preprocessing data, transformasi data, data mining atau pemodelan, dan evaluasi (Alijoyo, 2024).

Berikut adalah tahapannya:



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan alur penelitian yang digunakan dalam penelitian ini. Tahapan penelitian diawali dengan proses pengumpulan data melalui situs Kaggle. Selanjutnya dilakukan preprocessing data sebelum diterapkan algoritma K-Means dan K-Medoids untuk proses pengelompokan. Tahap terakhir adalah evaluasi hasil *Clustering* untuk membandingkan kinerja kedua algoritma. Berikut detail proses alur penelitian ini:

2.1 Pengumpulan Data

Pada tahap ini peneliti melakukan pengumpulan dataset Kaggle yang berisi informasi perusahaan dan lowongan pekerjaan IT di Vietnam. Dataset tersebut diperoleh melalui proses crawling dari situs itviec.com berjumlah 2042 data. Data yang dikumpulkan memiliki atribut seperti `company_id`, `company_name`, `average_rating`, `num_review`, `city`, `type`, `num_employee`, `country`, `working_day`, `OT`, `overview`, `expertise`, `benefit`, `logo_link`.

2.2 Preprocessing

Tahap *preprocessing* adalah proses untuk membersihkan data yang akan digunakan dalam proses analisis. Pada tahapan ini akan dilakukan beberapa kegiatan seperti data cleaning untuk menghapus data yang tidak lengkap, duplikat, atau tidak sesuai. Selanjutnya data transformation yang meliputi perubahan format data menjadi nilai numerik sehingga algoritma pengelompokan dapat berjalan dengan baik. Terakhir dilakukan reduksi data dengan memilih atribut-atribut paling sesuai seperti jumlah karyawan, lokasi operasional dan jenis layanan.

2.3 Pemodelan

Pada tahap pemodelan peneliti menerapkan metode pengelompokan menggunakan dua algoritma yaitu K-Means dan K-Medoids. Adapun kedua algoritma ini akan digunakan untuk mengelompokkan perusahaan IT di Vietnam berdasarkan karakteristik yang telah dipilih selama tahap preprocessing.

2.4 Evaluasi

Pada tahap evaluasi peneliti memecahnya menjadi dua tahap yaitu evaluasi model dan analisis SWOT. Pertama peneliti melakukan pengujian terhadap hasil pengelompokan yang diperoleh dari algoritma K-Means dan K-Medoids. Evaluasi ini dilakukan dengan menggunakan Silhouette Score untuk mengukur kualitas *Cluster* yang terbentuk. Selain itu hasil *Cluster* juga akan dianalisis secara deskriptif untuk memberikan informasi yang bermanfaat sesuai dengan karakteristik perusahaan IT di Vietnam berdasarkan hasil pengelompokan. Terakhir peneliti melakukan analisis SWOT untuk memberikan gambaran kondisi perusahaan serta pemetaan secara keseluruhan.

III. HASIL DAN PEMBAHASAN

3.1 Evaluasi Model

Tahap *preprocessing* adalah tahap untuk melakukan pembersihan data yang masih mentah menjadi siap untuk diolah dalam pemodelan. Pada penelitian akan dilakukan beberapa proses yaitu pemeriksaan data kosong dan data duplikat, encoding pada data kategori serta normalisasi data. Adapun dari 13 atribut yang terdapat yaitu *company_id*, *company_name*, *average_rating*, *num_review*, *city*, *type*, *num_employee*, *country*, *working_day*, *OT*, *overview*, *expertise*, *benefit*, dan *logo_link* maka peneliti hanya akan menggunakan tiga atribut yaitu *company_name*, *num_employee*, *city*, dan *type*. Atribut lain seperti *company_id* dihilangkan karena merupakan identitas perusahaan dan sudah diwakilkan oleh pemakaian atribut *company_name*. Atribut seperti *average_rating*, *num_review*, *overview*, *expertise*, *benefit*, dan *logo_link* juga tidak digunakan karena lebih bersifat deskriptif atau teks yang tidak sesuai dengan tujuan penelitian untuk pengelompokan karakteristik perusahaan. Sementara itu atribut, *num_employee*, *city*, dan *type* dipilih karena paling representative untuk menggambarkan profil perusahaan IT di Vietnam yaitu dari segi skala perusahaan (*num_employee*), lokasi geografis (*city*), dan jenis operasional perusahaan (*type*) yang secara bersama-sama dapat mencerminkan karakteristik perusahaan.

Pada tahap pemodelan akan dilakukan proses penerapan algoritma k-means dan k-medoids pada data yang sebelumnya telah melalui proses preprocessing. Adapun kedua algoritma memiliki prinsip pembagian data yang sama yaitu membagi data ke dalam beberapa kelompok berdasarkan kesamaan karakteristik. Maka untuk menentukan jumlah *Cluster* yang paling baik penelitian ini

akan menggunakan evaluasi metode Silhouette Score untuk mengukur seberapa baik *Cluster* yang terbentuk. Adapun semakin tinggi nilai Silhouette Score mendekati 1 maka semakin baik hasil pengelompokan yang diperoleh.

Proses pemodelan pertama yang akan dilakukan adalah penerapan Algoritma K-Means. Algoritma k-means bekerja dengan cara membagi data ke dalam k *Cluster* berdasarkan jarak terdekat terhadap centroid atau titik pusat *Cluster*. Adapun nilai centroid dihitung berdasarkan nilai rata-rata dari setiap data dalam *Cluster*. Pada penelitian ini akan dilakukan pengujian nilai k terbaik berdasarkan Silhouette Score dengan range 2-5. Kode program dalam proses pemodelan yaitu sebagai berikut:

Tabel 1. Kode Program Pemodelan K-means

No	Kode Program Identifikasi Data Kosong
1	# === 3. Evaluasi Silhouette Score (K-Means) ===
2	from sklearn.metrics import silhouette_score
3	import matplotlib.pyplot as plt
4	from sklearn.preprocessing import StandardScaler
5	
6	results_kmeans = []
7	
8	for k in range(2, 6):
9	kmeans = KMeans(n_Clusters=k, random_state=42)
10	labels = kmeans.fit_predict(X_scaled)
11	sil_score = silhouette_score(X_scaled, labels)
12	results_kmeans.append((k, sil_score))
13	print(f"k = {k} Silhouette Score = {sil_score:.4f}")
14	
15	# === hasil keseluruhan ===
16	print("\n=== HASIL EVALUASI K-MEANS ===")
17	print("K\tSilhouette Score")
18	for k, score in results_kmeans:
19	print(f"{k}\t{score:.4f}")
20	
21	# === Visualisasi ===
22	k_values = [r[0] for r in results_kmeans]
23	si_values = [r[1] for r in results_kmeans]
24	
25	plt.figure(figsize=(6,4))

No	Kode Program Identifikasi Data Kosong
26	<code>plt.plot(k_values, si_values, marker='o', color='blue')</code>
27	<code>plt.title('Evaluasi K-Means Berdasarkan Silhouette Score')</code>
28	<code>plt.xlabel('Jumlah Cluster (k)')</code>
29	<code>plt.ylabel('Silhouette Score')</code>
30	<code>plt.grid(True)</code>
31	<code>plt.show()</code>
32	
33	<code># === Evaluasi ===</code>
34	<code>best_k = max(results_kmeans, key=lambda x: x[1])</code>
35	<code>print(f"\nK terbaik untuk K-Means: k = {best_k[0]} dengan SI = {best_k[1]:.4f}")</code>

Pada Tabel 1 menunjukkan kode program penerapan algoritma K-Means pada penelitian ini. Adapun jumlah *Cluster* yang paling optimal berdasarkan nilai Silhouette Score. Peneliti melakukan iterasi dengan menentukan nilai k dari 2 hingga 5. Untuk setiap nilai k maka algoritma K-Means dijalankan dengan menggunakan data yang telah ditransformasi sebelumnya dimana fungsi `fit_predict()` digunakan untuk mengelompokkan data dan memberikan label *Cluster* pada setiap objek. Selanjutnya nilai Silhouette Score dihitung menggunakan fungsi `silhouette_score(X_scaled, labels)` untuk mengukur seberapa baik data dalam satu *Cluster* saling berdekatan dan seberapa jauh jaraknya dengan *Cluster* lain.

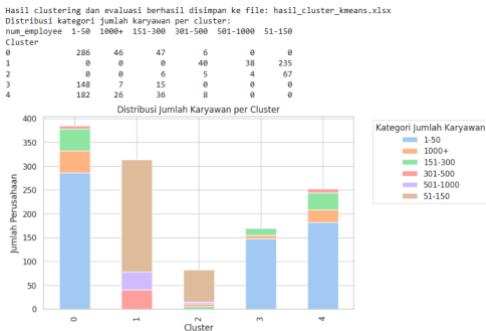
Pada tahapan evaluasi akan dilakukan untuk penentuan model *Clusterisasi* terbaik antara K-Means dan K-Medoids dengan membandingkan nilai Silhouette Score pada hasil pengujian nilai k yang telah dilakukan sebelumnya. Adapun nilai Silhouette Score mengukur seberapa baik data terkelompok berdasarkan jarak antar objek dalam satu *Cluster* dan antar *Cluster* lain. Cara kerja dalam evaluasi ini yaitu jika nilai Silhouette Score semakin mendekati angka 1 maka hasil *Clusterisasi* semakin baik karena data dalam satu *Cluster* lebih homogen dan berbeda jauh dari *Cluster* lainnya. Adapun Silhouette Score dipilih sebagai metrik evaluasi dikarenakan mampu memberikan penilaian akurat terhadap kualitas hasil *Clusterisasi*. Metode ini dapat menunjukkan seberapa baik suatu data berada di dalam *Cluster* yang tepat dengan menilai tingkat kekompakan antar data dalam satu *Cluster* serta jarak pemisahan dengan *Cluster* lainnya. Setelah itu beberapa penelitian sebelumnya seperti oleh (Guntur et al., 2025) juga menggunakan Silhouette Score untuk menentukan jumlah *Cluster* optimal karena metrik ini terbukti efektif, mudah diinterpretasikan, dan dapat menggambarkan kualitas *Cluster*.

Berikut adalah perbandingan hasil evaluasi yang dihasilkan:

Tabel 2. Perbandingan Silhouette Score

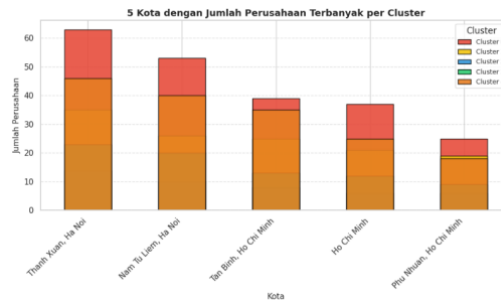
Jumlah <i>Cluster</i> (K)	Silhouette Score K-Means	Silhouette Score K Medoids
2	0.3497	0.4385
3	0.4404	0.4404
4	0.4656	0.4788
5	0.5162	0.4879

Berdasarkan tabel 2 maka dapat dilihat bahwa nilai Silhouette Score untuk pemodelan dengan menggunakan algoritma K-Means memiliki nilai evaluasi tertinggi sebesar 0.5162 pada k=5. Hal ini menunjukkan bahwa pemodelan K-Means mampu menghasilkan pemisahan *Cluster* yang lebih baik dan lebih stabil dibandingkan dengan K-Medoids. Hasil analisa *Cluster* terbaik yaitu *Cluster* 5 menggunakan algoritma K-Means yaitu sebagai berikut:



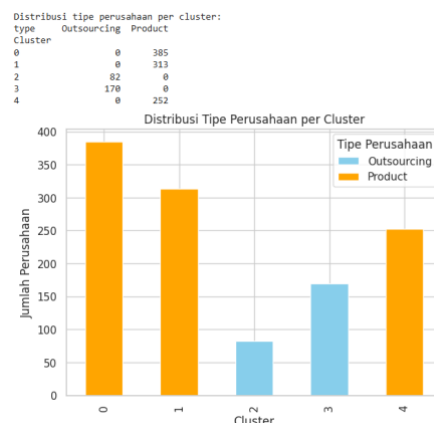
Gambar 2. Distribusi Kategori Jumlah Karyawan Per *Cluster*

Berdasarkan Gambar 2 menunjukkan hasil *Clustering* k=5 menggunakan K-Means untuk pengelompokan perusahaan IT Vietnam berdasarkan ukuran tenaga kerja. Hasil distribusi karyawan menunjukkan bahwa *Cluster* 0, *Cluster* 3, dan *Cluster* 4 didominasi oleh perusahaan skala mikro dengan 1-50 karyawan walaupun terdapat juga perusahaan dengan jumlah karyawan diatas 1000. *Cluster* 1 merupakan kelompok perusahaan menengah-besar dengan mayoritas memiliki 51-150 karyawan sehingga menunjukkan perusahaan sedang bertumbuh. *Cluster* 2 adalah kelompok dengan distribusi karyawan merata yang menunjukkan bahwa perusahaan dalam fase transisi menuju skala lebih besar.



Gambar 3. Distribusi 5 Kota Terbanyak Per *Cluster*

Berdasarkan Gambar 3 menunjukkan distribusi geografis perusahaan IT berdasarkan *Cluster* yang terbentuk dimana wilayah dengan perusahaan IT terbanyak di setiap *Cluster* yaitu di wilayah Ho Chi Minh dengan tiga area yaitu Tan Binh, pusat kota Ho Chi Minh dan Phu Nhuan. Pola ini menunjukkan bahwa wilayah kota Ho Chi Minh menjadi pusat utama industri IT di Vietnam dikarenakan memiliki infrastruktur yang lebih maju, akses terhadap tenaga kerja terampil serta ekosistem bisnis yang sudah terbentuk dengan baik. Ketiga faktor tersebut membuat banyak perusahaan IT memilih untuk berlokasi dan berkembang di wilayah ini.



Gambar 4. Distribusi Tipe Perusahaan Per Cluster

Berdasarkan Gambar 8 menunjukkan distribusi tipe perusahaan yang memberikan informasi kelompok jenis perusahaan yang sangat jelas di setiap *Cluster*. Adapun *Cluster* 0, 1, dan 4 terdiri dari perusahaan product dengan jumlah masing-masing 385, 313 dan 252 perusahaan. Selanjutnya *Cluster* 2 dan 3 didominasi oleh tipe perusahaan outsourcing dengan jumlah masing-masing 82 dan 170 perusahaan. Hasil ini menunjukkan bahwa perusahaan IT bertipe product cenderung memiliki distribusi yang lebih banyak dibandingkan perusahaan IT bertipe outsourcing.

3.2 Analisis SWOT

Dalam pengambilan keputusan investasi yang tepat maka diperlukan analisis yang mampu menggambarkan kondisi perusahaan secara komprehensif. Maka dalam penelitian ini dilakukan

pengelompokan perusahaan teknologi informasi di Vietnam untuk memperoleh Gambaran karakteristik Perusahaan yang berpotensi sebagai tujuan investasi.

Berdasarkan hasil pengelompokan perusahaan IT yang telah dilakukan maka informasi tersebut dapat dimanfaatkan sebagai dasar dalam analisis strategis untuk mengevaluasi kekuatan, kelemahan, peluang dan ancaman pada masing-masing klaster perusahaan menggunakan analisis SWOT. Hal ini nantinya dapat membantu investor atau pemerintah untuk memperoleh gambaran kondisi perusahaan juga pemetaan sektor IT di Vietnam secara keseluruhan. Analisis ini bersifat deskriptif dan bertujuan untuk memberikan gambaran kondisi yang ada sehingga dapat menjadi dasar pertimbangan bagi investor sebelum mengambil keputusan investasi di Negara Vietnam.

Analisis SWOT sendiri adalah metode analisis strategis untuk mengevaluasi kondisi perusahaan atau fenomena berdasarkan faktor internal dan eksternal. Faktor internal meliputi kekuatan dan kelemahan sedangkan faktor eksternal meliputi peluang dan ancaman (Jannah et al., 2024). Berikut adalah 4 faktor analisis SWOT (Rasta et al., 2024).

- a. *Strength* atau kekuatan adalah kondisi yang menunjukkan keunggulan yang dimiliki oleh organisasi atau Perusahaan seperti kemampuan teknologi dan mutu yang lebih baik dibandingkan pihak lain.
- b. *Weakness* atau kelemahan adalah kondisi internal yang menjadi keterbatasan atau kekurangan perusahaan. Faktor ini dapat menghambat perkembangan usaha dan perlu dianalisis untuk mengetahui penyebabnya.
- c. *Opportunities* atau peluang adalah faktor eksternal yang bersifat menguntungkan dan berpotensi mendorong perkembangan perusahaan di masa mendatang.
- d. *Threat* atau ancaman adalah kondisi eksternal yang berpotensi menimbulkan risiko atau tekanan bagi perusahaan. Jika tidak diantisipasi dengan baik maka ancaman tersebut dapat mengganggu usaha serta berdampak negatif baik dalam jangka pendek maupun jangka panjang.

Matriks SWOT disusun dengan membandingkan faktor internal yaitu *Strengths* dan *Weaknesses* dengan faktor eksternal yaitu *Opportunities* dan *Threats* (Simanjuntak & Nadya, 2026). Hasil matriks ini akan memberikan informasi bagi investor dalam mengambil Keputusan bisnis serta meningkatkan daya saing.

Tabel 3. Matriks Strategi SO dan WO

No	Faktor Internal
Memanfaatkan kemampuan inovasi perusahaan product untuk memenuhi permintaan pasar IT yang saat ini meningkat.	Membantu perusahaan skala mikro meningkatkan modal dan SDM untuk dapat memanfaatkan peluang pasar yang luas

No	Faktor Internal
Pemerintah membantu potensi pertumbuhan perusahaan menengah/transisi dengan dukungan untuk ekspansi dan transformasi digital.	Mengurangi ketergantungan perusahaan outsourcing pada klien dengan perbedaan layanan atau pasar
Memanfaatkan peluang konsentrasi perusahaan di Ho Chi Minh City untuk membangun kolaborasi antar perusahaan.	Menyediakan pelatihan dan pengembangan teknologi bagi perusahaan kecil untuk mampu bersaing di pasar global.

Tabel 4. Matriks Strategi ST dan WT

Strategi ST (<i>Strength</i> + Threat)	Strategi WT (<i>Weakness</i> + Threat)
Memanfaatkan fleksibilitas dan kemampuan inovasi perusahaan product untuk menghadapi persaingan yang ketat.	Meningkatkan kapasitas SDM dan modal perusahaan mikro untuk dapat bertahan menghadapi ancaman pasar dan teknologi.
Menggunakan konsentrasi lokasi di Ho Chi Minh City untuk memperkuat kolaborasi menghadapi persaingan.	Diversifikasi pasar dan produk bagi perusahaan outsourcing sehingga risiko ketergantungan klien berkurang.
Memanfaatkan peluang konsentrasi perusahaan di Ho Chi Minh City untuk membangun kolaborasi antar perusahaan.	Menyediakan pelatihan dan pengembangan teknologi bagi perusahaan kecil untuk mampu bersaing di pasar global.

Temuan penelitian ini adalah hasil pengelompokan menggunakan algoritma K-Means dan K-Medoids dapat menjadi dasar dalam memetakan faktor *Strength*, *Weakness*, *Opportunity* dan *Threat* sehingga menghasilkan rekomendasi strategi yang sesuai dengan karakteristik masing-masing klaster perusahaan. Hasil penelitian ini sejalan dengan penelitian oleh (Al-Fahmi et al., 2023) yang menyatakan bahwa hasil *Clustering* dapat dipetakan ke dalam analisis SWOT sebagai dasar perumusan strategi bisnis serta penelitian tentang pengelompokan perusahaan IT di Vietnam yang menunjukkan bahwa hasil klaster dapat dimanfaatkan untuk mengevaluasi kekuatan, kelemahan, peluang dan ancaman sebagai dasar pengambilan keputusan strategis.

IV. KESIMPULAN

Berdasarkan penelitian yang dilakukan mengenai pengelompokan perusahaan IT di Vietnam menggunakan algoritma K-Means dan K-Medoids maka diambil kesimpulan bahwa penelitian ini berhasil menerapkan algoritma K-Means dan K-Medoids untuk melakukan pengelompokan perusahaan IT di Vietnam. Hasil evaluasi menunjukkan bahwa metode K-Means menghasilkan Cluster terbaik dengan nilai silhouette score tertinggi sebesar 0,5162 pada jumlah 5 Cluster sedangkan K-Medoids pada jumlah Cluster yang sama hanya menghasilkan skor 0,4879. Hal ini menunjukkan bahwa metode K-Means lebih unggul daripada metode K-Medoids pada pengelompokan data IT Jobs. Hasil pengelompokan menunjukkan karakteristik perusahaan

IT di Vietnam berdasarkan jenis perusahaan dan lokasi, serta mengidentifikasi Ho Chi Minh City sebagai pusat industri teknologi. Temuan penelitian ini adalah hasil pengelompokan menggunakan algoritma K-Means dapat menjadi dasar dalam memetakan faktor Strength, Weakness, Opportunity, dan Threat (SWOT) untuk mendukung penyusunan strategi. Keterbatasan penelitian ini adalah penggunaan atribut dan metrik evaluasi yang masih terbatas. Oleh karena itu penelitian selanjutnya disarankan menambahkan atribut perusahaan yang lebih beragam serta menggunakan metrik evaluasi lain seperti Davies–Bouldin Index untuk meningkatkan kualitas hasil pengelompokan..

DAFTAR PUSTAKA

- Afifah, A. N., Nurcahyawati, V., & Hananto, V. R. (2023). Analisis Clustering dan Pemetaan Sebaran Pelanggan Perusahaan Properti di Sidoarjo. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 9(3), 502. <https://doi.org/10.26418/jp.v9i3.67935>
- Al-Fahmi, B. M., Rahmawati, E., & Sagirani, T. (2023). Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(2), 141–149. <https://doi.org/10.25077/TEKNOSI.v9i2.2023.141-149>
- Alijoyo, F. A. (2024). Optimasi Pemilihan Metode Pengajaran Dosen Menggunakan Data Mining, dan Algoritma K-Means dalam Proses Bisnis Pendidikan. *Informatics and Digital Expert (INDEX)*, 6(2), 105–109. <https://doi.org/10.36423/index.v6i2.1883>
- Cholanda, Y. (2024). *Masa Depan Bisnis Teknologi Asia Tenggara di Genggaman Vietnam*. Tirto.Id. <https://tirto.id/masa-depan-bisnis-teknologi-asia-tenggara-di-genggaman-vietnam-g6Ca>
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, 12(1), 64–77. <https://doi.org/10.34010/jati.v12i1.6674>
- Effendi, H., Syahril, A., Prayoga, S., & Hidayat, W. D. (2021). Penerapan Metode K-Means Clustering untuk Pengelompokan Lahan Sawit Produktif pada PT Kasih Agro Mandiri APPLICATION OF K-MEANS CLUSTERING METHOD FOR GROUPING PRODUCTIVE PALM OIL LAND AT PT KASIH AGRO MANDIRI. *Teknomatika*, 11(02), 1–5. <https://doi.org/10.57152/ijirse.v3i1.488>
- Fitri, E. M., Suryono, R. R., & Wantoro, A. (2023). Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz. *Jurnal Komputasi*, 11(2), 157–168. <https://doi.org/10.23960/komputasi.v11i2.12582>
- Guntur, J. M., Kurniawan, R., & Hayati, U. (2025). *Algoritma K-Means untuk Meningkatkan Silhouette Score pada Pengelompokan Data Stok Bahan Manufaktur di PT . XYZ*. 11–21.
- Hermadi, A. O., & Alexander, A. D. (2023). Implementasi Algoritma K-Medoids Clustering Untuk Mencari Keuntungan Sementara Dalam Laporan Keuangan. *Jurnal Teknologi Dan Ilmu Komputer Prima (Jutikomp)*, 6(1), 6–11.
- Jannah, M., Indraputri, A. J., & Puspita, V. E. (2024). *Pentingnya Analisis Swot dalam Suatu Perencanaan dan Pengembangan Bisnis*. 2(1), 9–17.
- Rafi Nahjan, M., Nono Heryana, & Apriade Voutama. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101–104. <https://doi.org/10.36040/jati.v7i1.6094>

- Rasta, J., Pohan, T. D., & Nurbaiti. (2024). Analisis Swot Dalam Strategi Pemasaran Pada Mcdonald ' s Juwairiazizah Rasta Tia Deja Pohan. *Jurnal Rimba : Riset Ilmu Manajemen Bisnis Dan Akuntansi*, 2(1).
- Simanjuntak, E. F., & Nadya, Y. (2026). Perencanaan Strategi Pemasaran Menggunakan Analisis Strengths Weaknesses Opportunities Threat) Dan Quantitative Strategic Planning Matrix. *JENIUS: Jurnal Terapan Teknik Industri*, 7(1), 176–189. <https://doi.org/10.37373/jenius.v7i1.1227>
- Sugiono, S., Wahanani, N. A., & Hendriani, L. J. (2022). Pengelompokan Perusahaan Berdasarkan Statistik Penelitian dan Pengembangan Melalui Metode K-Medoids. *JOINS (Journal of Information System)*, 7(1), 41–52. <https://doi.org/10.33633/joins.v7i1.5935>
- Wisna, N., Nanda, A., & Fahrudin, T. (2023). Klasterisasi Perusahaan Sub Kontraktor Berdasarkan Rasio Likuiditas Menggunakan K-Means Clustering. *Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA)*, 7(2), 139–150. <https://doi.org/10.31955/mea.v7i2.2994>