

BAB II TINJAUAN PUSTAKA

A. Penelitian Terkait

Penelitian terkait adalah serangkaian sumber penelitian sebelumnya yang dikumpulkan dan digabungkan untuk mendukung penelitian, memberikan acuan kepada penulis, dan mencegah hasil penelitian yang sama. Berikut penelitian terkait yang digunakan sebagai referensi dalam penelitian ini ditunjukkan pada tabel berikut:

Tabel 1. Penelitian Terkait

No.	Peneliti	Keterangan Penelitian
1	Tuti Hartati, Odi Nurdiawan, Eko Wiyandi (2021)	Judul: Analisis Dan Penerapan Algoritma <i>K-Means</i> Dalam Strategi Promosi Kampus Akademi Maritim Suaka Bahari Metode: <i>Knowledge Discovery in Databases</i> (KDD) dan metode <i>K-Means Clustering</i> . Software: RapidMiner Hasil Penelitian: Temuan dalam penelitian ini menunjukkan bahwa penerapan metode <i>K-Means clustering</i> dapat digunakan untuk merumuskan strategi promosi yang efektif, seperti pemasaran langsung (<i>direct marketing</i>), pemanfaatan media sosial (<i>social network</i>), dan periklanan (<i>advertising</i>). Berdasarkan pengukuran performa algoritma K-Means dengan parameter jumlah cluster, diperoleh hasil bahwa <i>Cluster</i> ke-16 memiliki 115 anggota.
2	Dwi Fitri Miliani, Wawan Joko Pranoto (2021)	Judul: Penerapan <i>K-Means Cluster</i> Dalam Memilih Strategi Promosi Penerimaan Mahasiswa Baru Metode: <i>K-Means Clustering</i> . Software: Google Colab Hasil Penelitian: Hasil penelitian membuktikan bahwa metode <i>K-Means Clustering</i> efektif dalam merumuskan strategi promosi yang dapat diterapkan pada periode berikutnya dengan media online dan media cetak sebagai metode promosi tahun 2022/2023.

No.	Peneliti	Keterangan Penelitian
3	Neng Sri Lathifah Zulfa, Ana Hadiana (2021)	Judul: Kajian Data Mining Menggunakan <i>Algoritma K-Means</i> Dan <i>K-Medoids</i> Dalam Strategi Promosi Metode: <i>Algoritma K-Means Dan K-Medoids</i>. Software: Microsoft Excel dan Rapidminer Hasil Penelitian: <i>Algoritma K-Means</i> yang lebih menjadi baik dibandingkan dengan <i>K-Medoids</i>. berdasarkan dari nilai validasi dunn Index paling tinggi pada <i>K-Means</i> K=5 sebesar 0,999 dan dalam proses pengolahannya <i>K-Means</i> hanya membutuhkan waktu rata-rata 1 detik, sedangkan <i>algoritma K-Medoids</i> membutuhkan waktu rata rata ± 1 menit yang artinya apabila jumlah pengelompokkan ditentukan lebih besar, yang proses pengolahan data pun semakin lama.
4	Banu Harli Trimulya Suandi As, Lisna Zahrotun (2021)	Judul: Penerapan Data Mining dalam Mengelompokkan Data Riwayat Akademik Sebelum Kuliah dan Data Kelulusan Mahasiswa Menggunakan Metode Agglomerative Hierarchical Clustering Metode: <i>Agglomerative Hierarchical Clustering (AHC)</i> Software: Microsoft Excel dan Aplikasi Olah Data Hasil Penelitian: Menghasilkan 4 cluster terbaik dari masing-masing program studi (Informatika, Kimia, Industri, dan Elektro). Data rekomendasi berjumlah 158 mahasiswa, mayoritas berasal dari pulau Jawa dengan rata-rata nilai Matematika ≥ 80 (untuk Informatika dan Elektro) dan IPK ≥ 3.50

5	Anastasia L Maukar, Fitri Marisa, Anang Aris Widodo, Nurin kamilaningtyas, dan Novian Didik Nugraha (2022)	Judul: Analisis Data Penerimaan Mahasiswa Baru Berbasis <i>K-Means</i> Metode: <i>K-Means Clustering</i> Software: RapidMiner Hasil Penelitian: Hasil penelitian menunjukkan tiga <i>cluster</i> berdasarkan peminat jurusan di Universitas Widyagama Malang. <i>Cluster</i> pertama didominasi oleh peminat jurusan ekonomi, <i>cluster</i> kedua didominasi oleh peminat jurusan teknik, dan <i>cluster</i> ketiga didominasi oleh peminat jurusan hukum. Selain itu, terdapat pengetahuan bahwa minat terhadap jurusan belum merata dan asal sekolah peminat masing-masing jurusan juga belum merata. Peneliti menyarankan agar kampus melakukan promosi yang lebih mendetail terhadap jurusan yang ada untuk meningkatkan peminat, terutama di jurusan teknik dan dari madrasah.
6	Agneresa, April Lia Hananto, Shofa Shofiah Hilabi, Agustia Hananto, Tukino (2022)	Judul: Strategi Promosi Penerapan Data Mining Mahasiswa Baru dengan Metode <i>K-Means Clustering</i> Metode: <i>Knowledge Discovery in Databases</i> (KDD) dan metode <i>K-Means Clustering</i>. Software: RapidMiner Hasil Penelitian: Strategi promosi untuk penerimaan mahasiswa baru dapat disesuaikan dengan hasil pengelompokan (<i>cluster</i>) yang terbentuk berdasarkan asal sekolah dan program studi yang paling diminati di masing-masing sekolah, dengan menggunakan metode <i>K-Means Clustering</i>. Dari total 2.479 data, terbentuk dua <i>cluster</i>: <i>Cluster</i> 1 terdiri dari 1.954 data (78,82%) yang dikategorikan sebagai <i>cluster</i> dengan potensi tinggi, dan <i>Cluster</i> 2 terdiri dari 525 data (21,18%) yang termasuk dalam <i>cluster</i> dengan potensi rendah.

7	Dwina Pri Indini, Sartika R. Siburian, Nurhasanah, Dito Putro Utomo, Mesran (2022)	Judul: Implementasi Algoritma DBSCAN untuk Clustering Seleksi Penentuan Mahasiswa yang Berhak Menerima Beasiswa Yayasan Metode: <i>Density-Based Spatial Clustering of Applications with Noise (DBSCAN)</i> Software: Microsoft Excel dan Aplikasi Olah Data Hasil Penelitian: DBSCAN dianggap efektif karena mampu membentuk cluster dengan kerapatan berbeda dan menghasilkan 2 cluster dari 10 data mahasiswa angkatan 2019 berdasarkan kriteria: IPK rendah-sedang (1–6), Prestasi Non-Akademik rendah, Jumlah tanggungan orang tua tinggi, Penghasilan orang tua rendah
8	Herlina, Siti Mundari, Elsa Amanda Putri, Amanda Dian Rahmawati (2024)	Judul: Penerapan <i>K-Means Clustering</i> Untuk Strategi Promosi Untuk Meningkatkan Jumlah Perolehan Mahasiswa Baru Metode: <i>K-Means Clustering</i>. Software: Rapidminer Hasil Penelitian: Penelitian ini membuktikan bahwa <i>K-Means Clustering</i> efektif dalam mengolah data dan merumuskan strategi promosi yang didukung oleh pelayanan terbaik.
9	Devita Maulina Putri, Asri Samsiar Ilmananda, Nadita Prisanta (2024)	Judul: Penggunaan Algoritma <i>K-Means</i> dan <i>K-Medoids</i> untuk Pengembangan Strategi Promosi Penerimaan Mahasiswa Baru Metode: <i>K-Means</i> dan <i>K-Medoids</i>. Software: Microsoft Excel dan Rapidminer Hasil Penelitian: Algoritma <i>K-Means</i> dianggap lebih unggul dan direkomendasikan untuk digunakan dalam merumuskan strategi promosi penerimaan mahasiswa baru di perguruan tinggi. Untuk menargetkan promosi secara lebih efisien kepada segmen mahasiswa tertentu, dengan memusatkan upaya promosi pada wilayah-wilayah yang dinilai memiliki potensi lebih tinggi.

10	Muhammad Hanif Abdurrohman, Elin Haerani, Fadhilah Syafria, Lola Oktavia (2024)	Judul: Pendaftaran Mahasiswa Baru (Studi Kasus Universitas Abdurrab) Metode: <i>K-Means Clustering</i> . Software: Perangkat Lunak Statistik atau Pemrograman seperti Python Hasil Penelitian: Penelitian menunjukkan efektivitas <i>K-Means Clustering</i> dalam mengidentifikasi pola yang signifikan dalam data pendaftaran mahasiswa baru.
11	Christopher Kevin Herijanto, Yosefina Finsesia Riti (2024)	Judul: Perbandingan Algoritme <i>K-Means Clustering</i> dan <i>Fuzzy C-Means</i> untuk Penentuan Strategi Promosi Universitas Katolik Darma Cendika Metode: <i>Algoritme K-Means Clustering dan Fuzzy C-Means</i> Software: <i>Google Colab dengan Google Compute Engine dan Python v3.</i> Hasil Penelitian: Berdasarkan pengujian nilai Silhouette Coefficient, hasil dari <i>Algoritme K-Means</i> (0,2833) lebih unggul untuk dijadikan strategi promosi untuk Universitas Katolik Darma Cendika dalam mendapatkan hasil yang lebih valid dibandingkan dengan <i>Algoritme Fuzzy C-Means</i> (-0.1456). Oleh karena itu, strategi promosi yang berdasarkan hasil <i>K-Means</i> memberikan segmentasi target yang lebih jelas dan valid. Penggunaan <i>K-Means</i> dapat memberikan keuntungan dalam memahami karakteristik mahasiswa dan menyusun strategi promosi yang lebih terarah.

Penelitian ini memiliki perbedaan dibandingkan dengan studi-studi sebelumnya. Jika sebagian besar penelitian terdahulu lebih menekankan pada analisis deskriptif data mahasiswa, maka studi ini secara khusus menggunakan metode *K-Means clustering* untuk mengelompokkan calon mahasiswa berdasarkan wilayah asal mereka.

Pendekatan ini memungkinkan pengenalan segmen yang lebih terarah dan sesuai, serta menghubungkan hasil *clustering* dengan strategi bauran promosi yang lebih optimal. Oleh karena itu, penelitian ini tidak hanya memperkaya pemahaman dalam hal pengelompokan data, tetapi juga menawarkan rekomendasi praktis yang dapat diterapkan oleh UNUGHA dalam upaya menarik calon mahasiswa baru.

B. Landasan Teori

1. Clustering

Clustering merupakan metode *unsupervised learning* yang digunakan untuk mengelompokkan data berdasarkan kemiripan atribut yang dimiliki [12]. Berbeda dengan klasifikasi yang menggunakan variabel target, *clustering* tidak membutuhkan variabel tersebut dan digunakan sebagai langkah awal dalam analisis data [13].

2. Metode K-Means Clustering

Metode K-Means Clustering merupakan metode partisi yang digunakan untuk membagi data ke dalam beberapa kelompok yang terpisah satu sama lain [14]. Menurut K. Arai dan A. R. Barakbah, *algoritma K-Means* merupakan salah satu metode *clustering* yang paling sederhana dan banyak digunakan. Keunggulannya terletak pada kemampuannya dalam mengelompokkan data berukuran besar dengan waktu komputasi yang relatif cepat dan efisien. Namun demikian, algoritma ini memiliki kelemahan, terutama pada tahap penentuan pusat *cluster* awal yang dapat memengaruhi hasil akhir pengelompokan [15].

3. Strategi Promosi

Menurut Kotler, “*promotion, the fourth marketing mix tools, stand for various activities, the company undertakes to communicate its products merits and to persuade target customers to buy them*”. Definisi tersebut mempunyai pengertian bahwa promosi meliputi semua alat yang terdapat dalam bauran promosi yang peran utamanya adalah mengadakan komunikasi yang bersifat membujuk [16]. Menurut Kotler & Armstrong (dalam Chasanah [17]) variabel - variabel yang ada di dalam *promotional mix* ada lima, yaitu:

- a. Periklanan (*Advertising*).
- b. Penjualan Personal (*Personal Selling*).
- c. Promosi penjualan (*Sales Promotion*).
- d. Hubungan masyarakat (*Public Relation*).
- e. Pemasaran langsung (*Direct Marketing*).

4. PMB

Penerimaan mahasiswa baru (PMB) adalah tahapan penting yang menentukan kualitas dan masa depan institusi. Dimana, proses ini bukan sekadar pada formalitas administratif yang biasanya dilakukan, melainkan sebuah investasi yang bersifat jangka panjang yang akan berdampak pada berbagai aspek[18].

Proses ini berlangsung secara berulang setiap semester di perguruan tinggi. Dengan mengolah data tersebut, dapat diperoleh berbagai informasi yang bermanfaat untuk merumuskan strategi promosi penerimaan mahasiswa baru pada periode selanjutnya[19].

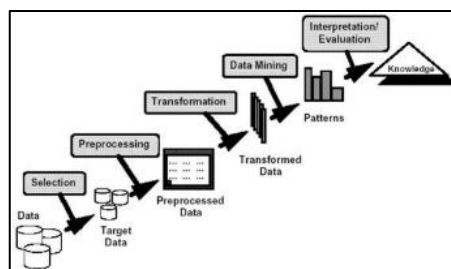
Secara umum, promosi penerimaan mahasiswa baru di perguruan tinggi dilakukan melalui berbagai saluran media, seperti media elektronik, cetak, serta media sosial seperti TikTok, Instagram, dan Facebook. Namun, dalam praktiknya, kegiatan promosi ini seringkali tidak mempertimbangkan data yang telah tersedia sebelumnya, seperti data demografis pendaftar, padahal informasi tersebut merupakan unsur penting dalam merancang strategi promosi yang tepat. [20].

5. *Google Colab*

Google Colab merupakan platform *cloud* gratis yang memungkinkan pengguna menulis dan menjalankan kode *Python* langsung dari browser menggunakan *library* populer seperti *TensorFlow*, *Pytorch*, *Scikit-learn*, dan *Numpy* [21].

6. *Knowledge Discovery in Database (KDD)*

Knowledge Discovery in Database (KDD) adalah proses analisis data yang digunakan dalam *data mining* dengan tujuan mengungkap serta mengenali pola-pola dalam basis data untuk menghasilkan informasi baru.



Gambar 2. Tahapan KDD

Tahapan KDD meliputi [22]:

a. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalan informasi dalam KDD dimulai. Tidak semua data dalam database digunakan, sehingga hanya data yang relevan dan sesuai untuk dianalisis yang akan diambil.

b. *Data Pre- processing / Cleaning*

Sebelum tahap data mining dilakukan, diperlukan proses pembersihan terhadap data yang menjadi fokus dalam KDD. Tahap ini meliputi penghapusan data yang duplikat serta pemeriksaan terhadap data yang tidak konsisten.

c. *Data Transformation*

Data perlu diubah atau disesuaikan ke dalam format yang kompatibel untuk proses data mining. Beberapa metode *data mining* membutuhkan format data sebelum dapat digunakan. Misalnya, metode seperti analisis *clustering* hanya dapat memproses data dalam bentuk kategorikal, sehingga data numerik yang bersifat kontinu harus dipecah menjadi beberapa interval terlebih dahulu.

d. *Data mining*

Data mining merupakan proses untuk menemukan pola atau informasi yang relevan dari data yang telah dipilih, dengan memanfaatkan berbagai teknik atau metode tertentu. Ragam metode dan algoritma dalam data mining sangat beragam, dan pemilihannya harus disesuaikan dengan tujuan serta keseluruhan tahapan dalam proses KDD.

e. *Interpretation / Evaluation*

Hasil pola informasi yang diperoleh dari proses data mining perlu disajikan dalam format yang mudah dipahami oleh pihak terkait. Pada tahap ini juga dilakukan evaluasi untuk memastikan apakah temuan tersebut sejalan atau bertentangan dengan fakta maupun hipotesis yang telah ada sebelumnya.

7. Pengujian *Silhouette Score*

Silhouette Score, atau yang dikenal juga sebagai *silhouette coefficient*, adalah metode yang digunakan untuk menilai kualitas dan kekompakan suatu *cluster*. Evaluasi ini mengukur seberapa efektif *algoritma clustering* dalam memisahkan data ke dalam kelompok-kelompok berbeda, dengan menilai sejauh mana setiap titik data sesuai dengan *cluster* tempatnya berada dibandingkan dengan *cluster* lainnya.

Nilai *Silhouette Coefficient (SC)* berada dalam rentang -1 hingga 1, di mana nilai positif menunjukkan bahwa data berada dalam *cluster* yang sesuai, sedangkan nilai negatif mengindikasikan bahwa data tersebut kurang tepat berada dalam *cluster* tersebut [23].

Berikut ini adalah interpretasi nilai *Silhouette Coefficient*:

- a. *High SC value* (mendekati angka 1):
Sampel terpisah dengan baik dengan *cluster* lainnya atau dengan kata lain, pengelompokan sampel sudah sangat sesuai.
- b. *SC value* (sekitar angka 0):
Sampel terkluster masih belum tepat atau dikatakan masih belum signifikan dikarenakan jarak antar sampel dengan *cluster* sekitarnya.
- c. *Low SC value* (mendekati angka -1):
Sampel terkluster dengan tidak tepat atau sampel ditempatkan pada *cluster* yang salah.

Jika nilai *silhouette coefficient* semakin menghampiri 1 maka objek sudah terletak pada klaster yang benar, sehingga jumlah klaster akan optimal. Kriteria pengukuran pengelompokan menurut *Kauffman* dan *Roesseeuw* menggunakan metode *silhouette coefficient* dapat dilihat pada Tabel 2 [24].

Tabel 2. Kriteria *Silhouette Score*

Nilai SC	Kriteria
0,71-1,00	Struktur Kuat
0,51-0,70	Struktur Baik
0,26-0,50	Struktur Lemah
$\leq 0,25$	Struktur Buruk

8. Google Maps

Google Maps merupakan salah satu produk dari google yang dapat diakses untuk memberikan informasi-informasi tempat dan penunjuk jalan kepada penggunanya[25]. Google Maps menyediakan layanan web mapping dan data geospasial seluruh dunia dengan adanya fasilitas layanan API (Application Programming Interface), yang dapat dimanfaatkan untuk mendapatkan peta dari lokasi tertentu tanpa perlu membayar dan menggunakan fitur yang disediakan termasuk untuk memvisualisasikan data yang berupa koordinat geografis[26]. Dalam konteks pengukuran jarak, Google Maps menggunakan metode yang membedakan antara jarak *Euclidean* dan jarak jalan. Jarak jalan dihitung berdasarkan rute yang tersedia dan kondisi lalu lintas, memberikan estimasi yang lebih akurat untuk perjalanan nyata.