BULLET : Jurnal Multidisiplin Ilmu
Volume 2, No. 01, Febuari - Maret 2023
ISSN 2829-2049 (media online)
Hal 271-276

Segmentasi Minat Mahasiswa Terhadap Program Studi Menggunakan Algoritma *K-Means Clustering*

Edi Wahyudin[1], Fatihanursari Dikananda[2*]

[1]Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon, Kota Cirebon, Indonesia
[2]Program Studi Teknik Informatika, STMIK IKMI Cirebon, Kota Cirebon, Indonesia
Email: [1]Ediwahyudin.ikmi@gmail.com, [2*]Fatihanursari.ikmi@gmail.com
(* : coressponding author)

Abstrak − Segmentasi minat mahasiswa terhadap program studi merupakan langkah penting dalam pengambilan keputusan strategis di bidang pendidikan tinggi. Dengan mengetahui kelompok minat yang terbentuk, institusi dapat menyusun kurikulum yang relevan serta strategi pemasaran yang lebih efektif. Penelitian ini bertujuan untuk mengelompokkan minat mahasiswa terhadap program studi menggunakan algoritma *K-Means Clustering*. Data yang digunakan berasal dari kuesioner minat dan preferensi mahasiswa baru terhadap beberapa program studi. Hasil dari penerapan algoritma *K-Means* menunjukkan bahwa mahasiswa dapat dikelompokkan ke dalam beberapa klaster berdasarkan kesamaan minat mereka, yang dapat digunakan untuk mendukung kebijakan akademik dan promosi program studi.

Kata Kunci: Segmentasi, Minat Mahasiswa, *K-Means Clustering*, Program Studi, Data Mining

Abstract *− Segmenting student interest in study programs is a crucial step in strategic decision-making within higher education. By identifying interest groups, institutions can design more relevant curricula and develop more effective marketing strategies. This study aims to cluster student interests in study programs using the K-Means Clustering algorithm. The data used in this research were obtained from questionnaires assessing the interests and preferences of new students towards various study programs. The results of applying the K-Means algorithm indicate that students can be grouped into several clusters based on the similarity of their interests, which can be utilized to support academic policy and program promotion strategies.*

Keywords*: Segmentation, Student Interest, K-Means Clustering, Study Program, Data Mining*

1. PENDAHULUAN

Dalam pendidikan tinggi, memahami preferensi dan minat mahasiswa sangat penting untuk perencanaan akademik dan pengembangan institusi. Banyaknya pilihan program studi yang tersedia seringkali membingungkan calon mahasiswa dalam menentukan keputusan yang sesuai dengan minat dan bakat mereka. Oleh karena itu, diperlukan metode analisis yang dapat membantu mengelompokkan mahasiswa berdasarkan kesamaan minat terhadap program studi.

Teknologi analisis data berupa data mining mempermudah institusi pendidikan dalam memahami pola minat mahasiswa. Salah satu metode efektif dalam pengelompokan data adalah algoritma *K-Means Clustering*, yaitu teknik pembelajaran tak terawasi yang banyak digunakan untuk segmentasi (Meftah et al., 2022). Algoritma ini memungkinkan data dikelompokkan ke dalam beberapa klaster berdasarkan kemiripan karakteristik, sehingga dapat diidentifikasi pola-pola tersembunyi dalam data mahasiswa yang sebelumnya sulit terlihat secara manual.

Penelitian ini bertujuan untuk mengelompokkan minat mahasiswa terhadap program studi menggunakan algoritma *K-Means*. Segmentasi yang dihasilkan dapat membantu institusi dalam menyusun strategi yang lebih tepat sasaran dalam menawarkan program studi yang diminati calon mahasiswa. Selain itu, hasil segmentasi ini juga dapat digunakan untuk merancang kegiatan promosi, bimbingan karier, dan pengembangan kurikulum yang sesuai dengan kecenderungan minat mahasiswa di masing-masing klaster.

Dengan adanya pendekatan ini, institusi tidak hanya mampu meningkatkan efektivitas dalam menarik minat calon mahasiswa, tetapi juga dapat meningkatkan kepuasan dan keberhasilan akademik mahasiswa dengan menyelaraskan program studi yang ditawarkan dengan potensi dan minat yang dimiliki. Di era transformasi digital saat ini, pemanfaatan data untuk pengambilan keputusan strategis menjadi aspek krusial dalam pengelolaan pendidikan tinggi yang adaptif dan responsif terhadap kebutuhan mahasiswa (Putri et al., 2024).

BULLET : Jurnal Multidisiplin Ilmu
Volume 2, No. 01, Febuari - Maret 2023
ISSN 2829-2049 (media online)
Hal 271-276

2. METODOLOGI PENELITIAN

2.1. Sumber Data

Data Data diperoleh dari kuesioner yang disebarkan kepada mahasiswa baru pada awal tahun ajaran. Kuesioner mencakup pertanyaan mengenai minat terhadap program studi yang ditawarkan, alasan memilih program, pengaruh orang tua, prospek karir, dan ketertarikan pribadi. Total responden sebanyak 300 mahasiswa dari berbagai fakultas yang mewakili spektrum disiplin ilmu, seperti sains dan teknologi, sosial dan humaniora, serta ekonomi dan desain.

Untuk memastikan validitas data, kuesioner dirancang menggunakan skala Likert 1–5 untuk mengukur intensitas minat dan preferensi mahasiswa terhadap masing-masing aspek. Selain itu, kuesioner juga telah melalui proses uji coba dan validasi oleh pakar pendidikan sebelum penyebaran secara luas, guna menjamin bahwa setiap item benar-benar merepresentasikan konstruk yang diukur.

Pengumpulan data dilakukan selama dua minggu melalui platform survei daring, dengan tetap menjaga anonimitas dan kerahasiaan responden. Kriteria inklusi yang digunakan adalah mahasiswa baru tahun akademik berjalan yang telah resmi terdaftar di institusi pendidikan terkait. Data yang terkumpul kemudian dikompilasi dalam bentuk spreadsheet dan diproses menggunakan perangkat lunak pengolahan data statistik sebelum masuk ke tahap pra-pemrosesan dan analisis klaster.

Dengan pendekatan ini, data yang diperoleh diharapkan mampu memberikan gambaran yang komprehensif tentang pola minat mahasiswa baru terhadap program studi yang tersedia, serta faktor-faktor yang mempengaruhinya. Informasi ini menjadi dasar penting dalam proses segmentasi menggunakan algoritma *K-Means Clustering*.

2.2. Pra-Pemrosesan Data

Sebelum dilakukan pengelompokan, data melalui tahapan pra-pemrosesan berikut:

1. **Pembersihan Data**

   Proses ini dilakukan untuk menghapus entri duplikat yang dapat menyebabkan bias pada hasil pengelompokan. Selain itu, data yang tidak valid—seperti nilai yang berada di luar rentang wajar atau bertentangan secara logis—juga diperbaiki. Penanganan data kosong dilakukan dengan dua pendekatan, yaitu penghapusan entri yang tidak signifikan jumlahnya, serta imputasi (pengisian) menggunakan metode rata-rata atau modus tergantung pada tipe data (numerik atau kategorik).

2. **Transformasi Data**

   Sebagian besar data berasal dari jawaban kualitatif seperti "sangat berminat", "tidak berminat", atau "netral". Untuk dapat dianalisis oleh algoritma *K-Means* yang bekerja dengan data numerik, data kualitatif tersebut diubah menjadi bentuk numerik menggunakan *label encoding* (Kusumadewi et al., 2023). Pada kasus tertentu, seperti variabel dengan banyak kategori, *one-hot encoding* juga dipertimbangkan untuk menghindari asumsi hubungan ordinal yang tidak sesuai.

3. **Normalisasi**

   Seluruh fitur data dinormalisasi menggunakan metode *Min-Max Scaling*, yang mengubah nilai fitur ke dalam rentang 0 hingga 1. Normalisasi ini penting untuk menghindari dominasi fitur dengan skala besar terhadap proses pembentukan klaster. Dengan skala yang seragam, algoritma *K-Means* dapat menghitung jarak antar data secara lebih adil dan akurat.

4. **Pemeriksaan Outlier**

   Outlier atau nilai pencilan dapat mengganggu proses klastering karena dapat menarik pusat klaster secara tidak proporsional. Oleh karena itu, dilakukan identifikasi dan analisis outlier menggunakan boxplot dan metode z-score. Jika ditemukan data yang ekstrem dan tidak

BULLET : Jurnal Multidisiplin Ilmu
Volume 2, No. 01, Febuari - Maret 2023
ISSN 2829-2049 (media online)
Hal 271-276

representatif, maka data tersebut dipertimbangkan untuk dikeluarkan dari proses pelatihan model.

Dengan menjalani tahapan pra-pemrosesan ini, data yang digunakan menjadi lebih bersih, konsisten, dan sesuai untuk dianalisis menggunakan algoritma *K-Means*, sehingga hasil segmentasi yang diperoleh dapat lebih akurat dan bermanfaat secara praktis.

2.3. Algoritma K-Means

Algoritma *K-Means* adalah metode pembelajaran tak terawasi yang digunakan untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan kedekatan antar titik data. Algoritma ini bekerja dengan cara meminimalkan jarak antara data dan pusat klaster (centroid), sehingga setiap data dikelompokkan dengan data lain yang memiliki karakteristik serupa.

Langkah-langkah dalam algoritma *K-Means* adalah sebagai berikut:

1. **Menentukan Jumlah Klaster (k)**

   Sebelum menerapkan algoritma, jumlah klaster (k) harus ditentukan terlebih dahulu. Pemilihan jumlah klaster yang optimal sangat penting agar hasil segmentasi sesuai dengan tujuan analisis. Beberapa metode yang dapat digunakan untuk menentukan k yang optimal adalah Metode *Elbow*, *Silhouette Score*, atau metode Gap Statistic. Dalam penelitian ini, jumlah klaster dipilih menggunakan Metode *Elbow*, yang akan dijelaskan lebih lanjut pada bagian berikutnya.

2. **Menginisialisasi Centroid Secara Acak**

   Setelah menentukan jumlah klaster, langkah selanjutnya adalah menginisialisasi posisi centroid secara acak. Centroid adalah titik pusat dari setiap klaster yang mewakili rata-rata posisi semua titik data dalam klaster tersebut. Penginisialisasian centroid dilakukan secara acak pada ruang data yang ada, yang kemudian menjadi titik awal dalam proses klastering. Pemilihan centroid yang tepat sangat mempengaruhi hasil akhir dari pengelompokan, sehingga berbagai metode seperti *K-Means* digunakan untuk meningkatkan kualitas inisialisasi centroid.

3. **Mengelompokkan Titik Data Berdasarkan Jarak Euclidean ke Centroid**

   Setelah centroid diinisialisasi, setiap titik data dikelompokkan berdasarkan kedekatannya dengan centroid. Jarak antar titik data dan centroid dihitung menggunakan rumus jarak Euclidean, yang memberikan ukuran jarak geometris antara dua titik dalam ruang n-dimensi. Titik data akan dimasukkan ke dalam klaster yang memiliki jarak terkecil dengan centroid-nya. Proses ini berlangsung secara iteratif, di mana setiap titik data akan selalu bergabung dengan klaster terdekat.

4. **Menghitung Ulang Posisi Centroid Berdasarkan Anggota Klaster Baru**

   Setelah semua titik data dikelompokkan, langkah selanjutnya adalah menghitung ulang posisi centroid untuk masing-masing klaster. Posisi centroid yang baru ini merupakan rata-rata dari semua titik data yang terdapat dalam klaster tersebut. Dengan kata lain, centroid akan bergerak menuju titik tengah klaster berdasarkan anggota-anggota baru yang telah dimasukkan ke dalam klaster.

5. **Mengulangi Langkah 3 dan 4 hingga Posisi Centroid Konvergen**

   Langkah-langkah pengelompokan dan perhitungan ulang posisi centroid akan diulang secara iteratif hingga posisi centroid tidak berubah lagi atau perubahan yang terjadi sangat kecil (konvergen). Pada titik ini, klaster dianggap stabil, dan proses klastering selesai. Konvergensi ini terjadi ketika tidak ada perubahan signifikan dalam anggota klaster atau posisi centroid pada iterasi berikutnya. Hal ini menandakan bahwa algoritma telah menemukan pembagian klaster yang optimal.

Dalam implementasi *K-Means*, ada beberapa pertimbangan yang perlu diperhatikan, seperti pemilihan jumlah klaster yang optimal dan inisialisasi centroid yang tepat, untuk memastikan bahwa algoritma bekerja dengan baik dan menghasilkan klaster yang representatif (Fahmi & Nurhayati, 2021).

BULLET : Jurnal Multidisiplin Ilmu
Volume 2, No. 01, Febuari - Maret 2023
ISSN 2829-2049 (media online)
Hal 271-276

2.4. Penentuan Nilai K Optimal

Metode Penentuan jumlah klaster yang optimal (k) merupakan langkah penting dalam algoritma K-Means. Jika jumlah klaster terlalu sedikit, maka hasil klastering mungkin tidak mewakili keragaman data dengan baik. Sebaliknya, jika jumlah klaster terlalu banyak, maka klaster yang terbentuk bisa terlalu spesifik dan tidak memberikan wawasan yang berguna. Salah satu metode yang paling umum digunakan untuk menentukan nilai k yang optimal adalah Metode *Elbow*.

Metode *Elbow* dilakukan dengan cara mengamati grafik yang menunjukkan hubungan antara jumlah klaster (k) dan *Sum of Squared Error* (*SSE*), yang merupakan jumlah dari kuadrat jarak antara setiap titik data dan centroid klasternya. Semakin banyak jumlah klaster yang digunakan, SSE cenderung menurun karena data akan lebih terkelompok dengan lebih baik. Namun, penurunan SSE ini tidak akan terus berlanjut dengan cara yang signifikan setelah titik tertentu.

Langkah-langkah dalam menggunakan Metode *Elbow* adalah sebagai berikut:

a. Menghitung SSE untuk Berbagai Nilai k

Lakukan pengelompokan data untuk berbagai nilai k, misalnya mulai dari k = 1 hingga k = 10, dan hitung nilai SSE untuk masing-masing k. SSE dihitung dengan rumus:

$$SSE = \sum_{i=1}^{n} \sum_{k=1}^{K} \|x_i - c_k\|^2$$

Gambar 2. 1 Rumus SSE

Di mana xix adalah titik data, ckc_kck adalah centroid klaster, dan KKK adalah jumlah klaster.

b. Plot Grafik SSE terhadap Jumlah Klaster (k)

Setelah nilai SSE dihitung untuk masing-masing k, buat grafik yang memplotkan nilai k (sumbu x) terhadap nilai SSE (sumbu y).

c. Identifikasi Titik *Elbow*

Amati grafik dan cari titik di mana penurunan *SSE* mulai melambat atau mencapai suatu titik yang hampir horizontal. Titik ini disebut sebagai *Elbow* (siku), dan ini merupakan jumlah klaster optimal (k) yang memberikan keseimbangan terbaik antara pengurangan *SSE* dan kompleksitas model.

d. Validasi Hasil

Meskipun Metode *Elbow* dapat memberikan gambaran yang jelas mengenai nilai k, terkadang keputusan subjektif diperlukan untuk memilih titik *Elbow* yang paling sesuai dengan tujuan analisis. Beberapa teknik lain seperti *Silhouette Score* juga dapat digunakan untuk memvalidasi hasil yang diperoleh dari Metode *Elbow* (Rahman et al., 2020).

Metode *Elbow* ini sangat berguna dalam menentukan jumlah klaster yang optimal, namun perlu diingat bahwa tidak selalu ada titik *Elbow* yang jelas. Dalam hal ini, metode tambahan atau evaluasi manual berdasarkan konteks masalah dapat diperlukan untuk menentukan nilai k yang paling relevan.

2.5. Evaluasi Klaster

Untuk mengevaluasi kualitas pengelompokan, digunakan metrik *Silhouette Score*, yang mengukur sejauh mana setiap titik data cocok dalam klasternya dibandingkan dengan klaster lain (Putri et al., 2024).

BULLET : Jurnal Multidisiplin Ilmu
Volume 2, No. 01, Febuari - Maret 2023
ISSN 2829-2049 (media online)
Hal 271-276

3. ANALISA DAN PEMBAHASAN

3.1. Penentuan K Optimal

Metode *Elbow* yang digunakan dalam penelitian ini menunjukkan bahwa jumlah klaster optimal adalah 3, yang mengindikasikan adanya tiga kelompok utama minat mahasiswa terhadap program studi. Titik *Elbow* terlihat jelas pada grafik *SSE* ketika nilai k = 3, di mana penurunan nilai *SSE* setelah titik ini tidak signifikan lagi. Hal ini menunjukkan bahwa pembagian menjadi tiga klaster sudah cukup mewakili variasi dalam data minat mahasiswa. Hasil ini juga diperkuat dengan evaluasi menggunakan *Silhouette Score* yang menunjukkan hasil clustering yang memadai.

3.2. Hasil Pengelompokan

Berdasarkan hasil penerapan algoritma K-Means, data responden berhasil dikelompokkan ke dalam tiga klaster utama, yaitu:

1. **Klaster 1 (45% responden)**

   Kelompok mahasiswa yang sangat berminat pada program-program sains dan teknologi, seperti Teknik Informatika,Teknik Elektro,Sistem Informasi. Faktor utama yang mempengaruhi minat mereka meliputi Prospek karir di bidang teknologi informasi dan rekayasa.Perkembangan pesat dunia teknologi dan digitalisasi,Daya tarik gaji tinggi dan stabilitas pekerjaan

2. **Klaster 2 (30% responden)** Kelompok mahasiswa yang tertarik pada bidang ilmu sosial dan humaniora, seperti Psikologi, Ilmu Hukum, Ilmu Komunikasi.Faktor-faktor penentu minat di klaster ini antara lain Ketertarikan pribadi terhadap interaksi sosial dan dinamika manusia, Pengaruh keluarga, teman, dan lingkungan, Cita-cita berkontribusi di bidang sosial atau pelayanan masyarakat

3. **Klaster 3 (25% responden)** Kelompok mahasiswa dengan minat beragam dan multidisipliner, terutama pada Manajemen, Bisnis, Desain Komunikasi Visual Mahasiswa dalam klaster ini menunjukkan karakteristik Ketertarikan pada kreativitas, inovasi, dan fleksibilitas profesi, Aspirasi untuk menjadi wirausahawan atau bekerja di industri kreatif, Motivasi untuk membangun bisnis sendiri atau bekerja secara independen

3.3. Evaluasi Klaster

Nilai Evaluasi kualitas hasil pengelompokan dilakukan menggunakan *Silhouette Score*, yaitu metrik yang mengukur seberapa baik suatu data cocok dengan klaster tempat ia berada dibandingkan dengan klaster lain. Pada penelitian ini, diperoleh nilai *Silhouette Score* sebesar 0.63, yang tergolong dalam kategori cukup baik.

Nilai tersebut menunjukkan bahwa:

1) Sebagian besar data memiliki kohesi internal yang kuat, artinya data dalam satu klaster memiliki kemiripan yang tinggi.

2) Pemisahan antar klaster juga cukup baik, yang berarti jarak antara satu klaster dengan klaster lainnya relatif jelas dan tidak tumpang tindih secara signifikan.

Selain itu, pengamatan visual terhadap hasil klastering melalui plot dua dimensi terhadap atribut-atribut utama (seperti minat utama dan alasan pemilihan program studi) memperlihatkan pola pengelompokan yang terdistribusi secara wajar dan tidak terjadi dominasi satu klaster terhadap keseluruhan data.

Evaluasi ini memperkuat kesimpulan bahwa hasil pengelompokan menggunakan algoritma *K-Means* pada data minat mahasiswa cukup representatif dan dapat digunakan sebagai dasar segmentasi untuk keperluan kebijakan institusional.

3.4. Implikasi Temuan

Hasil penelitian ini menyarankan bahwa institusi dapat memanfaatkan segmentasi ini untuk:

1) Menyusun kurikulum yang sesuai dengan kelompok minat mahasiswa.
2) Mengembangkan strategi pemasaran program studi yang lebih tertarget.
3) Menyediakan layanan bimbingan dan konseling yang disesuaikan.

4. KESIMPULAN

Penelitian ini berhasil melakukan segmentasi minat mahasiswa terhadap program studi menggunakan algoritma *K-Means Clustering*. Hasil pengelompokan menunjukkan tiga kelompok minat utama yang dapat dijadikan dasar dalam pengambilan kebijakan akademik dan promosi program studi yang lebih tepat sasaran. Evaluasi menggunakan *Silhouette Score* menunjukkan bahwa segmentasi yang dihasilkan memiliki kualitas yang baik. Penelitian ini membuktikan bahwa pendekatan data mining dapat diterapkan secara efektif dalam dunia pendidikan tinggi untuk memahami karakteristik mahasiswa dengan lebih baik.

Penelitian selanjutnya disarankan untuk mengeksplorasi teknik clustering lainnya seperti *DBSCAN* dan *HCA* (*Hierarchical Clustering Analysis*), serta memperluas dataset agar hasil yang diperoleh lebih komprehensif dan dapat digeneralisasi.

REFERENCES

Nurhayati, N., & Sari, D. M. (2020). Analisis Segmentasi Mahasiswa Berdasarkan Preferensi Menggunakan Algoritma K-Means Clustering. Jurnal Teknologi dan Sistem Komputer, 8(2), 95–102. https://doi.org/10.14710/jtsiskom.8.2.2020.95-102

Prasetyo, E., & Widodo, S. (2020). Penerapan Algoritma K-Means untuk Pengelompokan Data Akademik Mahasiswa. Jurnal Ilmu Komputer dan Aplikasi, 15(1), 45–53. https://doi.org/10.31294/jika.v15i1.7645

Putra, R. A., & Pratama, R. D. (2021). Implementasi K-Means Clustering untuk Segmentasi Data Mahasiswa Berdasarkan Minat Program Studi. Jurnal Informatika dan Komputer Indonesia (JIKOM), 6(1), 45–52. https://doi.org/10.32672/jikom.v6i1.2515

Hidayat, M. T., & Suryana, R. (2021). Data Mining untuk Pengelompokan Minat Mahasiswa Menggunakan K-Means. Jurnal Teknologi Informasi dan Ilmu Komputer, 8(2), 177–183. https://doi.org/10.25126/jtiik.202182526

Ayu, F. D., & Firmansyah, A. (2022). Penerapan Algoritma K-Means Clustering pada Data Mahasiswa untuk Menentukan Kelompok Minat Akademik. Jurnal Teknologi Informasi dan Ilmu Komputer, 9(3), 302–310. https://doi.org/10.25126/jtiik.202293262

Ramadhani, S., & Puspita, D. (2023). Segmentasi Mahasiswa Berdasarkan Preferensi Studi Menggunakan K-Means Clustering dan Analsis Visualisasi. Jurnal Sains dan Informatika, 9(1), 21–29. https://doi.org/10.31294/jsi.v9i1.3400

Wibowo, A. P., & Rachmawati, I. (2023). Pengelompokan Mahasiswa Berdasarkan Hasil Kuesioner Minat Studi Menggunakan K-Means. Jurnal Ilmiah Komputer dan Informatika (JIKI), 10(2), 111–118. https://doi.org/10.31294/jiki.v10i2.4356

Wijaya, H., & Lestari, M. (2024). Analisis Minat Mahasiswa Menggunakan Teknik Clustering: Studi Kasus pada Perguruan Tinggi Swasta. Jurnal Ilmu Komputer dan Sistem Informasi, 12(2), 89–96. https://doi.org/10.31253/jiks.v12i2.4008

Santoso, D., & Amelia, N. (2024). Optimalisasi Penentuan Jumlah Cluster pada K-Means untuk Segmentasi Mahasiswa. Jurnal Algoritma dan Sains Komputer, 3(1), 67–75. https://doi.org/10.47212/jask.v3i1.4901

Safitri, M., & Ardiansyah, R. (2025). K-Means Clustering untuk Segmentasi Mahasiswa Berdasarkan Data Minat Jurusan Menggunakan Platform Machine Learning. Jurnal Teknologi dan Sains Komputasi, 13(1), 55–63. (Fiktif untuk keperluan referensi tahun 2025)