



Analisis Rekomendasi Media Promosi PPDB SMA Nurul Falah Pekanbaru dengan Algoritma *K-Means Clustering*

Muhammad Taufiq Hidayat^{*1}, Emansa Hasri Putra² Dini Nurmalasari³

Email: ¹taufiq22mttk@mahasiswa.pcr.ac.id, ²emansa@pcr.ac.id, ³dini@pcr.ac.id.

¹ Magister Terapan Teknik Komputer, Politeknik Caltex Riau

² Teknologi Industri, Politeknik Caltex Riau

² Teknologi Informasi, Politeknik Caltex Riau

Diterima: 12 November 2025 | Direvisi: - | Disetujui: 11 Desember 2025

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Perkembangan teknologi informasi telah memberikan dampak besar terhadap cara lembaga pendidikan dalam melakukan kegiatan promosi dan penerimaan peserta didik baru (PPDB). Perubahan perilaku masyarakat menuju era digital menuntut sekolah untuk mengoptimalkan strategi promosi berbasis data yang lebih adaptif dan efektif. SMA Nurul Falah Pekanbaru sebagai salah satu lembaga pendidikan swasta mengalami penurunan jumlah pendaftar pasca pandemi *Covid-19*. Promosi konvensional seperti baliho, spanduk, dan brosur cetak terbukti kurang efektif dalam menjangkau calon siswa secara luas. Penelitian ini bertujuan untuk menganalisis efektivitas media promosi berdasarkan karakteristik calon peserta didik menggunakan algoritma *K-Means Clustering* sebagai metode segmentasi. Data penelitian diperoleh dari hasil pendaftaran PPDB selama tiga tahun yang mencakup informasi demografis, sosial ekonomi, asal sekolah, dan sumber informasi promosi. Proses analisis dilakukan melalui tahapan *data preprocessing*, eksplorasi data (EDA), penentuan jumlah kluster optimal menggunakan metode *Elbow* dan *Silhouette*, serta pengembangan sistem rekomendasi berbasis *web* dengan *Python*, *PHP*, dan *MySQL*. Hasil penelitian menunjukkan konfigurasi optimal pada $k=4$ dengan nilai *Silhouette Score* sebesar 0,351. Empat kluster yang terbentuk menggambarkan perbedaan perilaku calon siswa dalam mengakses informasi pendidikan, dengan media digital menjadi saluran paling efektif. Sistem rekomendasi yang dikembangkan mampu membantu sekolah dalam merancang strategi promosi yang lebih efisien, terukur, dan tepat sasaran berbasis data analitik.

Kata kunci: PPDB, *K-Means Clustering*, Media Promosi, Segmentasi, Data Mining

Analysis of Promotional Media Recommendations for Nurul Falah High School Admissions in Pekanbaru Using the K-Means Clustering Algorithm

Abstract

The advancement of information technology has significantly transformed how educational institutions conduct promotional activities and student admissions. The shift toward digital behavior among society requires schools to adopt more adaptive and data-driven marketing strategies. SMA Nurul Falah Pekanbaru, as one of the private high schools, has experienced a decrease in student enrollment after the *Covid-19* pandemic. Conventional promotional methods such as banners, billboards, and printed brochures have proven less effective in reaching prospective students. This research aims to analyze the effectiveness of promotional media based on student characteristics using the *K-Means Clustering* algorithm as a segmentation method. The dataset was obtained from three years of PPDB registration records, including demographic, socioeconomic, school origin, and promotion media information. The analysis process involved several stages, namely *data preprocessing*, exploratory data analysis (EDA), determination of the optimal number of clusters using the *Elbow* and *Silhouette* methods, and the development of a web-based recommendation system using *Python*, *PHP*, and *MySQL*. The results indicate that the optimal number of clusters

is $k=4$ with a Silhouette Score of 0.351. The four clusters represent distinct behavioral patterns in accessing educational information, with digital media emerging as the most effective channel. The developed recommendation system provides decision support for the school in designing promotional strategies that are more efficient, measurable, and accurately targeted through data analytics-based insights..

Keywords: PPDB, K-Means Clustering, Promotional Media, Segmentation, Data Mining

1. PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk dalam bidang Pendidikan [1], [2], [3]. Pergeseran perilaku masyarakat terhadap penggunaan media digital mendorong lembaga pendidikan untuk melakukan inovasi dalam strategi promosi dan komunikasi [4] [5]. Di era digital saat ini, calon peserta didik dan orang tua lebih aktif mencari informasi melalui media daring seperti media sosial, situs web, dan platform berbagi video [6]. Kondisi ini menuntut sekolah untuk beradaptasi dengan pendekatan promosi yang lebih interaktif, terukur, dan sesuai dengan karakteristik calon siswa yang semakin beragam.

Perubahan pola konsumsi informasi tersebut juga sejalan dengan berkembangnya konsep digital marketing di sektor pendidikan, di mana strategi promosi tidak lagi bersifat satu arah, tetapi berorientasi pada pengalaman pengguna (user experience) [7]. Penelitian oleh Septiana et al. [8] menunjukkan bahwa pemanfaatan media sosial mampu meningkatkan jangkauan promosi lembaga pendidikan hingga 65% dibanding metode tradisional. Namun demikian, efektivitas promosi digital sangat bergantung pada pemahaman terhadap profil calon siswa dan media yang mereka gunakan [9]. Oleh karena itu, diperlukan metode analisis yang dapat mengelompokkan calon siswa secara objektif berdasarkan pola dan karakteristik tertentu.

SMA Nurul Falah Pekanbaru sebagai salah satu lembaga pendidikan swasta di Kota Pekanbaru menghadapi tantangan serupa. Dalam beberapa tahun terakhir, terutama pasca pandemi Covid-19, terjadi penurunan jumlah pendaftar baru yang cukup signifikan. Pembatasan kegiatan tatap muka dan perubahan perilaku masyarakat terhadap aktivitas digital berdampak langsung terhadap efektivitas metode promosi konvensional seperti pamflet, baliho, dan brosur cetak [10]. Strategi promosi tradisional tersebut tidak lagi mampu menjangkau segmen calon siswa secara luas maupun tepat sasaran. Situasi ini menuntut pihak sekolah untuk mengoptimalkan sumber daya yang tersedia melalui pendekatan berbasis data agar dapat meningkatkan efisiensi dan ketepatan sasaran dalam kegiatan promosi.

Dalam konteks tersebut, pendekatan berbasis data mining menjadi solusi yang potensial untuk mengidentifikasi pola tersembunyi dalam data pendaftaran siswa [11]. Dengan metode ini, lembaga pendidikan dapat memahami preferensi calon siswa berdasarkan atribut-atribut tertentu seperti asal sekolah, tingkat ekonomi, serta sumber informasi promosi yang digunakan [2]. Salah satu algoritma yang paling populer digunakan dalam tahap segmentasi adalah K-Means Clustering, karena kemampuannya dalam membagi data menjadi beberapa kelompok dengan karakteristik serupa secara efisien dan cepat [12]. Pendekatan ini telah terbukti efektif dalam berbagai studi, termasuk pada bidang pemasaran digital, perilaku konsumen, dan analisis Pendidikan [13].

Melalui penerapan algoritma K-Means Clustering, proses mengidentifikasi kelompok calon siswa berdasarkan kesamaan karakteristik, serta memberikan rekomendasi media promosi yang paling efektif untuk setiap kluster [14]. Hasil dari analisis ini dapat membantu SMA Nurul Falah Pekanbaru dalam meningkatkan efisiensi penggunaan anggaran promosi, memperluas jangkauan calon pendaftar, serta memperkuat daya saing sekolah di tengah ketatnya kompetisi antar lembaga pendidikan. Selain itu, penelitian ini diharapkan dapat menjadi contoh penerapan nyata integrasi data science dalam pengambilan keputusan strategis di bidang pendidikan.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining dengan algoritma *K-Means Clustering* sebagai metode utama untuk melakukan segmentasi calon peserta didik baru berdasarkan kemiripan karakteristik tertentu. Pendekatan *unsupervised learning* ini dipilih karena mampu mengelompokkan data tanpa memerlukan label kelas, sehingga efektif dalam menganalisis data pendaftaran siswa yang bersifat heterogen dan tidak terstruktur [15].

1. Jenis dan Sumber Data

Data penelitian diperoleh dari basis data Penerimaan Peserta Didik Baru (PPDB) di SMA Nurul Falah Pekanbaru selama tiga tahun terakhir (2021–2023). Data terdiri dari 1.253 entri yang mencakup beberapa atribut penting, yaitu:

Data demografis: jenis kelamin, asal sekolah, dan domisili calon siswa.

Data sosial ekonomi: pekerjaan dan penghasilan orang tua.

Sumber informasi promosi: media sosial, brosur, spanduk, kunjungan sekolah, dan rekomendasi keluarga atau alumni.

Seluruh data dikumpulkan dalam format *spreadsheet* dan diekspor ke format *Comma-Separated Values* (CSV) untuk kemudian diolah menggunakan bahasa pemrograman Python pada lingkungan *Google Colab*.

2. Tahapan Pra-pemrosesan Data (*Data Preprocessing*)

Sebelum proses analisis dilakukan, data mentah perlu melalui beberapa tahap pra-pemrosesan untuk memastikan konsistensi dan validitas. Tahapan ini mencakup:

- A. *Data Cleaning*: Menghapus data duplikat dan memperbaiki nilai kosong (*missing values*) dengan pendekatan *mean imputation* untuk data numerik serta *mode imputation* untuk data kategorikal.
- B. *Data Transformation*: Mengonversi variabel kategorikal seperti “media promosi” menjadi nilai numerik menggunakan metode *label encoding* agar dapat diolah oleh algoritma *K-Means*.
- C. *Normalization*: Menggunakan metode Min-Max Scaling untuk menyamakan rentang nilai tiap variabel antara 0 dan 1, sehingga perbedaan skala tidak memengaruhi hasil pengelompokan.
- D. *Exploratory Data Analysis* (EDA): Dilakukan untuk memahami distribusi data dan mengidentifikasi potensi *outlier* melalui visualisasi histogram, diagram batang, serta analisis korelasi antar variabel menggunakan pustaka *Matplotlib* dan *Seaborn*.

3. Proses *Clustering* Menggunakan *K-Means*

Setelah data siap, proses *Clustering* dilakukan menggunakan algoritma *K-Means*. Proses ini diawali dengan menentukan jumlah kluster (*k*) optimal melalui dua metode evaluasi, yaitu:

Metode *Elbow*, untuk mengidentifikasi titik penurunan (*Elbow point*) yang signifikan pada grafik *inertia*,

Metode *Silhouette Coefficient*, untuk menilai tingkat kohesi internal dan separasi antar kluster.

Pada tahap ini, dilakukan iterasi sebanyak 10 kali untuk setiap nilai *k* (mulai dari *k*=2 hingga *k*=10) guna memperoleh stabilitas hasil. Nilai *k* terbaik diperoleh pada *k*=4, dengan *Silhouette Score* sebesar 0,351 yang menunjukkan struktur kluster cukup baik.

Secara konseptual, proses *K-Means* dapat dijelaskan melalui alur kerja berikut:

- A. Menentukan jumlah kluster (*k*) dan titik pusat awal (*centroid*).
- B. Menghitung jarak *Euclidean* antara setiap data dan *centroid*.
- C. Menetapkan setiap data ke kluster dengan jarak terdekat.
- D. Memperbarui posisi *centroid* berdasarkan rata-rata anggota setiap kluster.
- E. Mengulangi langkah 2–4 hingga posisi *centroid* tidak lagi berubah secara signifikan.

Tahapan tersebut menghasilkan empat kelompok utama calon siswa dengan karakteristik berbeda berdasarkan preferensi media promosi dan latar belakang sosial ekonomi.

4. Pengembangan Sistem Rekomendasi Berbasis *Web*

Untuk mendukung interpretasi hasil kluster, penelitian ini juga mengembangkan sistem rekomendasi berbasis web. Sistem dibangun menggunakan kombinasi *Python* (untuk analisis data), *PHP* (untuk tampilan antarmuka pengguna), dan *MySQL* (untuk penyimpanan data). *Dashboard* sistem menampilkan:

Grafik distribusi calon siswa per kluster,

Statistik demografis tiap kluster,

Rekomendasi media promosi yang sesuai untuk setiap segmen.

Integrasi data dilakukan secara real-time melalui *Application Programming Interface (API)* sederhana, sehingga sekolah dapat memperbarui hasil analisis setiap periode PPDB berikutnya tanpa perlu melakukan pemrosesan manual.

5. Evaluasi Model

Evaluasi hasil dilakukan menggunakan dua metrik utama:

Inertia (Within-Cluster Sum of Squares – WCSS) untuk menilai kekompakan tiap klaster,

Silhouette Score, untuk menilai sejauh mana objek berada dalam klaster yang tepat.

Nilai *inertia* sebesar 309,8 menunjukkan jarak antar anggota klaster relatif kecil, sedangkan *Silhouette Score* sebesar 0,351 menandakan struktur klaster cukup stabil. Hasil ini menjadi dasar bahwa *K-Means* mampu memisahkan kelompok calon siswa dengan baik sesuai karakteristik media promosi yang diakses.

3. HASIL DAN PEMBAHASAN

Hasil analisis *Clustering* menggunakan algoritma *K-Means* menunjukkan bahwa jumlah klaster optimal diperoleh pada $k = 4$, dengan nilai *inertia* sebesar 309,8 dan *Silhouette Score* sebesar 0,351. Nilai tersebut menandakan bahwa pemisahan antar klaster sudah cukup baik, dengan tingkat homogenitas yang memadai dalam setiap kelompok data. Proses penentuan nilai k dilakukan menggunakan metode *Elbow* untuk melihat titik perubahan signifikan pada grafik *inertia*, dan dikonfirmasi menggunakan *Silhouette Coefficient* untuk memastikan stabilitas hasil klaster.

Empat klaster yang terbentuk menggambarkan karakteristik calon peserta didik berdasarkan latar belakang sosial ekonomi, sumber informasi, dan preferensi media promosi:

1. Klaster 1 terdiri dari calon siswa dengan latar belakang ekonomi menengah ke atas yang aktif menggunakan media sosial seperti Instagram dan TikTok sebagai sumber utama informasi sekolah. Kelompok ini memiliki tingkat literasi digital yang tinggi dan lebih responsif terhadap konten promosi yang bersifat visual dan interaktif.
2. Klaster 2 mencakup calon siswa dari kelompok ekonomi menengah dengan preferensi utama pada promosi langsung melalui kegiatan sekolah seperti pameran pendidikan, kunjungan ke SMP, atau *open house*. Klaster ini menunjukkan pentingnya interaksi personal dan kehadiran fisik dalam membangun kepercayaan calon siswa dan orang tua.
3. Klaster 3 didominasi oleh calon siswa dari wilayah pinggiran yang masih mengandalkan media cetak seperti brosur, baliho, dan spanduk sebagai sumber informasi utama. Hal ini menunjukkan bahwa media tradisional masih relevan untuk segmen dengan akses internet terbatas.
4. Klaster 4 berisi calon siswa yang memperoleh informasi melalui testimoni keluarga, teman, atau alumni sekolah. Segmen ini memperlihatkan kuatnya pengaruh *word of mouth* dan jejaring sosial dalam pengambilan keputusan memilih sekolah.

Hasil pengelompokan ini memberikan wawasan strategis bagi pihak sekolah untuk menyesuaikan pendekatan promosi terhadap karakteristik masing-masing segmen. Misalnya, promosi digital dapat difokuskan untuk klaster 1 melalui kampanye media sosial dan *content marketing*, sementara kegiatan tatap muka dan kolaborasi dengan sekolah asal lebih sesuai untuk klaster 2. Adapun klaster 3 dapat dioptimalkan dengan distribusi media cetak di wilayah dengan penetrasi internet rendah, sedangkan klaster 4 dapat diperkuat melalui program *alumni ambassador* dan testimoni digital.

Implementasi hasil analisis ini dituangkan dalam dashboard berbasis *web* yang dikembangkan menggunakan *Python* dan *PHP*. *Dashboard* tersebut menampilkan distribusi calon siswa per klaster dalam bentuk visualisasi grafik dan peta interaktif, sehingga memudahkan pihak sekolah dalam melakukan analisis cepat dan pengambilan keputusan berbasis data. Fitur sistem juga menyediakan rekomendasi otomatis terkait media promosi yang paling efektif untuk setiap segmen, berdasarkan hasil klasterisasi.

Secara keseluruhan, penerapan *K-Means Clustering* berhasil membantu SMA Nurul Falah Pekanbaru memahami pola perilaku calon siswa secara lebih mendalam. Dengan dukungan sistem rekomendasi visual, proses perencanaan dan evaluasi strategi promosi menjadi lebih efisien, terukur, dan berorientasi pada data (*data-driven decision making*).

4. KESIMPULAN

Implementasi algoritma *K-Means Clustering* dalam analisis media promosi PPDB SMA Nurul Falah Pekanbaru berhasil mengidentifikasi empat kelompok utama calon siswa dengan karakteristik yang berbeda. Media digital terbukti menjadi saluran promosi paling efektif menjangkau seluruh segmen. Sistem rekomendasi yang dibangun dapat membantu pihak sekolah dalam merumuskan strategi promosi yang lebih terarah, efisien, dan berbasis data. Pengembangan lanjutan disarankan untuk menambahkan variabel perilaku daring dan integrasi dengan sistem promosi *real-time*.

DAFTAR PUSTAKA

- [1] O. Rangga, D. T. Bilo, and D. Yuliana, "Penggunaan Teknologi Informasi dalam Pembelajaran Pendidikan Agama Kristen," *Didache: Jurnal Teologi dan Pendidikan Kristiani*, vol. 5, no. 2, pp. 127–140, Aug. 2024, doi: 10.55076/didache.v5i2.164.
- [2] M. Ibnu Sholeh, "Menghadapi Persaingan Sengit Lembaga Pendidikan: Strategi Diferensiasi dalam Pemasaran Lembaga Pendidikan Islam di Indonesia," *Jurnal Manajemen Pendidikan Islam*, vol. 1, no. 3, pp. 192–222, Jul. 2020, doi: 10.37348/aksi.v1i3.259.
- [3] R. Syerlita and I. Siagian, "Dampak Perkembangan Revolusi Industri 4.0 Terhadap Pendidikan Di Era Globalisasi Saat Ini," *Journal on Education*, vol. 07, no. 01, pp. 3507–3515, Jul. 2024.
- [4] A. Hidayat, N. Alifah, and A. A. Rodiansjah, "Strategi Manajemen Pemasaran Pendidikan Menggunakan Platform Media Sosial," *Cakrawala Repositori IMWI*, vol. 7, no. 4, pp. 1228–1239, Jul. 2024, doi: 10.52851/cakrawala.v7i4.713.
- [5] S. Mulyani, A. Idi, I. P. Pratama, and Y. Yuniar, "Transformasi Branding Sekolah melalui Digital Marketing: Studi di SMA Negeri 3 Prabumulih," *Jurnal Ilmiah Global Education*, vol. 6, no. 3, pp. 2079–2093, Sep. 2025, doi: 10.55681/jige.v6i3.3911.
- [6] N. Janattaka and P. Chika Cecilia Adella, "Peran Platform Digital dalam Pembelajaran Daring," *Almufi Jurnal Pendidikan*, vol. 1, no. 3, pp. 138–146, Dec. 2021, [Online]. Available: <http://almufi.com/index.php/AJP/article/view/74>
- [7] J. D. S. Amory, M. Mudo, and R. J., "Transformasi Ekonomi Digital dan Evolusi Pola Konsumsi: Tinjauan Literatur tentang Perubahan Perilaku Belanja di Era Internet," *Jurnal Minfo Polgan*, vol. 14, no. 1, pp. 28–37, Feb. 2025, doi: 10.33395/jmp.v14i1.14608.
- [8] V. Septiana, B. S. E. Murtiningsih, and N. Astagini, "Efektivitas Media Sosial dalam Mempromosikan Pendidikan Tinggi Berbasis Vokasional Kepada Calon Mahasiswa," *Jurnal Pustaka Komunikasi*, vol. 8, no. 1, pp. 63–73, Mar. 2025, doi: 10.32509/pustakom.v8i1.4428.
- [9] Unzil Al Huda, "Analisis Efektivitas Kampanye Iklan Digital dalam Meningkatkan Citra Lembaga Pendidikan Islam (Studi kasus : MI. Bahrul Ulum Watupanjang - Krucil)," *Advances In Education Journal*, vol. 2, no. 1, pp. 276–286, Aug. 2025, [Online]. Available: <https://journal.al-afif.org/index.php/aej/article/view/169>
- [10] J. Ilmiah and K. Makna, "Pemanfaatan New Media untuk Efektivitas Komunikasi di Era Pandemi," *JURNAL ILMIAH KOMUNIKASI MAKNA*, vol. 10, no. 1, pp. 28–42, Feb. 2022, doi: 10.30659/jikm.10.1.28-42.
- [11] A. Ermillian and K. Nugroho, "Perancangan Model Deteksi Potensi Siswa Putus Sekolah Menggunakan Metode Logistic Regression dan Decision Tree," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 9, no. 3, pp. 281–295, Dec. 2024, doi: 10.30591/jpit.v9i3.8007.
- [12] A. Fauzan, N. Suarna, I. Ali, and H. Susana, "Penerapan Algoritma *K-Means Clustering* untuk Meningkatkan Model Pengelompokan dan Kinerja Jaringan Wi-Fi Secara Optimal," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6272.
- [13] S. Mutiah, Y. Hasnataeni, A. Fitrianto, E. Erfiani, and L. M. R. D. Jumansyah, "Perbandingan Metode Klastering *K-Means* dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat," *Teorema: Teori dan Riset Matematika*, vol. 9, no. 2, p. 247, Sep. 2024, doi: 10.25157/teorema.v9i2.16290.
- [14] N. L. Hasanah, A. A. Ningrum, and K. Kamarudin, "Penerapan *K-Means Clustering* Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Universitas Muhammadiyah Banjarmasin," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 2, Jun. 2025, doi: 10.35957/jatisi.v12i2.11488.

- [15] R. S. Nurhalizah, R. Ardianto, and P. Purwono, "Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review," *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 61–72, Aug. 2024, doi: 10.54082/jiki.168.