



Pengelompokan Pengguna Media Sosial Berdasarkan Pola Interaksi Menggunakan K-Means

Muhamad Ilhan Mansiz^{1*}, Zaehol Fatah²,

¹ Teknologi Informasi, Universitas Ibrahimy

²Sistem Informasi, Universitas Ibrahimy

[1*ilhanmansiz2022@gmail.com](mailto:ilhanmansiz2022@gmail.com), [2zaeholfatah@gmail.com](mailto:zaeholfatah@gmail.com),

Abstrak

Mengelompokkan pengguna media sosial berdasarkan pola interaksi dengan menggunakan metode K-Means Clustering. Dalam era digital saat ini, memahami perilaku pengguna media sosial menjadi sangat penting bagi platform digital untuk meningkatkan keterlibatan dan retensi pengguna. Dengan mengidentifikasi interaksi pengguna, platform dapat merancang strategi yang lebih efektif untuk memenuhi kebutuhan dan preferensi pengguna. Dataset yang digunakan dalam penelitian ini mencakup atribut-atribut seperti BounceRates dan ExitRates untuk mencerminkan tingkat ketertarikan awal dan durasi interaksi pengguna, serta atribut Month, Region, VisitorType, dan Weekend untuk menangkap faktor temporal, geografis, dan perilaku kunjungan. Data tersebut diolah melalui proses normalisasi dan pengkodean untuk memastikan kompatibilitas dalam analisis K-Means. Hasil clustering menunjukkan adanya dua kelompok utama pengguna media sosial dengan karakteristik interaksi yang berbeda. Kelompok pertama terdiri dari pengguna dengan keterlibatan awal yang tinggi dan interaksi yang lebih lama, sedangkan kelompok kedua cenderung memiliki minat awal yang rendah dan durasi interaksi yang singkat. Temuan ini memberikan wawasan bagi pengelola platform media sosial untuk mengembangkan strategi yang lebih personalisasi dalam meningkatkan pengalaman pengguna, baik melalui konten yang lebih relevan maupun pendekatan pemasaran yang terarah. Dengan adanya pengelompokan ini, diharapkan platform media sosial dapat meningkatkan efektivitas interaksi pengguna dan memperkuat loyalitas pengguna, yang pada akhirnya berdampak positif pada pertumbuhan platform secara keseluruhan.

Kata Kunci: Algoritma *K-Means Clustering*, media sosial, Data Mining

PENDAHULUAN

Media sosial telah menjadi bagian penting dari kehidupan sehari-hari, memungkinkan pengguna untuk terhubung, berbagi informasi, dan berinteraksi secara real-time. Dengan jutaan pengguna yang aktif setiap harinya, platform media sosial mengumpulkan jumlah data interaksi yang sangat besar, seperti jumlah posting, komentar, like, share, serta pola pertemanan. Pola interaksi ini mencerminkan perilaku dan preferensi pengguna, yang dapat digunakan untuk berbagai tujuan seperti pemasaran, personalisasi konten, hingga pengelolaan komunitas.(Kanavos, A., Kiourt, C., Sakas, D. P., & Giannakopoulos, 2023)

Namun, dengan volume data yang sangat besar, memahami dan mengelompokkan pengguna berdasarkan pola interaksi menjadi tantangan yang kompleks. Salah satu metode yang efektif untuk mengatasi tantangan ini adalah K-Means Clustering, sebuah algoritma pembelajaran mesin yang berfungsi untuk membagi data ke dalam kelompok-kelompok (*clusters*) berdasarkan kesamaan karakteristik.(Kumar, A., & Singh, 2021) Di dalam clustering terdapat beberapa metode salah satunya adalah metode K-means. K-means adalah salah satu proses metode pengelompokan data ke dalam beberapa kelompok atau klaster berdasarkan kesamaan fitur.(Zulyan Pratama et al., 2024)

Dengan menggunakan K-Means,dapat diidentifikasi grup pengguna yang memiliki pola interaksi yang serupa. Sebagai contoh, pengguna yang sering memberikan like namun jarang berkomentar mungkin termasuk dalam kelompok tertentu, sementara pengguna yang aktif memposting konten tetapi jarang berinteraksi dengan postingan orang lain dapat berada di kelompok yang berbeda. Pengelompokan ini dapat membantu platform media sosial atau perusahaan yang beroperasi di dalamnya untuk memahami perilaku pengguna lebih baik dan menciptakan strategi yang lebih efektif untuk meningkatkan keterlibatan pengguna, personalisasi konten, atau bahkan monetisasi platform.

Pendekatan ini dapat diterapkan pada berbagai metrik interaksi, seperti frekuensi posting, intensitas komentar, interaksi dengan teman, serta preferensi terhadap jenis konten tertentu. (Almoqbel, M. I., & Moradi, 2022)

Penelitian ini akan menjelaskan bagaimana algoritma K-Means digunakan untuk mengelompokkan pengguna media sosial berdasarkan pola interaksi, seperti frekuensi posting, jumlah interaksi dengan teman, jenis konten yang sering dibagikan, dan respons terhadap konten. Selain itu, hasil pengelompokan ini dapat memberikan wawasan lebih dalam tentang preferensi dan kecenderungan pengguna, yang berguna untuk pengembangan strategi pemasaran, personalisasi konten, dan peningkatan keterlibatan. Pengelompokan ini juga dapat membantu pengelola platform dalam mengenali pola tertentu yang menunjukkan tren perilaku dalam berbagai kelompok pengguna.

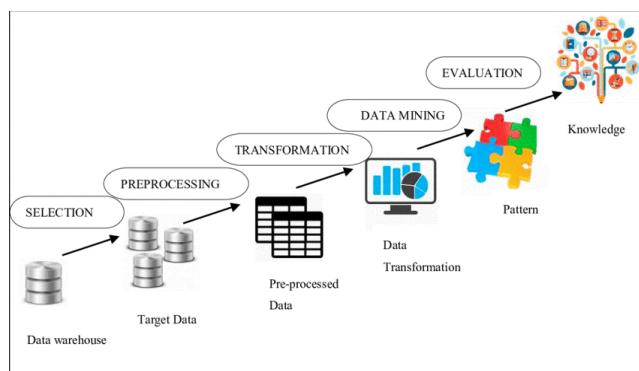
METODE

Jenis Penelitian

Jenis penelitian ini adalah deskriptif kuantitatif dengan pendekatan rekayasa softwareserta Library Online Research. Pengumpulan data dilakukan dengan teknik kepustakaan. Penelitian deskriptif kuantitatif dipilih agar dapat menggambarkan, mendeskripsikan serta menjelaskan sesuatu secara objektif serta dapat menarik kesimpulan dari data berupa angka-angka yang telah disajikan, dan dengan cara ini peneliti menghimpun, mengelola data serta menganalisis data-data yang terkait dengan pola interaksi pengguna medsos dengan metode K-Means (Zulfikar et al., 2022)nm

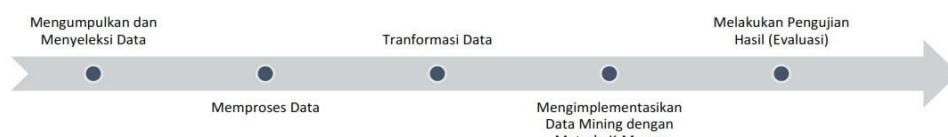
Knowledge Discovery in Database (KDD)

Knowledge discovery in database (KDD) adalah metode teknis yang berguna untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, pola yang sudah ditemukan bersifat sah dan baru sehingga dapat bermanfaat dan dapat dimengerti.(Arrohman & Fatah, 2024) Proses dalam KDD terdapat 5 tahapan yaitu seleksi data dari data sumber ke data target, tahap *pre-processing*, *transformasi*, *data mining* dan tahap evaluasi. Tahap seleksi dilakukan untuk menargetkan data yang digunakan untuk penelitian, tahap *pre-procesing* dapat dilakukan integrasi data atau pengabungan data serta dilakukan *cleaning* data, yaitu dengan menhilangkan *noise*, data *redundant*, *inkonsistensi* data, serta data yang tidak relavan, tahap transformasi adalah pengabungan data dan penyesuaian format data agar dapat diproses pada tahap data mining, tahap data mining dilakukan menggunakan algoritma yang cocok untuk permasalahan dalam data, serta tahap evaluasi yang digunakan untuk pengujian dalam data.(Yogianto et al., 2024) seperti ditampilkan pada gambar 1 dibawah ini :



Gambar 1 Metode dan Tahapan Penelitian

Pada gambar 1 digambarkan alur mengenai tahapan metode KDD, setelah disesuaikan dengan tahapan-tahapan KDD dihasilkan kerangka kerja penelitian yang dijabarkan pada Gambar 2.



Gambar 2 Kerangka Kerja Penelitian

Media Sosial

Media sosial adalah bentuk mediaonline yang mencakup blog, jejaring sosial, wiki, dan dunia virtual, memungkinkan partisipasi,berbagai,dan penciptaan konten pengguna. Andreas Kaplan dan Michael Haenlein mendefinisikannya sebagai kelompok aplikasi internet di atas ideologi dan teknologi web 2.0, memfasilitasi pertukaran konten pengguna.jejaring sosial, seperti facebook dan Twitter, memungkinkan pembuatan halaman web pribadi dan berbagai informasi. Media sosial menggunakan internet, mengundang partisipasi terbuka dan umpan balik. Dengan perkembangan teknologi internet dan perangkat seluler, media sosial tumbuh pesat, mengubah peran media massa, dan mempercepat penyebaran informasi global. Ciri khasnya melibatkan pengiriman pesan ke banyak orang tanpa gatekeeper,penyebaran cepat,dan pengguna memiliki atas waktu interaksi, mendorong partisipasi aktif dan berbagai informasi secara terbuka(Gani, 2020)

Data Mining

Data mining merupakan proses kompleks yang menggunakan statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstrak informasi berharga dari database besar. Sebagaimana dinyatakan oleh (Sukarna & Ansori, 2022), data mining bukan hanya mengenai pengumpulan data, tetapi juga melibatkan analisis dan prediksi informasi yang ingin ditampilkan. Informasi yang terkumpul diolah dalam database untuk mendukung pengambilan keputusan. (Oon

Wira Yuda et al., 2022) merinci tahap-tahap data mining, termasuk pembersihan data, integrasi data dari berbagai sumber, pemilihan data yang relevan, transformasi data, proses mining, evaluasi pola, dan penyajian pengetahuan. Tahap akhir melibatkan formulasi keputusan berdasarkan hasil analisis yang diperoleh.

Algoritma K-Means

Algoritma K-Means, sebagaimana dijelaskan dalam teks, merupakan suatu algoritma Teknik clustering yang dimulai dengan pemilihan secara acak terhadap cluster yang diinginkan untuk membentuk kelompok dari data yang akan di-cluster. K-Means dikenal sebagai salah satu algoritma paling popular dalam clustering karena sifatnya yang sederhana dan efisien, seperti yang diakui oleh penelitian (Abdul Rohman, 2020) yang menempatkannya sebagai salah satu dari 10 algoritma data mining teratas. Algoritma ini efektif dalam mengelompokkan data ke dalam cluster-cluster yang saling berdekatan, memfasilitasi analisis pola dan struktur data dengan cara yang mudah dimengerti dan diterapkan (Kamila, 2021).

Clustering

Clustering adalah metode efektif untuk mengorganisir sejumlah besar dokumen teks yang tidak berurutan menjadi kluster yang bermakna dan koheren. Tujuannya adalah menciptakan struktur yang memungkinkan navigasi dan penelusuran informasi yang intuitif serta informatif. Proses ini membantu dalam mengidentifikasi keterkaitan dokumen, mempermudah pemahaman konten terkait, dan meningkatkan efisiensi akses terhadap informasi yang tersebar luas(Widaningrum et al., 2022)

Rapid Miner

RapidMiner merupakan perangkat lunak yang digunakan untuk analisis data dan pembelajaran mesin. Aplikasi ini menawarkan berbagai operator yang membentuk alur kerja, mencakup seluruh tahap dalam proses data mining, mulai dari pembersihan data, pemilihan fitur, hingga pemodelan(Amril Mutoi Siregar, S.Kom., M.Kom. DAN Puspabhuana, S.Kom., 2021)

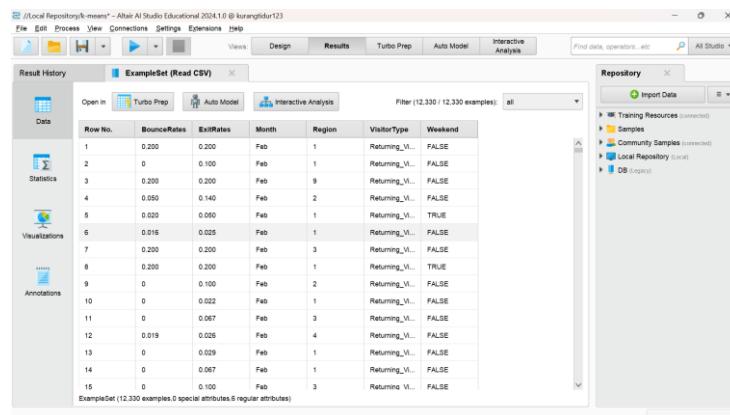
HASIL DAN PEMBAHASAN

Implementasi penelitian ini menggunakan perhitungan *Algoritma K-Means Clustering* yang dilakukan dengan tools Rapid Miner Studio. Rapid Miner sebagai tools yang mampu memudahkan usernya dalam melakukan operator. Operator berfungsi untuk memodifikasi data dan selanjutnya dihubungkan dengan node – node operator.

Dari proses penelitian yang dilakukan didapatkan hasil dari 12330 dataset dari Cluster Pertama terdiri dari pengguna yang memiliki tingkat ketertarikan awal (BounceRates) dan durasi interaksi (ExitRates) yang rendah. Pengguna dalam kelompok ini menunjukkan keterlibatan yang lebih tinggi di platform, yang mungkin mencerminkan minat yang lebih besar dan potensi loyalitas pengguna yang lebih tinggi. Cluster Kedua mencakup pengguna dengan tingkat ketertarikan awal dan durasi interaksi yang lebih tinggi. Hal ini dapat menunjukkan bahwa pengguna dalam kelompok ini cenderung berinteraksi lebih singkat atau memiliki minat yang lebih rendah terhadap konten yang tersedia, sehingga berpotensi meninggalkan platform dengan cepat.

Dataset

Dataset yang digunakan dalam penelitian ini merupakan data interaksi pengguna media sosial, mencakup atribut-atribut seperti tingkat ketertarikan awal (BounceRates), durasi interaksi (ExitRates), waktu penggunaan (Month), lokasi pengguna (Region), frekuensi kunjungan (VisitorType), dan aktivitas akhir pekan (Weekend). Data ini telah diproses dengan mengonversi atribut non-numerik menjadi numerik untuk memudahkan analisis menggunakan metode K-Means Clustering. Data interaksi yang dianalisis berjumlah 12.330 data dengan 6 variabel, yang diolah untuk menemukan pola interaksi spesifik setiap kelompok pengguna.seperti ditampilkan pada gambar 3 dibawah ini:

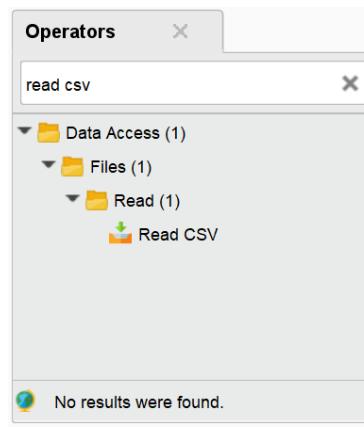


Gambar 3 Testing data- load current

Pengolahan Data dengan Rapid Miner

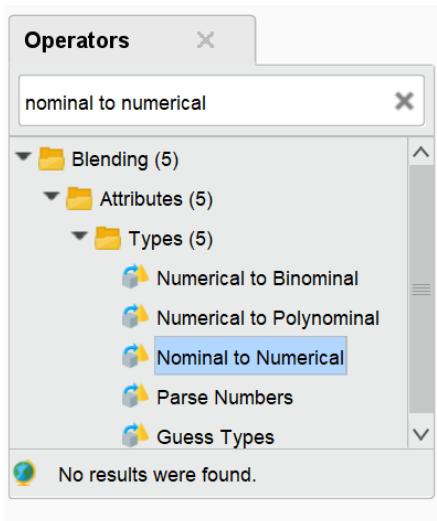
Untuk mengelompokkan data dalam penelitian ini, akan dilakukan dengan menggunakan software rapidminer sebagai berikut :

- Buka *Software Rapidminer*
- Lalu pilih menu file kemudian pilih *new proses*.
- Selanjutnya masuk kedalam panel operator kemudian ketik *Read CSV*, lalu drag operator *read CSV* agar operator tersebut muncul di panel proses dan bisa digunakan untuk memasukkan data *excel* yang telah melalui tahap data *cleansing* dan data *transformation*. Seperti yang terlihat pada gambar 4 berikut:



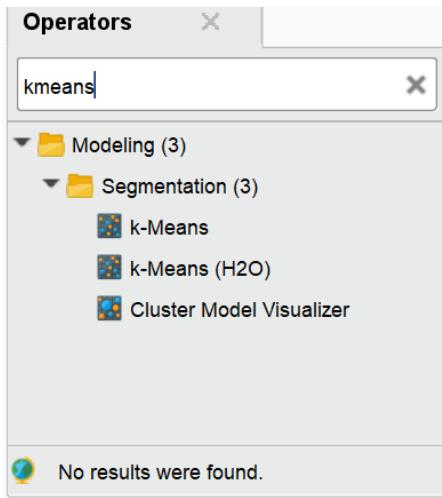
Gambar 4 Read CSV

- Langkah selanjutnya adalah encoding dengan cara atribut kategorikal seperti *Month*, *VisitorType*, dan *Weekend* diubah menjadi bentuk numerik agar dapat diproses dalam analisis K-Means. seperti ditampilkan pada gambar 5 dibawah ini :



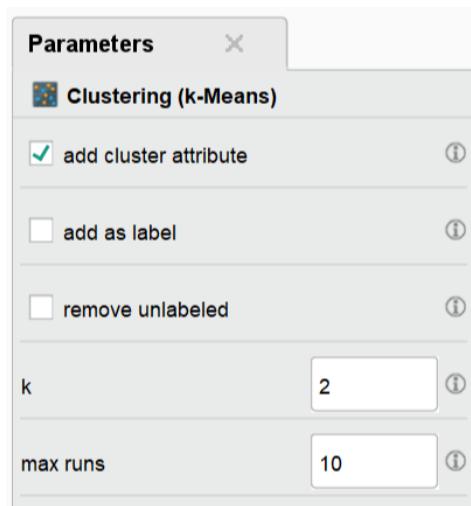
Gambar 5 Nominal To Numerical

- e) Langkah selanjutnya adalah memasukkan model yang akan digunakan untuk proses clustering data mining dengan cara masuk ke menu operator kemudian pilih K-Means seperti pada gambar 6 berikut:



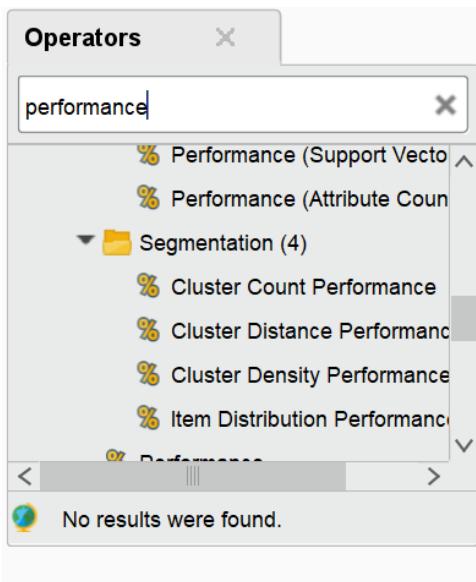
Gambar 6 Memilih Pemodelan Klustering

- f) Selanjutnya, pada proses *clustering* tentukan terlebih dahulu jumlah klasternya menjadi K=2 dan *mixed measure* menjadi *mixed Euclidean distance* seperti pada gambar 7 berikut:



Gambar 7 Menentukan jumlah k

- g) Langkah selanjutnya adalah memasukkan model yang akan mengukur seberapa baik hasil *clustering* yang dilakukan dengan cara masuk ke menu operator kemudian pilih *cluster distance performance* seperti pada gambar 8 berikut:



Gambar 8 Memilih Pemodelan Cluster Distance Performance

- h) Selanjutnya adalah menghubungkan konektor masing – masing proses pada main process seperti pada gambar 9 dan selanjutnya klik button run untuk memulai pengklusteran.



Gambar 9 Main proses Data Mining

Hasil Klasterisasi dengan Algoritma K-Means

a) Data View

Setelah melewati tahapan proses K-Means Clustering menggunakan aplikasi Rapidminer Studio maka proses berikutnya adalah menampilkan hasil dari penerapan tersebut. Pada gambar 10 akan ditampilkan data view hasil clusterisasi yang telah didapatkan sebagai berikut:

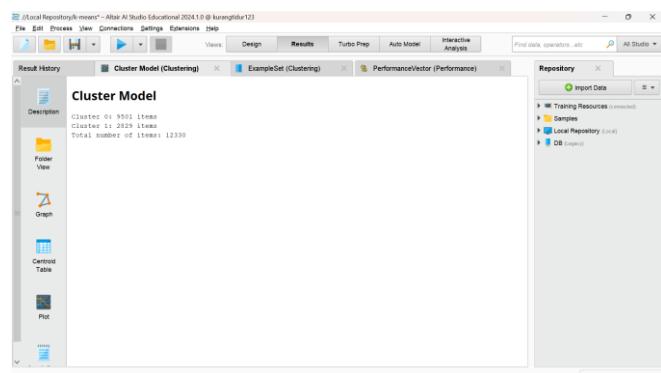
Row No.	id	cluster	Month = Feb	Month = Mar	Month = May	Month = Oct	Month = June	Month = J
1	1	cluster_0	1	0	0	0	0	0
2	2	cluster_0	1	0	0	0	0	0
3	3	cluster_1	1	0	0	0	0	0
4	4	cluster_0	1	0	0	0	0	0
5	5	cluster_0	1	0	0	0	0	0
6	6	cluster_0	1	0	0	0	0	0
7	7	cluster_0	1	0	0	0	0	0
8	8	cluster_0	1	0	0	0	0	0
9	9	cluster_0	1	0	0	0	0	0
10	10	cluster_0	1	0	0	0	0	0
11	11	cluster_0	1	0	0	0	0	0
12	12	cluster_0	1	0	0	0	0	0
13	13	cluster_0	1	0	0	0	0	0
14	14	cluster_0	1	0	0	0	0	0

Gambar 10 Tampilan Hasil Kluster pada Data View

Gambar 10 merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya berjumlah 12.330 data.

a) Cluster Model

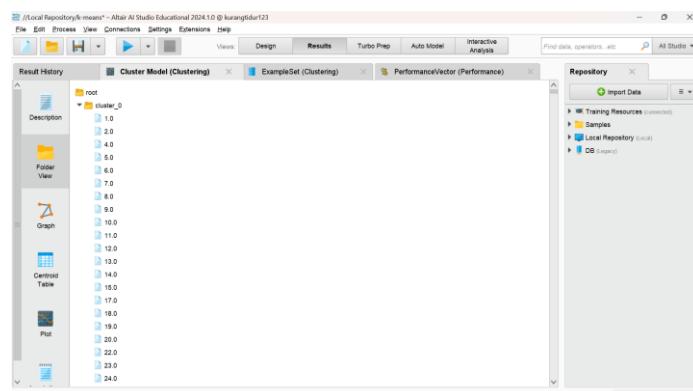
Tampilan hasil seperti gambar 11 berupa hasil pengujian dataset yang berjumlah 12.330 data menggunakan software rapidminer, sehingga berbentuk 2 cluster. Pada cluster 0 (cluster pertama) terdapat 9501 data, cluster 1 (cluster kedua) terdapat 2829 data.



Gambar 11 Cluster Model Hasil Data Mining

b) Folder View

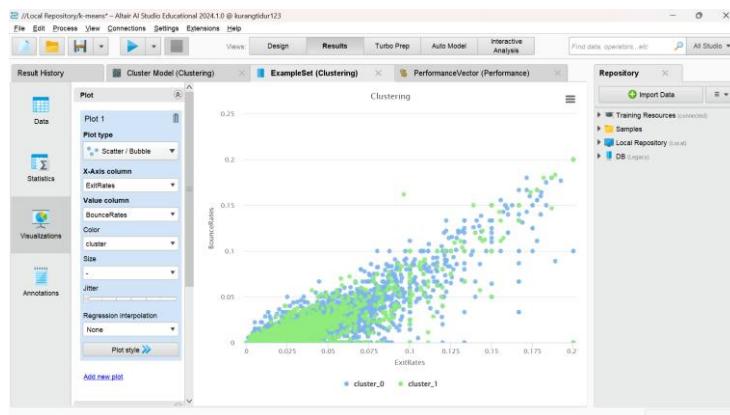
Pada gambar 12 berikut ini ditampilkan tampilan data pengguna media sosial berdasarkan pola interaksi, dengan fokus pada variabel seperti tingkat ketertarikan awal, durasi interaksi, dan aktivitas berdasarkan waktu serta frekuensi kunjungan. Data ini digunakan untuk mengidentifikasi pengguna yang memiliki potensi keterlibatan tinggi, yang memungkinkan platform untuk menyusun strategi peningkatan pengalaman pengguna berdasarkan cluster atau kelompok pola interaksi yang terbentuk:



Gambar 12 Folder View

c) Visualization Hasil Clustering

Chart di bawah ini menunjukkan hasil visualisasi pengelompokan atau clustering pengguna media sosial berdasarkan pola interaksi, menggunakan 6 variabel utama. Data diolah menjadi 2 cluster utama untuk menyoroti perbedaan karakteristik pengguna di setiap kelompok. Hasilnya ditampilkan dalam bentuk Scatter, yang membantu memperjelas distribusi dan komposisi pengguna dalam setiap cluster berdasarkan pola interaksi. Berikut ini tampilan visualization gambar Tampilan Scatter pada chart gambar 13 berikut :



Gambar 13 Scatter Chart

KESIMPULAN

Kesimpulan yang diperoleh dari hasil pengelompokan pengguna media sosial berdasarkan pola interaksi menggunakan metode K-Means menunjukkan adanya perbedaan yang signifikan dalam karakteristik pengguna. Dari analisis ini, dua kelompok utama (cluster) berhasil diidentifikasi:

1. Cluster Pertama terdiri dari pengguna yang memiliki tingkat ketertarikan awal (BounceRates) dan durasi interaksi (ExitRates) yang rendah. Pengguna dalam kelompok ini menunjukkan keterlibatan yang lebih tinggi di platform, yang mungkin mencerminkan minat yang lebih besar dan potensi loyalitas pengguna yang lebih tinggi.
2. Cluster Kedua mencakup pengguna dengan tingkat ketertarikan awal dan durasi interaksi yang lebih tinggi. Hal ini dapat menunjukkan bahwa pengguna dalam kelompok ini cenderung berinteraksi lebih singkat atau memiliki minat yang lebih rendah terhadap konten yang tersedia, sehingga berpotensi meninggalkan platform dengan cepat.

Korelasi yang terlihat antara BounceRates dan ExitRates mengindikasikan bahwa pengguna yang menunjukkan minat awal yang rendah juga cenderung meninggalkan platform lebih cepat. Informasi ini penting untuk merancang strategi yang lebih tepat sasaran, seperti peningkatan pengalaman pengguna atau kampanye pemasaran yang difokuskan untuk meningkatkan keterlibatan di kedua cluster. Dengan memahami pola interaksi ini, pengelola platform media sosial dapat menyusun kebijakan yang lebih efektif untuk mengoptimalkan retensi pengguna dan meningkatkan tingkat keterlibatan secara keseluruhan.

UCAPAN TERIMA KASIH

Puji dan Syukur saya panjatkan kehadirat Allah SWT. yang telah melimpahkan taufiq dan hidayah-nya kepada kami, sehingga saya dapat menyelesaikan jurnal ini. Terima kasih yang sebesar – besarnya kepada dosen pembimbing yang telah mengarahkan saya dengan penuh kesabaran. Dan juga terima kasih untuk segenap dukungan, khususnya kepada teman – teman yang selalu mengingatkan saya akan terselesaiannya jurnal ini. Tanpa adanya semua dukungan ini, sulit bagi saya untuk menyelesaikan jurnal ini.

DAFTAR PUSTAKA

- Abdul Rohman, M. R. (2020). No Title. *Implementasi Algoritma K-Means Untuk Clustering Kepuasan Mahasiswa Terhadap Pelayanan Akademik.*
- Almoqbel, M. I., & Moradi, F. (2022). No Title. *User Clustering in Social Networks Using K-Means and DBSCAN.*
- Amril Mutoi Siregar, S.Kom., M.Kom. DAN Adam Puspabhuana, S.Kom., M. K. (2021). *DATA MINING Pengolahan Data Menjadi Informasi dengan RapidMiner.* CV Kekata Group.
- Arrohman, S., & Fatah, Z. (2024). *Gudang Jurnal Multidisiplin Ilmu Prediksi Diabetes Menggunakan Algoritma Klasifikasi K-Nearest Neighbors (K-NN) pada Perempuan Indian Pima.* 2, 220–226.
- Gani, A. G. (2020). Pengaruh Media Sosial Terhadap Perkembangan Anak Remaja. *Jurnal Mitra Manajemen,* 7(2), 32–42.
- Kamila, C. (2021). Systematic Literature Review: Penggunaan Algoritma K-Means Untuk Clustering di Indonesia dalam Bidang Pendidikan. *Intech,* 2(1), 19–24. <https://doi.org/10.54895/intech.v2i1.866>
- Kanavos, A., Kiourt, C., Sakas, D. P., & Giannakopoulos, G. (2023). No Title. *Exploring Clustering Techniques for Analyzing User Engagement Patterns in Twitter Data.*
- Kumar, A., & Singh, S. (2021). No Title. *Social Media Data Analytics Using K-Means Clustering.*
- Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, & Susanti. (2022). Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest. *SATIN - Sains Dan Teknologi*

- Informasi*, 8(2), 122–131. <https://doi.org/10.33372/stn.v8i2.885>
- Sukarna, R. H., & Ansori, Y. (2022). Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Ilmiah Sains Dan Teknologi*, 6(1), 50–61. <https://doi.org/10.47080/saintek.v6i1.1467>
- Widaningrum, I., Mustikasari, D., Arifin, R., Tsaqila, S. L., & Fatmawati, D. (2022). Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen. *Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (SISFOTEK)*, 145–149.
- Yogianto, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1720–1728. <https://doi.org/10.33379/gtech.v8i3.4495>
- Zulfikar, Z., Podungge, E. S., Saleh, M. I., & ... (2022). Penerapan Data Mining Untuk Memprediksi Tingkat Kelulusan Siswa Menggunakan Algoritma Neural Network. *Jurnal Elektronik Sistem* ..., 5(1), 7–13. <http://jesik.web.id/index.php/jesik/article/view/91>
- Zulyan Pratama, I., Aziz, F., & Romadhan, B. K. (2024). Pengelompokan Berita Berdasarkan Kategori Like Action Pada Media Sosial Instagram Penamuda Menggunakan Algoritma K-means. *Jurnal Riset Informatika Dan Teknologi Informasi*, 1(2), 57–60. <https://doi.org/10.58776/jriti.v1i2.66>