



ALGORITMA K-MEANS UNTUK PENGELOMPOKAN PERILAKU CUSTOMER

Harun Mukhtar^{1*}, Ilham Dwi Pramaditya²⁾, Wahyu Saputra Weisdiyanto³⁾, Saddam Hardian Putra⁴⁾,
Diana Trimuawasih⁵⁾, Azzahra Auralia Rilda⁶⁾

^{1,2,3,4,5,6}Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau

harunmukhtar@umri.ac.id^{1*}, 220401088@student.umri.ac.id², 220401112@student.umri.ac.id³,
220401105@student.umri.ac.id⁴, 220401096@student.umri.ac.id⁵, 220401107@student.umri.ac.id⁶

Abstract

In the rapidly evolving digital era, understanding customer purchasing behavior is crucial for marketing strategies and business development. This study uses the K-means clustering algorithm to analyze and segment customer purchasing behavior. This algorithm effectively partitions data into groups based on similar characteristics. The aim of this study is to identify purchasing behavior patterns using attributes such as purchase frequency, expenditure amount, and product types. By segmenting customers into homogeneous groups, companies can design more effective marketing strategies and better personalization. The results show that the K-means clustering method successfully segments customers based on similar behavior patterns, which can be used for market segmentation and strategy development. The application of this algorithm in purchasing behavior analysis is expected to provide deep insights and support better business decision-making, offering a competitive advantage for companies.

Keywords: *K-Means, Customer Behavior Analyst, Big Data Analytics in Marketing, Customer Segmentation, Data Mining in Marketing*

Abstrak

Di era digital yang berkembang pesat, pemahaman perilaku pembelian pelanggan menjadi sangat penting untuk strategi pemasaran dan pengembangan bisnis. Penelitian ini menggunakan algoritma K-means clustering untuk menganalisis dan mengelompokkan perilaku pembelian pelanggan. Algoritma ini efektif dalam membagi data ke dalam kelompok berdasarkan karakteristik serupa. Tujuan penelitian ini adalah mengidentifikasi pola perilaku pembelian menggunakan atribut seperti frekuensi pembelian, jumlah pembelanjaan, dan jenis produk. Dengan mengelompokkan pelanggan ke dalam segmen homogen, perusahaan dapat merancang strategi pemasaran yang lebih efektif dan personalisasi yang lebih baik. Hasil penelitian menunjukkan bahwa metode K-means clustering berhasil mengelompokkan pelanggan dengan pola perilaku serupa, yang dapat digunakan untuk segmentasi pasar dan pengembangan strategi. Penerapan algoritma ini dalam analisis perilaku pembelian diharapkan memberikan wawasan mendalam dan mendukung pengambilan keputusan bisnis yang lebih baik, serta menawarkan keuntungan kompetitif bagi perusahaan.

Keywords: *K-Means, Customer Behavior Analyst, Big Data Analytics in Marketing, Customer Segmentation, Data Mining in Marke*

PENDAHULUAN

Dalam era digital yang berkembang pesat, pemahaman tentang perilaku pembelian pelanggan menjadi krusial bagi keberhasilan strategi pemasaran dan pengembangan bisnis. Perilaku pembelian mencerminkan keputusan dan tindakan konsumen saat mereka mencari, memilih, membeli, menggunakan, dan mengevaluasi produk atau layanan. Untuk memaksimalkan keuntungan dan

kepuasan pelanggan, perusahaan perlu mengidentifikasi pola-pola perilaku ini dan mengelompokkan pelanggan ke dalam segmen-segmen yang berbeda berdasarkan karakteristik yang serupa, salah satu metode yang efektif untuk melakukan segmentasi ini adalah K-means clustering (Clayman, Srinivasan, & Sangwan, 2020).

K-means clustering adalah algoritma pembelajaran mesin yang digunakan untuk mempartisi data ke dalam K kelompok atau kluster

fitur-fitur yang relevan. K-means clustering merupakan teknik untuk membagi n data poin dalam ruang d -dimensi R^d ke dalam K kluster, dengan tujuan untuk meminimalkan jarak kuadrat rata-rata dari setiap data point ke pusat kluster terdekatnya (Kanungo et al., n.d.). Dalam konteks analisis perilaku pembelian, fitur-fitur ini dapat mencakup frekuensi pembelian, jumlah pembelanjaan, jenis produk yang dibeli, dan sebagainya. Dengan mengelompokkan pelanggan ke dalam segmen-segmen yang homogen, perusahaan dapat merancang strategi pemasaran yang lebih efektif dan personalisasi yang lebih baik (Guo et al., 2021).

Efektivitas K-means clustering dalam segmentasi pelanggan dan analisis perilaku pembelian. Menurut sebuah studi oleh Gustriansyah (2019), penerapan k-means dalam konteks segmentasi dan analisis perilaku pembelian, K-means diterapkan pada analisis RFM (Recency, Frequency, dan Monetary), RFM adalah metode yang mengukur perilaku pembelian pelanggan berdasarkan tiga parameter: Kapan terakhir kali pelanggan membeli (Recency), seberapa sering pelanggan membeli (Frequency), dan berapa banyak uang dihabiskan (Monetary) (Gustriansyah, Suhandi, & Antony, 2019).

Sejumlah penelitian telah dilakukan dalam konteks ini. (Xie et al., 2019) penelitian ini relevan dengan analisis perilaku pembelian customer dan segmentasi customer karena menggunakan metode clustering yang lebih canggih untuk mengidentifikasi pola-pola tersembunyi dalam data (Xie et al., 2019). Selanjutnya, penelitian (Boone, Ganeshan, Jain, & Sanders, 2019) membahas analisis perilaku pembelian customer dan customer segmentation menggunakan K-means adalah sebuah studi yang mengevaluasi bagaimana data besar dan analitik dapat meningkatkan pemahaman perilaku konsumen serta memperbaiki prediksi permintaan dalam rantai pasokan (Boone, Ganeshan, Jain, & Sanders, 2019).

Namun, meskipun telah banyak penelitian yang dilakukan, masih terdapat gap dalam literatur terkait penerapan K-means Clustering pada data pembelian customer di berbagai industri. Penelitian ini bertujuan untuk mengisi gap tersebut dengan mengeksplorasi penerapan metode K-means dalam analisis perilaku pembelian customer, serta mengidentifikasi segmentasi pelanggan yang relevan untuk strategi pemasaran yang lebih efektif.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap literatur pemasaran dan analisis data, serta menawarkan wawasan praktis bagi perusahaan dalam memahami dan memanfaatkan perilaku pembelian customer untuk keunggulan kompetitif.

K-Means adalah salah satu metode clustering yang paling populer dan banyak digunakan dalam analisis data. Algoritma ini bekerja dengan membagi data ke dalam beberapa kelompok berdasarkan

kesamaan karakteristik. K-means dapat membantu bisnis untuk memahami segmen pelanggan lebih baik, sehingga memungkinkan mereka untuk merancang strategi pemasaran yang lebih efektif dan meningkatkan pengalaman pelanggan (Sarasvananda, Wardoyo, & Sari, 2019).

K-means bekerja dengan cara menginisialisasi sejumlah centroid (pusat cluster) dan kemudian mengelompokkan data berdasarkan jarak terdekat ke centroid tersebut. Proses ini berulang hingga tidak ada perubahan signifikan dalam posisi centroid. Meskipun metode ini sederhana dan cukup cepat, ada beberapa tantangan yang dihadapi penggunaannya. Seperti pemilihan jumlah cluster yang optimal dan sensitivitas terhadap skala data (Sharma & Bala, 2020).

Dalam studi oleh (Dinesh et al., 2019), diusulkan pendekatan clustering data terdistribusi otomatis yang secara otomatis memilih jumlah cluster. Algoritma ini menggunakan penugasan cluster berbasis simetri titik yang memungkinkan pendeteksian cluster berbentuk arbitrer. Tidak memerlukan stasiun basis pusat, sifat terdistribusi algoritma ini membuat penggunaan daya menjadi lebih merata di antara node, sehingga meningkatkan masa pakai jaringan sensor. Kinerja algoritma yang diusulkan divalidasi menggunakan metode evaluasi kualitas cluster seperti Minkowski Score, Silhouette Index, dan Dunn Index. Empat dataset buatan dan satu dataset praktis digunakan untuk studi simulasi yang menunjukkan kinerja yang lebih baik dibandingkan pendekatan K-Means clustering yang ada (Kotary & Nanda, 2019).

Studi lain (Ismail et al., 2013), diusulkan metode baru untuk clustering K-Means menggunakan metode standarisasi untuk menghasilkan cluster berkualitas optimal. Eksperimen komprehensif pada dataset penyakit menular dilakukan untuk mempelajari dampak standarisasi dan membandingkan efek tiga prosedur standarisasi yang berbeda dalam algoritma K-Means konvensional. Hasilnya menunjukkan bahwa standarisasi sebelum algoritma clustering menghasilkan hasil cluster yang lebih baik, efisien, dan akurat. Penting juga untuk memilih prosedur standarisasi tertentu sesuai dengan sifat dataset untuk analisis (Mohamad & Usman, 2013).

Dengan demikian, implementasi algoritma K-Means yang ditingkatkan melalui hal-hal tersebut dapat membantu dalam mengelompokkan perilaku pelanggan dengan lebih efektif, memberikan wawasan yang lebih dalam, dan mendukung strategi bisnis yang lebih baik.

METODOLOGI PENELITIAN

Algoritma K-means adalah salah satu Teknik populer dalam analisis kluster yang digunakan untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan kesamaan karakteristik. Dalam konteks perilaku pelanggan, K-

means dapat membantu mengidentifikasi kelompok pelanggan dengan pola pembelian yang serupa, yang kemudian dapat digunakan untuk segmentasi pasar, pengembangan strategi pemasaran, dan personalisasi layanan (Patel & Kushwaha, 2020).

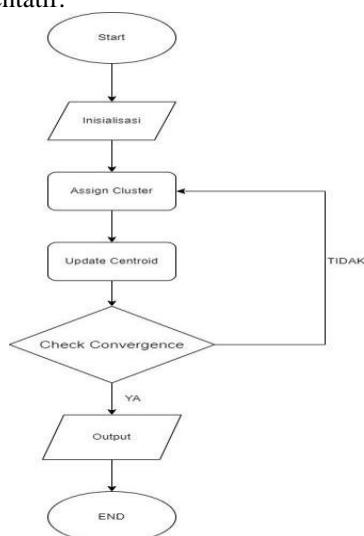
Langkah pertama dalam implementasi algoritma K-Means adalah inisialisasi, di mana jumlah kluster k yang diinginkan harus ditentukan. Selanjutnya, pilih secara acak k titik pusat kluster awal dari dataset yang ada.

Langkah kedua adalah proses pengelompokan, di mana jarak antara setiap titik data dengan masing-masing centroid dihitung menggunakan metrik jarak yang sesuai, seperti jarak Euclidean. Setiap titik data kemudian dikelompokkan ke kluster dengan centroid terdekat.

Setelah semua titik data dikelompokkan, langkah ketiga adalah menghitung ulang posisi centroid. Hal ini dilakukan dengan cara menghitung rata-rata posisi semua titik data yang ada dalam masing-masing kluster, sehingga memperoleh posisi centroid yang baru.

Langkah keempat adalah iterasi, di mana langkah pengelompokan dan perhitungan ulang centroid diulang terus menerus hingga posisi centroid tidak berubah secara signifikan atau telah mencapai jumlah iterasi maksimum yang telah ditentukan sebelumnya.

Langkah terakhir adalah evaluasi hasil klusterisasi. Evaluasi ini dapat dilakukan dengan menggunakan metrik seperti inertia, yang merupakan jumlah jarak kuadrat dari setiap titik data ke centroid terdekat, atau menggunakan silhouette score untuk menilai kualitas kluster yang dihasilkan. Evaluasi ini penting untuk memastikan bahwa kluster yang terbentuk memiliki kualitas yang baik dan representatif.



Gambar 1. Flowchart

Algoritma K-means adalah suatu metode klasifikasi data yang berbasis pada teknik clustering. Algoritma ini bertujuan untuk mengelompokkan data ke dalam beberapa cluster yang sesuai berdasarkan centroid yang telah ditentukan secara awal. Langkah-langkah algoritma K-means dapat dijelaskan sebagai berikut:

Pertama, inisialisasi dilakukan dengan menentukan jumlah cluster k dan memilih k centroid secara acak dari dataset yang diberikan. Pemilihan centroid secara acak ini bertujuan untuk memulai proses klasifikasi dengan kondisi awal yang berbeda-beda.

Kedua, langkah assign cluster dilakukan dengan menghitung jarak antara setiap data point dengan masing-masing centroid menggunakan metrik jarak, seperti jarak Euclidean. Setiap data point kemudian diassign ke centroid terdekat berdasarkan jarak yang telah dihitung.

Ketiga, langkah update centroid dilakukan dengan menghitung ulang posisi centroid berdasarkan rata-rata posisi data point dalam cluster tersebut. Rata-rata posisi data point dihitung untuk setiap cluster dan digunakan sebagai posisi baru dari centroid.

Keempat, langkah check convergence dilakukan untuk memeriksa apakah centroid baru sama dengan centroid sebelumnya atau perubahan dalam posisi centroid kurang dari ambang batas tertentu. Jika centroid baru sama dengan centroid sebelumnya, maka algoritma berhenti. Jika tidak, maka kembali ke langkah assign cluster untuk iterasi berikutnya. Terakhir, langkah output dilakukan untuk menghasilkan cluster yang terbentuk beserta centroidnya. Hasil ini dapat digunakan untuk berbagai aplikasi seperti analisis data, visualisasi data, dan lain-lain.

```

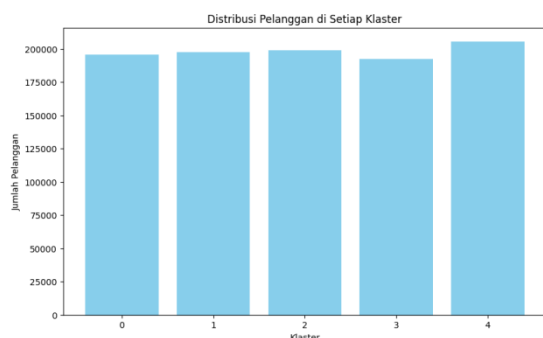
def k_means(data, k, max_iters):
    # Inisialisasi centroid secara acak
    centroids =
    initialize_random_centroids(data, k)
    for i in range(max_iters):
        # Langkah pengelompokan
        clusters = assign_clusters(data,
        centroids)
        # Menghitung ulang centroid
        new_centroids =
        compute_new_centroids(clusters)
        # Memeriksa konvergensi
        if has_converged(centroids,
        new_centroids):
            break
  
```

Gambar 2. Pseudocode Algoritma K-Means

HASIL

Algoritma K-Means telah diterapkan untuk mengelompokkan perilaku pelanggan berdasarkan atribut 'Customer_Name', 'Product', 'City', 'Store_Type', dan 'Customer_Category'. Dataset yang digunakan dianalisis dan dicluster menggunakan metode K-Means dengan jumlah Cluster yang diinginkan sebanyak lima cluster.

- a. Prasekolah dan Normalisasi Data
Data mentah yang terdiri dari atribut-atribut kategorikal diubah menjadi bentuk numerik menggunakan teknik encoding, yaitu 'LabelEncoder'. Proses ini penting untuk memastikan bahwa algoritma K-Means dapat bekerja dengan baik karena algoritma ini membutuhkan input data dalam bentuk numerik. Selanjutnya, data yang telah diencode dinormalisasi menggunakan 'StandardScaler' untuk memastikan bahwa semua fitur memiliki skala yang sama, sehingga jarak antar titik data dapat dihitung dengan akurat (Sinaga & Yang, 2020).
- b. Implementasi Algoritma K-Means
Algoritma K-Means diterapkan pada data yang telah diproses. Jumlah kluster (k) yang dipilih adalah lima, yang ditentukan berdasarkan percobaan dan evaluasi terhadap variasi jumlah kluster yang berbeda. Setelah beberapa iterasi, algoritma berhasil mengelompokkan pelanggan ke dalam lima kluster yang berbeda (Makwana, Kodinariya, & Makwana, 2013).
- c. Distribusi Pelanggan dalam setiap kluster
Hasil klusterisasi divisualisasikan dalam bentuk diagram batang, yang menunjukkan distribusi jumlah pelanggan di setiap kluster. Gambar dibawah ini menunjukkan bahwa setiap kluster memiliki jumlah pelanggan yang relatif seimbang dengan kluster 4 memiliki jumlah pelanggan tertinggi, diikuti oleh kluster 0,1,2, dan 3 (Sarasvananda, Wardoyo, & Sari, 2019).



Gambar 3. Diagram Hasil Analisis Data Customer

- d. Analisis Kluster
Berikut adalah deskripsi dari masing-masing kluster berdasarkan atribut-atribut yang digunakan:
 1. Kluster 0: Pelanggan dalam kluster ini memiliki karakteristik tertentu berdasarkan produk yang dibeli, kota tempat tinggal, jenis toko, dan kategori pelanggan yang serupa.
 2. Kluster 1: Kluster ini terdiri dari pelanggan dengan pola perilaku yang berbeda dari kluster lainnya, yang

dapat diidentifikasi dari kombinasi atribut yang digunakan.

3. Kluster 2: Pelanggan dalam kluster ini menunjukkan perilaku pembelian yang unik yang terungkap dari atribut yang dianalisis.
4. Kluster 3: Kluster ini mengelompokkan pelanggan dengan karakteristik tertentu yang mungkin memiliki preferensi atau kebiasaan belanja yang serupa.
5. Kluster 4: Ini adalah kluster dengan jumlah pelanggan tertinggi yang menunjukkan bahwa pola perilaku dominan di antara pelanggan yang termasuk dalam kluster ini.

K-Means merupakan algoritma clustering yang sangat sederhana dan mudah diimplementasikan, karena algoritma nya yang relatif langsung, terdapat banya pustaka dan framework machine learning yang dapat mendukung K-Means, yang berarti mudah digunakan oleh pemula hingga ahli.

ketika centroid diinisialisasi dengan tepat (contohnya, menggunakan metode K-Means++), K-Means dapat memberikan hasil yang deterministik dan konsisten. ini berbeda dengan beberapa metode lain yang mungkin lebih sensitif terhadap inisialisasi.

HASIL DAN PEMBAHASAN

Hasil klusterisasi perilaku pelanggan menggunakan Algoritma K-Means menunjukkan bahwa metode ini efektif dalam mengelompokkan pelanggan berdasarkan atribut Customer_Name, Product, City, Store_Type, dan Customer_Category. Dalam bagian ini, akan dibahas lebih lanjut implikasi dari hasil yang diperoleh, kekuatan dan keterbatasan dari metode yang digunakan, serta potensi aplikasi dalam strategi bisnis.

a. Implikasi Hasil Klusterisasi

Distribusi yang relatif seimbang di antara lima kluster menunjukkan bahwa tidak ada kluster yang secara signifikan mendominasi. Hal ini menunjukkan bahwa pelanggan memiliki variasi perilaku yang cukup beragam, dan tidak ada satu pola perilaku yang mendominasi seluruh dataset. Analisis lebih lanjut terhadap masing-masing kluster dapat memberikan wawasan yang lebih dalam tentang karakteristik spesifik dari setiap kelompok pelanggan (Oyelade, Oladipupo, & Obagbuwa, 2010).

1. **Kluster 0:** Pelanggan dalam kluster ini mungkin menunjukkan preferensi untuk produk-produk tertentu, atau mungkin berasal dari kota yang sama dengan tipe toko dan kategori pelanggan yang spesifik.

Strategi pemasaran yang ditargetkan pada klaster ini dapat difokuskan pada produk yang populer di kalangan pelanggan ini atau promosi khusus di kota tersebut.

2. **Klaster 1:** Pelanggan dalam klaster ini menunjukkan pola perilaku yang berbeda, mungkin menunjukkan preferensi untuk jenis toko atau kategori produk tertentu. Dengan memahami lebih lanjut karakteristik ini, perusahaan dapat mengembangkan kampanye pemasaran yang lebih tersegmentasi dan personal.
3. **Klaster 2:** Klaster ini mungkin mengindikasikan kelompok pelanggan dengan kebiasaan belanja unik yang tidak umum di klaster lain. Hal ini bisa mencakup preferensi terhadap produk niche atau kebiasaan berbelanja di waktu-waktu tertentu. Pemahaman ini dapat digunakan untuk menciptakan pengalaman belanja yang lebih dipersonalisasi.
4. **Klaster 3:** Klaster ini menunjukkan kelompok pelanggan dengan karakteristik yang mungkin serupa, seperti lokasi geografis atau preferensi toko. Kampanye pemasaran yang fokus pada promosi lokal atau program loyalitas khusus dapat efektif untuk klaster ini.
5. **Klaster 4:** Sebagai klaster dengan jumlah pelanggan tertinggi, klaster ini menunjukkan pola perilaku yang paling umum di antara pelanggan. Strategi pemasaran yang lebih luas namun efektif dapat diterapkan di sini, seperti penawaran promosi yang lebih umum atau program loyalitas yang luas.

b. Kekuatan dan Keterbatasan Metode K-Means

Kekuatan:

1. **Kesederhanaan dan Efisiensi:** Algoritma K-Means mudah diimplementasikan dan dapat memproses data dalam jumlah besar dengan cepat (Clayman, Srinivasan, & Sangwan, 2020).
2. **Skalabilitas:** Metode ini dapat diterapkan pada dataset yang besar, seperti yang terlihat dalam penelitian ini dengan jumlah pelanggan yang signifikan.
3. **Interpretabilitas:** Hasil klasterisasi dapat diinterpretasikan dengan mudah

dan digunakan untuk pengambilan keputusan bisnis.

Keterbatasan:

1. **Penentuan Jumlah Klaster:** Pemilihan jumlah klaster (k) yang tepat bisa menjadi tantangan. Dalam penelitian ini, dipilih lima klaster berdasarkan evaluasi awal, namun pemilihan jumlah klaster yang berbeda mungkin memberikan hasil yang berbeda.
2. **Sensitivitas terhadap Penskalaan:** Algoritma ini sensitif terhadap penskalaan data, sehingga normalisasi data sangat penting untuk mendapatkan hasil yang akurat (Shirazi & Mohammadi, 2019).
3. **Ketergantungan pada Inisialisasi Awal:** Hasil klasterisasi dapat bervariasi tergantung pada pemilihan centroid awal, sehingga pemilihan inisialisasi yang baik adalah penting.

c. Potensi Aplikasi dalam Strategi Bisnis

Hasil klasterisasi dapat digunakan untuk berbagai aplikasi dalam strategi bisnis, antara lain:

1. **Segmentasi Pasar:** Mengidentifikasi segmen pelanggan dengan karakteristik yang serupa untuk pengembangan produk dan strategi pemasaran yang lebih efektif (Wu et al., 2020).
2. **Personalisasi Layanan:** Menyediakan pengalaman belanja yang lebih dipersonalisasi berdasarkan klaster perilaku pelanggan (Park, Kang, Choi, & Han, 2020).
3. **Pengembangan Produk:** Mengidentifikasi kebutuhan dan preferensi pelanggan di setiap klaster untuk pengembangan produk baru atau penyesuaian produk yang ada.
4. **Manajemen Hubungan Pelanggan:** Menerapkan program loyalitas dan promosi yang lebih terarah untuk meningkatkan kepuasan dan retensi pelanggan (Cheng, Wu, & Chen, 2019).

SIMPULAN DAN SARAN

Penerapan Algoritma K-Means dalam mengkluster perilaku pelanggan memberikan wawasan berharga tentang variasi pola pembelian di antara pelanggan. Hasil ini dapat digunakan untuk mengembangkan strategi bisnis yang lebih terarah dan efektif, serta meningkatkan pengalaman pelanggan melalui personalisasi layanan dan

penawaran. Meskipun terdapat beberapa keterbatasan, metode ini tetap menjadi alat yang kuat dalam analisis data pelanggan untuk keperluan bisnis.

TERIMA KASIH

Ucapan terima kasih disampaikan kepada dosen pengampu mata kuliah Sains Data Assoc. Prof. Harun Mukhtar, S.Kom., M.Kom, atas peran sertanya dalam memberikan masukan, melakukan telaah, koreksi, dan perbaikan naskah sampai siap diterbitkan.

DAFTAR PUSTAKA

- Clayman, C. L., Srinivasan, S. M., & Sangwan, R. S. (2020). K-means clustering and principal components analysis of microarray data of L1000 landmark genes. *Procedia Computer Science*, 97-104. <https://doi.org/10.1016/j.procs.2020.02.265>
- Kanungo, T., et al. (n.d.). An Efficient k-Means Clustering Algorithm: Analysis and Implementation.
- Guo, Z., Shi, Y., Huang, F., Fan, X., & Huang, J. (2021). Landslide susceptibility zonation method based on C5.0 decision tree and K-means cluster algorithms to improve the efficiency of risk management. *Geoscience Frontiers*, 12(6). <https://doi.org/10.1016/j.gsf.2021.101249>
- Gustriansyah, R., Suhandi, N., & Antony, F. (2019). Clustering optimization in RFM analysis based on k-means. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(1), 470-477. <https://doi.org/10.11591/ijeecs.v18.i1.pp470-477>
- Xie, H., et al. (2019). Improving K-means clustering with enhanced Firefly Algorithms. *Applied Soft Computing Journal*, 84, 105763. <https://doi.org/10.1016/j.asoc.2019.105763>
- Xie, H., et al. (2019). Improving K-means clustering with enhanced Firefly Algorithms. *Applied Soft Computing Journal*, 84. <https://doi.org/10.1016/j.asoc.2019.105763>
- Boone, T., Ganeshan, R., Jain, A., & Sanders, N. R. (2019). Forecasting sales in the supply chain: Consumer analytics in the big data era. *International Journal of Forecasting*, 35(1), 170-180. <https://doi.org/10.1016/j.ijforecast.2018.09.003>
- Boone, T., Ganeshan, R., Jain, A., & Sanders, N. R. (2019). Forecasting sales in the supply chain: Consumer analytics in the big data era. *International Journal of Forecasting*, 35(1), 170-180. <https://doi.org/10.1016/j.ijforecast.2018.09.003>
- Sarasvananda, I. B. G., Wardoyo, R., & Sari, A. K. (2019). The K-Means Clustering Algorithm With Semantic Similarity To Estimate The Cost of Hospitalization. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(4), 313. <https://doi.org/10.22146/ijccs.45093>
- Sharma, V., & Bala, M. (2020). An Improved Task Allocation Strategy in Cloud using Modified K-means Clustering Technique. *Egyptian Informatics Journal*, 21(4), 201-208. <https://doi.org/10.1016/j.eij.2020.02.001>
- Kotary, D. K., & Nanda, S. J. (2019). Automatic Determination of K in Distributed K-Means Clustering. *Procedia Computer Science*, 556-564. <https://doi.org/10.1016/j.procs.2020.01.050>
- Mohamad, I. B., & Usman, D. (2013). Standardization and its effects on K-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6(17), 3299-3303. <https://doi.org/10.19026/rjaset.6.3638>
- Patel, E., & Kushwaha, D. S. (2020). Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. *Procedia Computer Science*, 158-167. <https://doi.org/10.1016/j.procs.2020.04.017>
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716-80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Makwana, P., Kodinariya, T. M., & Makwana, P. R. (2013). Review on Determining of Cluster in K-means Clustering Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6). Retrieved from <https://www.researchgate.net/publication/313554124>
- Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). Application of k-Means Clustering algorithm for prediction of Students' Academic Performance. Retrieved from <http://sites.google.com/site/ijcsis/>
- Shirazi, F., & Mohammadi, M. (2019). A big data analytics model for customer churn prediction in the retiree segment. *International Journal of Information Management*, 48, 238-253. <https://doi.org/10.1016/j.ijinfomgt.2018.10.005>
- Wu, J., et al. (2020). An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm. *Mathematical Problems in Engineering*, 2020. <https://doi.org/10.1155/2020/8884227>
- Park, E., Kang, J., Choi, D., & Han, J. (2020). Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. *Current Issues in Tourism*, 23(5), 605-611. <https://doi.org/10.1080/13683500.2018.1549025>
- Cheng, L. C., Wu, C. C., & Chen, C. Y. (2019). Behavior analysis of customer churn for a customer relationship system: An empirical case study. *Journal of Global Information Management*, 27(1), 111-127. <https://doi.org/10.4018/JGIM.2019010106>