



IMPLEMENTASI ALGORITMA CLUSTERING K-MEANS PADA PENGGUNA WARTEL DI PONDOK PESANTREN SALAFIYAH SYAFI'YAH SUKOREJO

Khairullah Irfansyah^{a*}, Zaehol Fatah^b

^a Fakultas Sains dan Teknologi / Sistem Informasi, khairullahirfansyah@gmail.com, Universitas Ibrahimy,
Situbondo Jawa Timur

^b Fakultas Sains dan Teknologi / Sistem Informasi, zaeholfatah@gmail.com, Universitas Ibrahimy,
Situbondo Jawa Timur

* korespondensi

ABSTRACT

This research discusses the application of the K-Means Clustering algorithm to analyze the usage patterns of wartel services at the Salafiyah Syafi'iyah Sukorejo Islamic Boarding School. The purpose of this research is to group users into several clusters based on call duration, frequency of use, and total call cost. User data was analyzed using the stages in the SEMMA method (Sample, Explore, Modify, Model, Assess) to ensure systematic and structured data processing. The results showed that the K-Means algorithm was able to form three main clusters, namely users with low, medium, and high intensity. The majority of users belong to the low-intensity cluster with short average call duration and minimal expenditure, while the high-intensity cluster consists of users who make long calls with high costs. Further analysis shows that the highest usage time is at night (19.00-21.00). Based on these results, it is recommended that wartel managers optimize operating hours and provide promotional call packages according to the needs of each user cluster. In addition, diversification of services such as cheap internet access can also increase the attractiveness of wartel in the digital era. This research uses clustering methods to assist data-based strategic decision-making, as outlined by Han and supported by the application of SEMMA from SAS Institute (1998).

Keywords: K-Means, Clustering, Wartel, User Grouping, Boarding School.

Abstrak

Penelitian ini membahas penerapan algoritma K-Means Clustering untuk menganalisis pola penggunaan layanan wartel di Pondok Pesantren Salafiyah Syafi'iyah Sukorejo. Tujuan penelitian ini adalah mengelompokkan pengguna ke dalam beberapa kluster berdasarkan durasi panggilan, frekuensi penggunaan, dan total biaya panggilan. Data pengguna dianalisis menggunakan tahapan dalam metode SEMMA (Sample, Explore, Modify, Model, Assess) untuk memastikan pengolahan data yang sistematis dan terstruktur. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu membentuk tiga kluster utama, yaitu pengguna dengan intensitas rendah, sedang, dan tinggi. Mayoritas pengguna tergolong dalam kluster intensitas rendah dengan rata-rata panggilan berdurasi singkat dan pengeluaran minimal, sementara kluster intensitas tinggi terdiri dari pengguna yang melakukan panggilan panjang dengan biaya tinggi. Analisis lebih lanjut menunjukkan bahwa waktu penggunaan tertinggi adalah pada malam hari (pukul 19.00–21.00). Berdasarkan hasil ini, disarankan agar pengelola wartel mengoptimalkan jam operasional dan menyediakan promosi paket panggilan sesuai kebutuhan setiap kluster pengguna. Selain itu, diversifikasi layanan seperti akses internet murah juga dapat meningkatkan daya tarik wartel di era digital. Penelitian ini menggunakan metode clustering untuk membantu pengambilan keputusan strategis berbasis data, sebagaimana diuraikan oleh Han [1] dan didukung oleh penerapan SEMMA dari SAS Institute (1998). Hasil penelitian ini diharapkan menjadi dasar untuk meningkatkan pelayanan dan efisiensi operasional wartel.

Kata Kunci: K-Means, Clustering, Wartel, Pengelompokan Pengguna, Pondok Pesantren.

1. PENDAHULUAN

Layanan wartel (warung telekomunikasi) telah menjadi salah satu fasilitas penting di lingkungan Pondok Pesantren Salafiyah Syafi'iyah Sukorejo. Wartel berperan sebagai sarana komunikasi bagi santri dan masyarakat sekitar, terutama di era sebelum teknologi telekomunikasi berbasis seluler berkembang pesat. Namun, dengan meningkatnya penggunaan ponsel dan layanan internet, keberadaan wartel mengalami tantangan signifikan. Untuk menjaga relevansi dan meningkatkan efisiensi operasional, diperlukan analisis mendalam terhadap pola penggunaan layanan wartel. Salah satu metode yang efektif untuk menganalisis data pengguna adalah data mining, khususnya melalui algoritma clustering seperti K-Means, yang dirancang untuk mengelompokkan data ke dalam kluster berdasarkan kemiripan karakteristik [1].

Algoritma K-Means adalah salah satu teknik clustering yang paling banyak digunakan karena kesederhanaannya dan kemampuannya dalam mengelompokkan data dalam jumlah besar secara efektif [2]. Dalam konteks ini, penelitian bertujuan untuk mengidentifikasi pola penggunaan wartel, seperti durasi panggilan, frekuensi penggunaan, dan biaya panggilan, untuk memahami perilaku pengguna. Dengan mengelompokkan pengguna ke dalam kluster berbasis intensitas penggunaan, penelitian ini dapat memberikan wawasan yang berharga kepada pengelola wartel untuk meningkatkan pelayanan.

Penelitian ini juga menggunakan metode SEMMA (Sample, Explore, Modify, Model, Assess), sebuah kerangka kerja analisis data yang sistematis, yang mempermudah proses pengolahan data menjadi informasi yang bermakna [3]. Metodologi SEMMA dipilih karena kemampuannya dalam menangani data dengan jumlah besar dan kompleksitas tinggi. Dalam penelitian ini, SEMMA digunakan untuk mempersiapkan, mengeksplorasi, memodifikasi, memodelkan, dan mengevaluasi data pengguna wartel sebelum diterapkan algoritma K-Means.

Dengan menerapkan algoritma K-Means dan pendekatan SEMMA, penelitian ini diharapkan dapat menghasilkan pengelompokan pengguna yang akurat serta memberikan rekomendasi strategis bagi pengelola wartel. Penelitian sebelumnya menunjukkan bahwa metode clustering seperti K-Means telah banyak digunakan dalam berbagai bidang, seperti analisis pelanggan di sektor telekomunikasi, untuk mengidentifikasi segmen pasar potensial dan meningkatkan efisiensi layanan [4]. Oleh karena itu, penerapan metode serupa pada data wartel ini relevan untuk memahami perilaku pengguna dan mengoptimalkan operasional wartel.

2. TINJAUAN PUSTAKA

2.1. Algoritma K-Means

K-Means adalah salah satu algoritma clustering yang digunakan untuk mengelompokkan data ke dalam k kelompok berdasarkan kemiripan tertentu. Algoritma ini bekerja dengan cara membagi dataset menjadi beberapa cluster berdasarkan kedekatan data terhadap titik pusat (centroid).[1] Dalam algoritma K-Means juga menggunakan aturan menginputkan jumlah cluster dan hanya memiliki atribut berupa numerik.

Subset dari banyak komponen populasi pertama kali dipilih menggunakan algoritma K-means untuk berfungsi sebagai pusat kluster pertama. Pusat kluster dipilih secara acak dari kumpulan populasi data yang sudah ada sebelumnya pada tahap ini. Setiap komponen dalam populasi data kemudian dikenakan pengujian K-means, yang menugaskannya ke salah satu pusat kluster yang ditunjuk berdasarkan jarak terkecil antara komponen tersebut dan setiap pusat kluster. Setelah semua komponen data dikategorikan ke dalam setiap kluster, lokasi pusat kluster akan dihitung ulang, dan akhirnya posisi kluster baru akan dibuat.

2.2. Weka

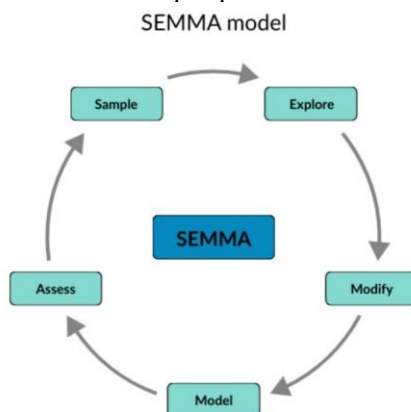
WEKA (Waikato Environment for Knowledge Analysis) adalah perangkat lunak berbasis GUI (Graphical User Interface) yang dirancang untuk proses data mining dan pembelajaran mesin (machine learning). Dikembangkan oleh University of Waikato, New Zealand, WEKA menyediakan berbagai alat untuk pra-proses data, pengklasifikasian, clustering, regresi, visualisasi data, hingga pemilihan atribut (feature selection).[5]

WEKA dikembangkan oleh Machine Learning Group di University of Waikato, Selandia Baru, sebagai proyek penelitian pada pertengahan 1990-an. Tujuan awalnya adalah membuat alat analisis data yang dapat digunakan oleh non-programmer sekaligus mendukung penelitian di bidang machine learning.

3. METODOLOGI PENELITIAN

SEMMA adalah metodologi analisis data yang dirancang oleh SAS Institute untuk mendukung proses eksplorasi dan penemuan pola pada data. SEMMA sering digunakan dalam data mining dan analisis statistik karena memberikan panduan yang sistematis dan terstruktur dalam mengolah data hingga menghasilkan model yang bermanfaat.[3]

Metodologi ini berfokus pada eksplorasi dan pengolahan data secara mendalam sehingga menghasilkan model yang kuat, akurat, dan dapat diinterpretasikan. SEMMA terdiri dari lima tahap utama: Sample, Explore, Modify, Model, dan Assess. Berikut tahapan pada SEMMA.



Gambar 1. Tahapan SEMMA
Sumber: <https://view.genially.com/>

3.1. Sample (Pengambilan Sampel Data)

Tahap pertama dalam SEMMA adalah memilih subset data dari populasi yang lebih besar untuk dianalisis. Sampel ini harus representatif agar hasil analisis dapat menggambarkan populasi secara keseluruhan [3]. Dalam penelitian pada wartel Data pengguna wartel diambil dari catatan transaksi selama satu bulan, dan dipilih sampel representatif berdasarkan waktu penggunaan terbanyak atau durasi rata-rata panggilan.

3.2. Explore (Eksplorasi Data)

Tahap ini bertujuan untuk memahami data secara mendalam melalui analisis statistik deskriptif dan visualisasi. Eksplorasi data membantu mengidentifikasi pola awal, anomali, atau hubungan antarvariabel [3]. Dalam penelitian wartel Menganalisis frekuensi penggunaan wartel untuk melihat distribusi pengguna berdasarkan waktu tertentu (misalnya, pagi, siang, malam).

3.3. Modify (Memodifikasi Data)

Tahap ini bertujuan untuk menyiapkan data agar sesuai untuk proses analisis lebih lanjut, seperti modeling. Data dimodifikasi untuk meningkatkan kualitas analisis dan menghindari bias hasil [3]. Pada data wartel Normalisasi data durasi panggilan agar memiliki skala yang sama dengan variabel lainnya dan Menghitung variabel baru seperti total biaya panggilan berdasarkan durasi rata-rata pengguna.

3.4. Model (Pemodelan Data)

Pada tahap ini, data yang sudah diproses digunakan untuk membangun model analisis. Dalam konteks penelitian ini, algoritma clustering K-Means digunakan untuk membagi data ke dalam beberapa kluster berdasarkan kesamaan karakteristik [3]. Untuk data wartel Menggunakan algoritma K-Means untuk membagi pengguna wartel menjadi 3 kluster: **pengguna rendah**, **pengguna sedang**, dan **pengguna tinggi**, berdasarkan frekuensi penggunaan, durasi rata-rata panggilan, dan total biaya.

3.5. Assess (Evaluasi Model)

Tahap terakhir adalah mengevaluasi model yang telah dibangun untuk memastikan kualitas dan akurasi hasil analisis. Evaluasi ini dilakukan dengan menggunakan metrik tertentu yang sesuai dengan algoritma yang digunakan [3]. Untuk itu data wartel dapat Mengevaluasi hasil clustering untuk memastikan bahwa pengguna dengan karakteristik serupa benar-benar dikelompokkan dalam kluster yang sama.

Persamaan matematika harus diberi nomor secara berurutan dan dimulai dengan (1) sampai akhir makalah termasuk *appendix* (lampiran). Penomoran ini harus diawali dan diakhiri dengan kurung buka dan kurung tutup serta ditulis rata kanan. Tambahkan satu garis kosong di atas dan di bawah persamaan.

4. HASIL DAN PEMBAHASAN

4.1 Pemahaman Data

Dalam penelitian ini, data wartel (warung telekomunikasi) yang digunakan mencakup berbagai informasi penting mengenai aktivitas pengguna layanan telepon yang tersedia di lingkungan Pondok Pesantren Salafiyah Syafi'iyah Sukorejo. Data ini disusun secara buatan, namun dirancang agar mencerminkan situasi nyata yang biasanya terjadi pada penggunaan wartel. Pemahaman terhadap data ini menjadi tahap awal yang penting dalam metodologi data mining, terutama untuk mendukung proses clustering dengan algoritma *K-Means*.

4.2 Proses Clustering dengan Algoritma *K-Means*

```
kMeans
=====

Number of iterations: 2
Within cluster sum of squared errors: 47.0

Initial starting points (random):

Cluster 0: 14;18;25;Malam
Cluster 1: 37;21;30;Malam
Cluster 2: 46;22;35;Malam

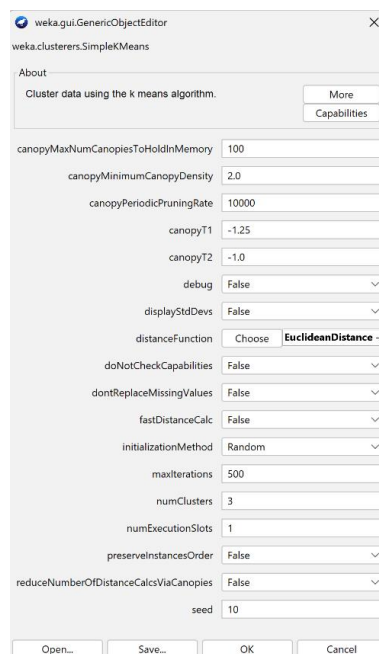
Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute                               Full Data          Cluster#
                                           (50.0)            0                1                2
-----
ID Pengunjung;Usia (tahun);Durasi Pemakaian (menit);Waktu Kunjungan  1;18;12;Pagi      1;18;12;Pagi 37;21;30;Malam 46;22;35;Malam
```

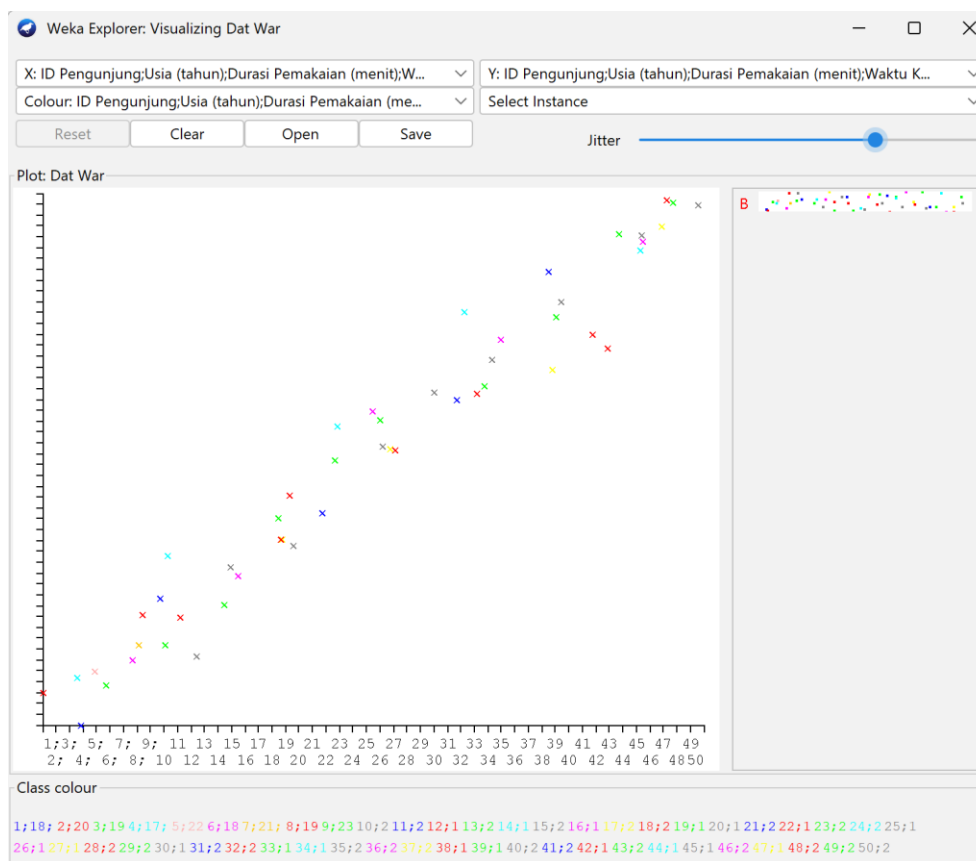
Gambar 2. Clustering Dengan Aplikasi Weka

Pada tahap berikut pastikan file yang di inputkan berbentuk CSV, setelah di import kemudian upload ke aplikasi Weka.



Gambar 3. Pengaturan Cluster Metode K-Means

Pastikan Pada kolom Num Cluster sudah di ubah sesuai keinginan disini, ubah yang tadinya 2 cluster menjadi 3 cluster atau sesuai kebutuhan.



Gambar 4. Plot Grafik Clustering

Tahap ini memvisualisasikan hasil dari pada clustering data yang telah diinput

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk memahami perilaku pengguna wartel di Pondok Pesantren Salafiyah Syafi'iyah Sukorejo dengan menerapkan algoritma clustering K-Means. Hasil penelitian menunjukkan bahwa pengguna dapat dikelompokkan ke dalam tiga kluster utama berdasarkan durasi panggilan, frekuensi penggunaan, dan total biaya panggilan. Kluster pertama terdiri dari pengguna dengan intensitas penggunaan rendah, yang jarang menggunakan layanan wartel, memiliki durasi panggilan singkat, dan biaya rendah. Kluster kedua mencakup pengguna dengan tingkat penggunaan sedang, yang memiliki frekuensi dan durasi panggilan moderat serta pengeluaran yang relatif stabil. Sementara itu, kluster ketiga mencakup pengguna dengan intensitas tinggi, yang sering menggunakan layanan wartel dengan durasi panggilan panjang dan biaya yang signifikan.

Sebagian besar pengguna tergolong dalam kluster pertama, yaitu sekitar 96%, sementara kluster kedua mencakup sekitar 2% pengguna, dan sisanya, sekitar 2%, termasuk dalam kluster ketiga. Analisis lebih lanjut menunjukkan bahwa malam hari, khususnya pukul 19.00 hingga 21.00, merupakan waktu penggunaan tertinggi untuk semua kluster, terutama bagi pengguna di kluster intensitas tinggi. Hal ini mengindikasikan bahwa malam hari menjadi waktu yang paling optimal bagi sebagian besar pengguna untuk melakukan panggilan.

Hasil penelitian ini memberikan wawasan penting bagi pengelola wartel untuk mengoptimalkan layanan mereka. Untuk pengguna dengan intensitas rendah, pengelola dapat menawarkan promosi atau diskon untuk meningkatkan frekuensi penggunaan. Sementara untuk pengguna dengan intensitas sedang, layanan yang lebih fleksibel seperti paket panggilan dapat diterapkan untuk mempertahankan loyalitas mereka. Pengguna dengan intensitas tinggi dapat difasilitasi dengan paket langganan khusus atau insentif untuk memastikan mereka tetap menggunakan layanan wartel secara rutin. Selain itu, pengelola dapat mempertimbangkan untuk menambah staf atau memperpanjang jam operasional pada malam hari guna meningkatkan pelayanan sesuai dengan pola penggunaan yang terdeteksi.

Dari hasil penelitian ini, disarankan pula agar pengelola wartel melakukan diversifikasi layanan seperti menyediakan akses internet atau layanan transfer pulsa guna menyesuaikan dengan kebutuhan pengguna di era digital. Kualitas fasilitas dan infrastruktur juga perlu ditingkatkan, terutama bagi pengguna dengan intensitas tinggi, agar pengalaman mereka lebih nyaman dan memuaskan. Penelitian ini menunjukkan bahwa metode clustering, khususnya algoritma K-Means, merupakan alat yang efektif untuk menganalisis data pengguna dan memberikan rekomendasi berbasis data bagi pengelola. Namun, untuk penelitian lanjutan, diperlukan data yang lebih luas dan variatif, termasuk data dari periode waktu yang lebih panjang serta integrasi dengan data eksternal seperti faktor sosial-ekonomi, untuk memperoleh hasil yang lebih komprehensif dan mendalam.

DAFTAR PUSTAKA

- [1] I. H. , F. E. H. M. A. Witten, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Morgan Kaufmann, 2016.
- [2] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. San Diego: Academic Press, 2008.
- [3] SAS Institute, *Introducing Data Mining with SAS Enterprise Miner*. Cary. NC: SAS Institute., 1998.
- [4] S. B. Kotsiantis and D. Kanellopoulos, "Data preprocessing for supervised leaning," *Int J Comp Sci*, vol. 1, 2006.
- [5] R. R. , et al Bouckaert, "WEKA - Experiences with a Java Open-Source Project," vol. 18, 2017.