

Penerapan Algoritma K-Means untuk Segmentasi Pemilih Berdasarkan Kelompok Usia (Studi Kasus: DPT Desa Kesambi)

Firda Intan Nur Alisa¹, Nanda Dea Anggi Maharani², Canes Nur Amandasari³, Anggita Nur Maulani⁴, Muhammad Arifin⁵

^{1,2,3,4,5} Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus
¹202353134@std.umk.ac.id, ²202353133@std.umk.ac.id, ³202353122@std.umk.ac.id,
⁴202353147@std.umk.ac.id, ⁵arifin.m@umk.ac.id,

Abstract

Voter participation is one of the key indicators of successful electoral administration at the village level. However, uniform socialization approaches that disregard voters' demographic characteristics are often ineffective. This study aims to segment voters by age group using the K-Means Clustering algorithm applied to the Permanent Voter List (DPT) data of Kesambi Village. The variables used in the clustering process are age and gender. A total of 1,465 records were used after preprocessing. The optimal number of clusters was determined using the Elbow Method, which identified $k = 3$ as the best value. The clustering results divided voters into three groups: Young Voters (aged 17–34 years) with 509 voters (34.7%), Productive Age (aged 35–53 years) with 592 voters (40.4%), and Elderly (aged 54–93 years) with 364 voters (24.8%). Evaluation using the Silhouette Score yielded a value of 0.5698, indicating a reasonably good clustering quality. These findings can serve as a basis for decision-making in designing more targeted socialization strategies and campaign approaches tailored to the demographic characteristics of each voter group.

Keywords : K-Means, Clustering, Voter Segmentation, Permanent Voter List, Elbow Method

Abstrak

Partisipasi pemilih merupakan salah satu indikator penting dalam keberhasilan penyelenggaraan pemilu di tingkat desa. Namun, pendekatan sosialisasi yang bersifat seragam tanpa mempertimbangkan karakteristik demografis pemilih seringkali kurang efektif. Penelitian ini bertujuan untuk melakukan segmentasi pemilih berdasarkan kelompok usia menggunakan algoritma K-Means *Clustering* pada data Daftar Pemilih Tetap (DPT) Desa Kesambi. Variabel yang digunakan dalam proses *clustering* adalah usia dan jenis kelamin. Data yang digunakan berjumlah 1.465 rekod setelah melalui tahap *preprocessing*. Penentuan jumlah cluster optimal dilakukan menggunakan *Elbow Method*, yang menghasilkan nilai $k = 3$ sebagai jumlah cluster terbaik. Hasil *clustering* membagi pemilih ke dalam tiga kelompok, yaitu Pemilih Muda (usia 17–34 tahun) sebanyak 509 pemilih (34,7%), Usia Produktif (usia 35–53 tahun) sebanyak 592 pemilih (40,4%), dan Lansia (usia 54–93 tahun) sebanyak 364 pemilih (24,8%). Evaluasi menggunakan *Silhouette Score* menghasilkan nilai 0,5698 yang menunjukkan kualitas *clustering* dalam kategori cukup baik. Hasil penelitian ini dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam merancang strategi sosialisasi dan pendekatan kampanye yang lebih tepat sasaran sesuai dengan karakteristik demografis masing-masing kelompok pemilih.

Kata Kunci : K-Means, Clustering, Segmentasi Pemilih, DPT, Elbow Method

1. Pendahuluan

Daftar Pemilih Tetap (DPT) merupakan komponen krusial dalam penyelenggaraan pemilihan umum karena berfungsi sebagai dasar utama dalam menentukan jumlah serta karakteristik pemilih pada suatu wilayah. Keakuratan dan kelengkapan data DPT sangat berpengaruh terhadap kualitas proses demokrasi, khususnya dalam menjamin hak pilih masyarakat serta meningkatkan tingkat partisipasi pemilih [1]. Selain digunakan untuk kepentingan administratif, data DPT sebenarnya memiliki potensi besar untuk dianalisis lebih lanjut guna menggali informasi strategis, seperti distribusi demografis pemilih berdasarkan usia, jenis kelamin, maupun karakteristik lainnya [2]

Dalam penyelenggaraan pemilihan umum, kualitas data pemilih menjadi salah satu indikator penting dalam menciptakan proses demokrasi yang transparan dan akuntabel. Data pemilih yang tersusun dengan baik tidak hanya mempermudah proses administrasi pemilu, tetapi juga membantu pemerintah maupun penyelenggara pemilu dalam memahami kondisi sosial masyarakat secara lebih komprehensif. Dengan adanya data yang terstruktur, proses perencanaan kebutuhan logistik, distribusi surat suara, hingga pelaksanaan sosialisasi dapat dilakukan secara lebih efektif dan efisien. Oleh karena itu, pengelolaan data pemilih tidak dapat dipandang hanya sebagai aktivitas pencatatan administratif, melainkan juga sebagai sumber informasi strategis yang dapat mendukung proses pengambilan kebijakan.

Selain itu, perkembangan digitalisasi dalam tata kelola pemerintahan telah mendorong berbagai instansi untuk mulai memanfaatkan data sebagai dasar dalam menentukan kebijakan publik. Konsep pengambilan keputusan berbasis data (*data-driven decision making*) semakin banyak diterapkan karena dinilai mampu menghasilkan keputusan yang lebih objektif dan terukur. Dalam konteks data pemilih, penerapan analisis data dapat membantu mengidentifikasi pola tertentu yang sebelumnya sulit diamati secara manual. Informasi tersebut dapat berupa persebaran usia pemilih, dominasi kelompok demografis tertentu, hingga kecenderungan partisipasi masyarakat berdasarkan karakteristik wilayah.

Namun, pada praktiknya pemanfaatan data DPT masih cenderung terbatas pada fungsi administratif semata. Padahal, dengan perkembangan teknologi informasi dan data mining, data tersebut dapat diolah menjadi informasi yang bernilai guna dalam mendukung pengambilan keputusan berbasis data (*data-driven decision making*) [3]. Kurangnya pemanfaatan ini disebabkan oleh belum adanya pendekatan analisis yang sistematis, serta keterbatasan dalam penggunaan metode komputasional untuk mengolah data dalam jumlah besar [4].

Permasalahan kurang optimalnya pemanfaatan

data juga dipengaruhi oleh masih terbatasnya kemampuan dalam mengolah data berukuran besar secara sistematis. Pada umumnya, data DPT hanya disimpan dalam bentuk tabel administrasi tanpa dilakukan analisis lebih lanjut. Padahal, apabila dilakukan pengolahan menggunakan pendekatan komputasional, data tersebut dapat menghasilkan informasi baru yang lebih bernilai. Proses analisis data secara manual tentu membutuhkan waktu yang lama dan memiliki risiko kesalahan yang lebih tinggi, terutama ketika jumlah data terus bertambah setiap periode pemilu. Oleh sebab itu, diperlukan metode yang mampu membantu proses analisis secara otomatis, cepat, dan konsisten.

Segmentasi data pemilih juga menjadi hal yang penting karena karakteristik masyarakat tidak bersifat homogen. Setiap kelompok masyarakat memiliki pola perilaku, tingkat pemahaman politik, serta kebutuhan informasi yang berbeda. Dengan melakukan segmentasi, pihak penyelenggara dapat menentukan pendekatan yang lebih tepat sesuai dengan karakteristik masing-masing kelompok. Misalnya, kelompok pemilih muda lebih mudah dijangkau melalui media sosial dan platform digital, sedangkan kelompok usia lanjut cenderung membutuhkan pendekatan langsung melalui kegiatan sosialisasi tatap muka. Perbedaan karakteristik tersebut menunjukkan bahwa strategi komunikasi yang seragam sering kali kurang efektif dalam meningkatkan partisipasi masyarakat secara menyeluruh.

Permasalahan lain yang sering muncul dalam pengelolaan data pemilih adalah belum adanya proses pengelompokan (*segmentasi*) yang terstruktur. Tanpa adanya segmentasi, sulit bagi pihak penyelenggara pemilu maupun pemerintah desa untuk memahami karakteristik pemilih secara lebih mendalam [5]. Akibatnya, strategi sosialisasi yang dilakukan cenderung bersifat umum dan kurang efektif dalam menjangkau seluruh lapisan masyarakat [6]. Pendekatan yang seragam ini tidak mempertimbangkan perbedaan kebutuhan dan perilaku antar kelompok usia, sehingga berdampak pada rendahnya efektivitas komunikasi politik dan partisipasi pemilih [7].

Seiring dengan berkembangnya teknologi, teknik data mining menjadi salah satu solusi yang dapat digunakan untuk mengolah dan menganalisis data dalam jumlah besar secara otomatis dan efisien [8]. Salah satu teknik dalam data mining yang banyak digunakan adalah clustering, yaitu metode pengelompokan data berdasarkan tingkat kemiripan karakteristik yang dimiliki [9]. Clustering termasuk dalam kategori *unsupervised learning*, di mana proses pengelompokan dilakukan tanpa label data sebelumnya, sehingga cocok digunakan untuk eksplorasi data [10].

Metode clustering menjadi salah satu teknik yang cukup banyak digunakan dalam analisis data karena mampu menemukan pola tersembunyi tanpa

memerlukan label data sebelumnya. Teknik ini bekerja dengan mengelompokkan data yang memiliki karakteristik serupa ke dalam satu kelompok tertentu sehingga memudahkan proses interpretasi data. Dalam implementasinya, clustering sering digunakan dalam berbagai bidang seperti bisnis, pendidikan, kesehatan, dan pemerintahan. Kemampuan clustering dalam menyederhanakan data yang kompleks menjadikan metode ini relevan digunakan untuk membantu proses analisis pada data kependudukan maupun data pemilih.

Pada penelitian berbasis data kependudukan, clustering dapat memberikan gambaran mengenai distribusi masyarakat berdasarkan variabel tertentu. Informasi hasil pengelompokan tersebut dapat digunakan untuk mendukung proses evaluasi maupun penyusunan strategi pelayanan publik. Dalam konteks pemilu, hasil clustering dapat membantu pihak terkait dalam memahami persebaran karakteristik pemilih sehingga program sosialisasi dapat dilakukan secara lebih tepat sasaran. Selain itu, pengelompokan data juga membantu proses visualisasi informasi sehingga lebih mudah dipahami oleh pengambil Keputusan.

Algoritma K-Means merupakan salah satu metode clustering yang paling populer karena kesederhanaannya, kemudahan implementasi, serta efisiensi dalam menangani dataset berukuran besar [11]. K-Means bekerja dengan cara mengelompokkan data ke dalam sejumlah cluster berdasarkan kedekatan jarak terhadap centroid yang dihitung menggunakan metode jarak tertentu seperti Euclidean Distance [12]. Proses ini dilakukan secara iteratif hingga mencapai kondisi konvergen [13].

Pemilihan algoritma K-Means dalam penelitian ini didasarkan pada kemampuannya dalam melakukan pengelompokan data numerik secara efektif. Algoritma ini memiliki mekanisme kerja yang relatif sederhana dibandingkan metode clustering lainnya sehingga lebih mudah diimplementasikan dalam berbagai penelitian. Selain itu, K-Means juga memiliki waktu komputasi yang cukup cepat untuk dataset dengan jumlah data besar. Hal tersebut menjadi alasan mengapa algoritma ini sering digunakan dalam penelitian data mining yang berkaitan dengan segmentasi data.

K-Means bekerja dengan menentukan sejumlah pusat cluster (centroid) sebagai representasi dari setiap kelompok data. Setiap data kemudian akan dihitung jaraknya terhadap centroid menggunakan metode perhitungan tertentu, kemudian dimasukkan ke dalam cluster dengan jarak terdekat. Proses tersebut dilakukan secara berulang hingga posisi centroid stabil dan tidak mengalami perubahan yang signifikan. Melalui proses iteratif tersebut, algoritma K-Means mampu menghasilkan pengelompokan data yang memiliki tingkat kemiripan tinggi dalam satu cluster dan perbedaan yang cukup jelas antar cluster.

Berbagai penelitian sebelumnya menunjukkan bahwa algoritma K-Means dapat diterapkan dalam berbagai bidang, seperti segmentasi pelanggan,

analisis perilaku pengguna, serta pengelompokan data sosial [14]. Dalam konteks pemilihan, penerapan clustering dapat membantu mengidentifikasi pola distribusi pemilih dan memberikan wawasan yang lebih mendalam terkait karakteristik demografis masyarakat [15]. Namun demikian, penerapan metode ini pada data DPT di tingkat desa masih relatif terbatas, khususnya dalam segmentasi berbasis usia [16].

Segmentasi pemilih berdasarkan usia menjadi penting karena setiap kelompok usia memiliki karakteristik, preferensi, serta tingkat partisipasi yang berbeda dalam proses pemilu [17]. Kelompok usia muda cenderung lebih aktif dalam penggunaan teknologi digital, sedangkan kelompok lansia membutuhkan pendekatan yang lebih personal dan langsung [18].

Usia menjadi salah satu variabel penting dalam penelitian ini karena berkaitan erat dengan pola partisipasi masyarakat dalam pemilu. Setiap kelompok usia memiliki tingkat pengalaman, pola komunikasi, serta cara menerima informasi yang berbeda-beda. Kelompok usia muda umumnya lebih cepat menerima informasi melalui media digital dan memiliki tingkat mobilitas yang tinggi. Sebaliknya, kelompok usia lanjut lebih mengutamakan interaksi langsung dan cenderung memiliki keterbatasan dalam penggunaan teknologi digital. Perbedaan karakteristik tersebut menyebabkan pendekatan sosialisasi yang dilakukan perlu disesuaikan agar dapat menjangkau seluruh lapisan masyarakat secara optimal.

Selain berpengaruh terhadap strategi sosialisasi, segmentasi usia juga dapat membantu pemerintah desa dalam memahami komposisi penduduk usia produktif dan nonproduktif. Informasi tersebut dapat dimanfaatkan sebagai bahan evaluasi dalam penyusunan program pembangunan masyarakat. Dengan demikian, hasil penelitian tidak hanya bermanfaat dalam konteks pemilu, tetapi juga dapat memberikan kontribusi terhadap pengelolaan data kependudukan secara lebih luas.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam melakukan segmentasi pemilih berdasarkan kelompok usia menggunakan data DPT Desa Kesambi. Dengan memanfaatkan teknik clustering, diharapkan dapat diperoleh pengelompokan pemilih yang lebih terstruktur dan informatif. Hasil segmentasi ini diharapkan dapat memberikan kontribusi dalam mendukung pengambilan keputusan, khususnya dalam merancang strategi sosialisasi dan pendekatan kampanye yang lebih efektif dan tepat sasaran [19].

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik data mining, khususnya metode clustering untuk mengelompokkan data pemilih berdasarkan tingkat

kemiripan karakteristiknya. Clustering merupakan metode unsupervised learning yang bertujuan untuk membagi data menjadi kelompok-kelompok berdasarkan pola kemiripan internal. Algoritma K-Means dipilih dalam penelitian ini karena memiliki proses iteratif yang sederhana, implementasi yang mudah, serta kemampuan untuk mengolah dataset berukuran besar secara efisien.

Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada pengolahan data numerik dan analisis statistik sederhana untuk menghasilkan pola pengelompokan data. Metode kuantitatif memungkinkan proses analisis dilakukan secara objektif berdasarkan hasil perhitungan matematis dan komputasional. Dalam penelitian data mining, pendekatan kuantitatif sering digunakan karena mampu mengolah data dalam jumlah besar secara sistematis sehingga menghasilkan informasi yang lebih akurat dan terukur.

Penggunaan teknik clustering pada penelitian ini bertujuan untuk menemukan pola distribusi pemilih tanpa adanya label kategori sebelumnya. Berbeda dengan metode klasifikasi yang memerlukan data pelatihan, clustering bekerja dengan mengidentifikasi kemiripan antar data secara otomatis. Oleh karena itu, metode ini sangat sesuai digunakan dalam eksplorasi data pemilih yang sebelumnya belum memiliki kategori segmentasi tertentu.

Data primer penelitian bersumber dari Daftar Pemilih Tetap (DPT) Desa Kesambi yang berisi informasi nama, jenis kelamin, dan usia pemilih. Data sekunder ini diperoleh melalui studi dokumentasi dari instansi terkait dan disimpan dalam format Microsoft Excel untuk memfasilitasi proses preprocessing dan analisis selanjutnya [20]. Penggunaan format spreadsheet dipilih karena memudahkan proses pengelolaan data awal sebelum diintegrasikan ke dalam lingkungan pemrograman Python. Selain itu, format tersebut juga mendukung proses pengecekan manual terhadap kemungkinan kesalahan input data. Pengolahan data dilakukan menggunakan bahasa pemrograman Python berbasis platform Google Colab sebagai lingkungan komputasi cloud. Sebelum tahap clustering, data melalui proses pembersihan (data cleaning) untuk menghilangkan nilai null, data duplikat, dan outlier yang dapat memengaruhi kualitas hasil pengelompokan [20]. Tahap preprocessing menjadi bagian penting dalam penelitian karena kualitas data sangat memengaruhi hasil clustering yang dihasilkan. Data yang tidak konsisten dapat menyebabkan proses pengelompokan menjadi kurang optimal dan menghasilkan interpretasi yang tidak akurat.

Proses transformasi data dilakukan dengan mengkonversi variabel kategorikal menjadi numerik, khususnya variabel jenis kelamin dari kategori "L" (laki-laki) dan "P" (perempuan) menjadi nilai biner 0 dan 1. Transformasi ini esensial karena algoritma K-Means berbasis perhitungan jarak Euclidean yang hanya dapat diterapkan pada data numerik kontinu.

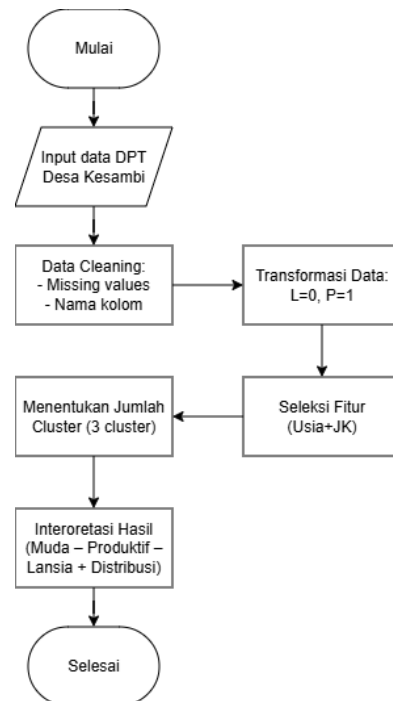
Seleksi fitur kemudian dilakukan dengan memilih dua atribut utama, yaitu usia dan jenis kelamin, sebagai input clustering[21].

Proses transformasi data dilakukan untuk memastikan seluruh atribut dapat diproses oleh algoritma K-Means. Dalam penelitian ini, atribut usia digunakan dalam bentuk numerik asli, sedangkan atribut jenis kelamin dikonversi menjadi representasi angka agar dapat dihitung menggunakan metode Euclidean Distance. Tahap transformasi ini penting karena algoritma clustering berbasis jarak memerlukan format data yang seragam agar proses perhitungan dapat dilakukan secara optimal.

Selain transformasi data, penelitian ini juga menerapkan proses normalisasi sederhana untuk menjaga konsistensi rentang nilai antar atribut. Hal ini dilakukan agar atribut dengan rentang nilai lebih besar tidak mendominasi proses perhitungan jarak. Dengan demikian, hasil clustering yang diperoleh dapat merepresentasikan karakteristik data secara lebih seimbang

Tahapan utama penelitian ini meliputi:

1. Pembersihan data untuk menghilangkan nilai kosong dan anomali.
2. Seleksi fitur clustering (usia dan jenis kelamin).
3. Implementasi algoritma K-Means untuk pengelompokan data.
4. Distribusi hasil clustering
5. Interpretasi dan validasi hasil clustering.



Gambar 1 Flowchart Tahapan Penelitian

Alur tahapan penelitian secara keseluruhan dapat dilihat pada Gambar 1 yang menggambarkan proses mulai dari pengolahan data hingga interpretasi hasil clustering.

Proses clustering dilakukan menggunakan algoritma K-Means dengan jumlah cluster sebanyak tiga kelompok ($k = 3$). Algoritma ini bekerja dengan cara menghitung jarak antara setiap data dengan pusat cluster (centroid) dan mengelompokkan data berdasarkan jarak terdekat. Perhitungan jarak dilakukan menggunakan metode Euclidean Distance sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

di mana x merupakan data, y merupakan centroid, dan n adalah jumlah atribut. Selanjutnya, nilai centroid diperbarui secara iteratif dengan menghitung rata-rata dari seluruh data dalam satu cluster menggunakan rumus:

$$C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$$

di mana C_k merupakan centroid pada cluster ke- k dan n_k adalah jumlah data dalam cluster tersebut. Proses ini dilakukan berulang hingga tidak terjadi perubahan signifikan pada centroid atau telah mencapai kondisi konvergen [22].

Penentuan jumlah cluster optimal dilakukan menggunakan metode Elbow dengan menganalisis nilai *Within Cluster Sum of Squares* (WCSS). Metode ini digunakan untuk mengidentifikasi titik optimal jumlah cluster berdasarkan perubahan nilai WCSS yang membentuk pola siku (elbow).

Metode Elbow digunakan karena mampu memberikan gambaran visual mengenai jumlah cluster yang paling optimal berdasarkan nilai WCSS. Semakin kecil nilai WCSS menunjukkan bahwa data dalam cluster memiliki tingkat kemiripan yang semakin tinggi. Namun, penambahan jumlah cluster secara terus-menerus juga dapat menyebabkan model menjadi terlalu kompleks. Oleh sebab itu, diperlukan titik keseimbangan yang ditunjukkan melalui pola siku (elbow) pada grafik hasil evaluasi.

Penggunaan metode evaluasi dalam clustering menjadi penting untuk memastikan bahwa hasil pengelompokan benar-benar mampu merepresentasikan pola data yang ada. Tanpa adanya proses evaluasi, jumlah cluster yang dipilih dapat bersifat subjektif dan kurang optimal. Dengan adanya metode Elbow, proses penentuan cluster dapat dilakukan secara lebih terukur dan sistematis.

Hasil clustering yang diperoleh kemudian dianalisis untuk mengetahui karakteristik masing-masing kelompok berdasarkan rata-rata usia, sehingga diperoleh tiga kategori utama, yaitu kelompok usia muda, usia produktif, dan usia lanjut. Selain itu, dilakukan analisis distribusi jenis kelamin dalam setiap cluster untuk melihat pengaruh variabel demografis terhadap hasil pengelompokan.

Penelitian ini menggunakan Google Colab sebagai lingkungan pengembangan dengan bahasa pemrograman Python. Library yang digunakan meliputi Pandas untuk manipulasi data, Matplotlib

untuk visualisasi, serta Scikit-learn untuk implementasi algoritma K-Means dan proses evaluasi clustering [23].

Visualisasi hasil clustering dilakukan menggunakan grafik dan diagram untuk mempermudah proses interpretasi data. Visualisasi membantu menunjukkan persebaran data pada masing-masing cluster sehingga pola pengelompokan dapat diamati dengan lebih jelas. Selain itu, penggunaan visualisasi juga mempermudah penyampaian informasi hasil penelitian kepada pembaca yang tidak memiliki latar belakang teknis dalam bidang data mining.

Seluruh tahapan penelitian dilakukan secara bertahap mulai dari pengumpulan data, preprocessing, transformasi data, implementasi algoritma, evaluasi cluster, hingga interpretasi hasil. Dengan tahapan yang sistematis, penelitian diharapkan mampu menghasilkan segmentasi data pemilih yang akurat dan dapat digunakan sebagai referensi dalam mendukung proses pengambilan keputusan berbasis data.

3. Hasil dan Pembahasan

3.1. Pengolahan Data

Data DPT Desa Kesambi terlebih dahulu dilakukan preprocessing, yaitu: (a) membersihkan nama kolom dari spasi, (b) mengubah atribut Jenis Kelamin menjadi numerik ($L = 0$, $P = 1$), dan (c) menghapus data yang memiliki nilai kosong (missing values). Dari total 1.466 rekod awal, ditemukan 1 rekod yang memiliki nilai kosong sehingga dihapus. Data yang digunakan dalam proses clustering berjumlah 1.465 rekod pemilih, terdiri dari 714 pemilih laki-laki dan 751 pemilih perempuan. Rentang usia pemilih berkisar antara 17 hingga 93 tahun, dengan rata-rata usia sebesar 42,29 tahun.

Variabel yang digunakan dalam proses clustering adalah usia dan jenis kelamin. Pemilihan variabel ini bertujuan untuk melihat segmentasi pemilih berdasarkan karakteristik demografis dasar.

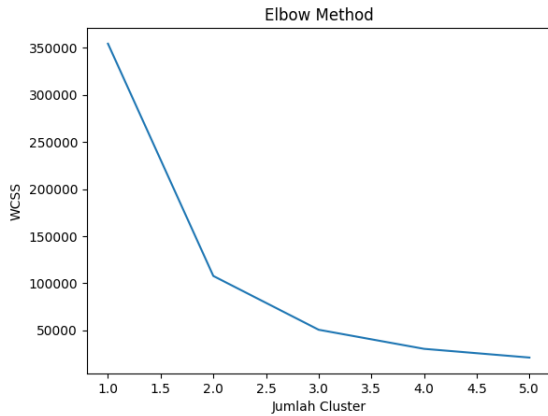
3.2. Penentuan Jumlah Cluster

Penentuan jumlah cluster dilakukan menggunakan metode Elbow Method dengan menghitung nilai *Within Cluster Sum of Squares* (WCSS) untuk jumlah cluster 1 sampai 5. Nilai WCSS yang diperoleh untuk setiap jumlah cluster disajikan pada Tabel 1.

Tabel 1. Nilai WCSS per Jumlah Cluster

Jumlah Cluster (k)	Nilai WCSS
1	354.308,46
2	107.668,53
3	50.067,56
4	30.020,65
5	20.565,45

Penurunan nilai WCSS mulai melambat secara signifikan pada $k = 3$, di mana penurunan dari $k=2$ ke $k=3$ sebesar 57.600,97 poin, sementara dari $k=3$ ke $k=4$ hanya sebesar 20.046,91 poin.

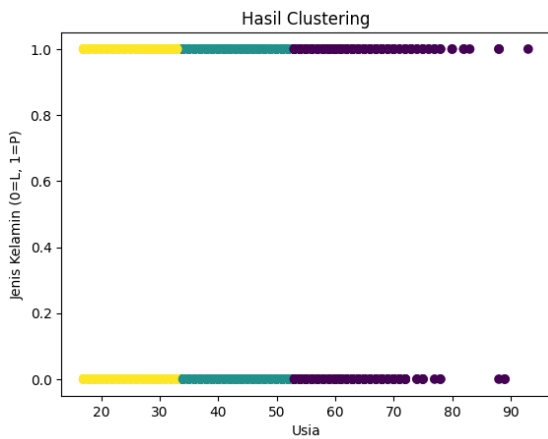


Gambar 1. Grafik Penentuan Jumlah Cluster menggunakan Elbow Method

Berdasarkan hal tersebut, titik siku (elbow) teridentifikasi pada $k = 3$, sehingga jumlah cluster optimal yang digunakan adalah 3 cluster

3.3. Hasil Clustering K-Means

Proses clustering dilakukan menggunakan algoritma K-Means dengan jumlah cluster sebanyak 3. Hasil clustering divisualisasikan dalam bentuk scatter plot berdasarkan usia dan jenis kelamin.



Gambar 2. Hasil Clustering

Nilai centroid akhir dari masing-masing cluster setelah proses iterasi K-Means konvergen ditampilkan pada Tabel 2.

Tabel 2. Nilai Centroid Hasil Clustering

Cluster	Label	Rata-Rata Usia	Rata-rata Jenis Kelamin
0	Lansia	63,00 tahun	0,5137
1	Produktif	44,12 tahun	0,5203
2	Muda	25,35 tahun	0,5029

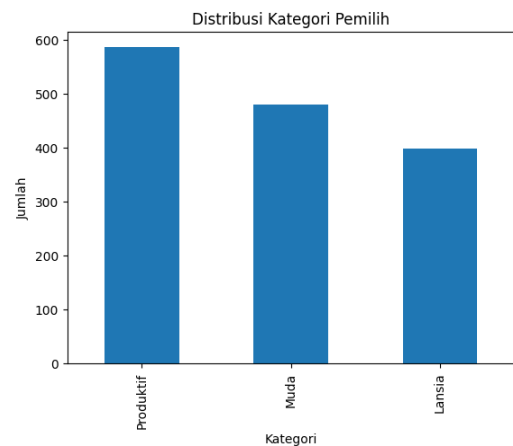
Berdasarkan tabel di atas, Cluster 0 memiliki rata-rata usia tertinggi yaitu 63,00 tahun, sehingga dikategorikan sebagai kelompok Lansia (rentang usia 54–93 tahun). Cluster 1 dengan rata-rata usia 44,12 tahun dikategorikan sebagai kelompok Usia Produktif (rentang usia 35–53 tahun), sedangkan Cluster 2 dengan rata-rata usia terendah 25,35 tahun dikategorikan sebagai kelompok Pemilih Muda (rentang usia 17–34 tahun)..

3.4 Distribusi Hasil Segmentasi

Berdasarkan hasil clustering, diperoleh distribusi jumlah pemilih pada masing-masing kategori sebagaimana disajikan pada Tabel 3.

Tabel 3. Distribusi Jumlah Pemilih per Kategori

Cluster	Label	Jumlah Pemilih	Persentase
0	Lansia	364	24,8%
1	Produktif	592	40,4%
2	Muda	509	34,7%
Total		1465	100%



Gambar 3. Distribusi Kategori Pemilih

Kelompok Usia Produktif merupakan kelompok terbesar dengan 592 pemilih (40,4%), diikuti oleh kelompok Pemilih Muda sebanyak 509 pemilih (34,7%), dan kelompok Lansia sebanyak 364 pemilih (24,8%). Hal ini menunjukkan bahwa struktur pemilih di Desa Kesambi didominasi oleh usia yang masih aktif secara sosial dan ekonomi.

3.5 Interpretasi dan Validasi Hasil Clustering

Untuk mengukur kualitas hasil clustering, dilakukan evaluasi menggunakan Silhouette Score. Hasil evaluasi menunjukkan nilai Silhouette Score sebesar 0,5698, yang mengindikasikan bahwa pemisahan antar cluster termasuk dalam kategori cukup baik (rentang 0,51–0,70). Nilai ini menunjukkan bahwa

algoritma K-Means dengan $k=3$ mampu membentuk kelompok yang cukup kohesif dan terpisah dengan baik.

Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan data pemilih secara efektif berdasarkan usia dan jenis kelamin. Beberapa poin penting dari hasil ini:

- 1) Usia menjadi faktor dominan dalam pembentukan cluster dibanding jenis kelamin, hal ini terlihat dari nilai rata-rata jenis kelamin yang relatif seragam di ketiga cluster (0,50–0,52), menunjukkan distribusi gender yang hampir merata di setiap kelompok usia.
- 2) Pengelompokan menjadi 3 kategori (Muda, Produktif, Lansia) sesuai dengan segmentasi demografis umum dalam analisis pemilih.
- 3) Informasi ini dapat dimanfaatkan oleh pihak terkait untuk menentukan strategi sosialisasi, menyesuaikan pendekatan kampanye, dan mengoptimalkan partisipasi pemilih.

Namun demikian, terdapat beberapa keterbatasan: (a) variabel yang digunakan masih terbatas pada usia dan jenis kelamin; (b) tidak mempertimbangkan faktor lain seperti pendidikan, pekerjaan, atau lokasi dusun. Segmentasi yang dihasilkan dapat diinterpretasikan sebagai berikut: (1) Pemilih Muda (34,7%): potensial untuk peningkatan partisipasi melalui media digital; (2) Usia Produktif (40,4%): kelompok paling strategis karena jumlah dominan; (3) Lansia (24,8%): memerlukan pendekatan khusus seperti sosialisasi langsung..

4. Kesimpulan

Penelitian ini telah berhasil menerapkan algoritma K-Means untuk melakukan segmentasi pemilih berdasarkan data Daftar Pemilih Tetap (DPT) Desa Kesambi dengan menggunakan variabel usia dan jenis kelamin. Beberapa kesimpulan yang dapat ditarik dari penelitian ini adalah sebagai berikut.

Pertama, proses preprocessing data menghasilkan 1.465 rekod pemilih yang valid dari total 1.466 data awal, setelah dilakukan penghapusan 1 rekod yang memiliki nilai kosong.

Kedua, penentuan jumlah cluster optimal menggunakan Elbow Method menunjukkan bahwa nilai $k = 3$ merupakan jumlah cluster yang paling optimal, ditandai dengan perlambatan signifikan penurunan nilai WCSS pada titik tersebut.

Ketiga, hasil clustering K-Means dengan $k = 3$ berhasil mengelompokkan pemilih Desa Kesambi ke dalam tiga segmen, yaitu: kelompok Pemilih Muda (usia 17–34 tahun) sebanyak 509 pemilih atau 34,7%, kelompok Usia Produktif (usia 35–53 tahun) sebanyak 592 pemilih atau 40,4%, dan kelompok Lansia (usia 54–93 tahun) sebanyak 364 pemilih atau

24,8%.

Keempat, evaluasi menggunakan Silhouette Score menghasilkan nilai 0,5698 yang mengindikasikan kualitas clustering dalam kategori cukup baik, sehingga hasil segmentasi ini dapat dinyatakan valid dan representatif.

Kelima, usia terbukti menjadi faktor dominan dalam pembentukan cluster, sementara distribusi jenis kelamin relatif merata di setiap kelompok. Temuan ini dapat menjadi acuan bagi pemerintah desa maupun penyelenggara pemilu dalam merancang strategi sosialisasi yang lebih tepat sasaran sesuai karakteristik masing-masing kelompok pemilih.

Untuk penelitian selanjutnya, disarankan untuk memperluas variabel yang digunakan seperti tingkat pendidikan, pekerjaan, dan lokasi dusun, serta membandingkan performa K-Means dengan algoritma clustering lain seperti DBSCAN atau Agglomerative Clustering guna memperoleh hasil segmentasi yang lebih komprehensif.

Daftar Rujukan

- [1] E. Jaatsiah, "Pengaruh # Indonesiagelap Terhadap Kesadaran Sosial Dan Tingkat Kepercayaan Mahasiswa Surabaya Pada Pemerintah," Vol. 17, No. 1, Pp. 19–31, 2026, Doi: 10.14710/Politika.17.1.2027.19-31.
- [2] R. A. Artisa, F. L. M., And M. Iqbal, "Citizen-Centric Governance And Public Service Transformation In The Digital Era : Case Of Banyuwangi , Indonesia," Vol. 17, No. 1, Pp. 65–101, 2026.
- [3] J. Apriyanto, "Penerapan Clustering Pada Rata-Rata Lama Sekolah Di Provinsi Indonesia Menggunakan Algoritma K-Means Dan Fuzzy C-Means," Vol. 19, No. 1, Pp. 36–42, 2024.
- [4] I. Engineering *Et AL.*, "Implementasi Algoritma K-Means Untuk Mengklaster," Vol. 6, No. 1, Pp. 39–48, 2022.
- [5] A. Alif, R. Mulyana, A. R. Juwita, A. M. Siregar, And A. Fauzi, "Penerapan Algoritma K-Means Clustering Dan Hierarchical Clustering Dalam Mengelompokkan Data Pengangguran Di Karawang," No. 2021, 2024, Doi: 10.33364/Algoritma/V.21-2.2155.
- [6] E. V. A. K. Farhan, D. Erowati, And M. D. Rizqi, "E-Govern M Ent Interoperability In Pro M Oting E-Governance : A Study Of The Super Mantab Application In Musi Rawas Regency," 2026.
- [7] R. Triyaningsih, P. Studi, S. Informasi, And U. Jambi, "Multi Proximity : Jurnal Statistika Universitas Jambi," Vol. 2, No. 2, Pp. 118–128, 2023.
- [8] I. R. A. Meiyenti And W. M. Assani, "Machine Learning – Driven Governance : Evaluating Village Fund Effectiveness With Nlp And Satellite Data Across 74 , 000 Indonesian Villages," Vol. 17, No. 1, 2026.
- [9] J. Zhang And J. B. T.-S. C. E. T. Li, "Chapter 14 - Intelligent Language Knowledge For Cognitive Engine," Academic Press, 2023, Pp. 187–197. Doi: <https://doi.org/10.1016/B978-0-323-95107-4.00012-3>.
- [10] R. S. Nurhalizah, R. Ardianto, And P. Purwono, "Analisis Supervised Dan Unsupervised Learning Pada Machine Learning: Systematic Literature Review," *J. Ilmu Komput. Dan Inform.*, Vol. 4, No. 1, Pp. 61–72, 2024, Doi: 10.54082/Jiki.168.
- [11] S. Sulistina, P. Eko Prasetyo Utomo, And B. Ferdian Hutabarat, "Klasterisasi Wilayah Rawan Kriminalitas Di

- Kota Jambi (2022-2024) Menggunakan Algoritma K-Means,” *J. Instek (Informatika Sains Dan Teknol.*, Vol. 10, No. 2, Pp. 546–560, 2025, Doi: 10.24252/Instek.V10i2.62051.
- [12] Matus Ivan Bimasena, I Gede Aris Gunadi, And I Made Agus Wirawan, “Studi Perbandingan Algoritma Euclidean, Manhattan Dan Chebysev Distance Untuk Optimasi Metode K-Means Clustering Pada Pengelompokan Data Penerimaan Mahasiswa Baru,” *J. Pendidik. Teknol. Dan Kejur.*, Vol. 22, No. 2, Pp. 116–127, 2025.
- [13] M. Iqbal, S. Syaripuddin, And M. N. Huda, “Implementasi Algoritma K-Means Clustering Dengan Jarak Euclidean Dalam Mengelompokkan Daerah Penyebaran Covid-19 Di Kabupaten Bogor,” *Basis J. Ilm. Mat.*, Vol. 2, No. 1, Pp. 47–56, 2023, Doi: 10.30872/Basis.V2i1.1019.
- [14] M. Sahira, A. Salsabila, S. Salsabila, A. N. Putri, K. D. Tania, And W. K. Sari, “Penerapan Metode K-Means Clustering Untuk Segmentasi Performa Pembalap F1 Season 2024,” *Infomatek*, Vol. 27, No. 1, Pp. 113–122, 2025, Doi: 10.23969/Infomatek.V27i1.24297.
- [15] Baiq Nikum Yulisasih, H. Herman, And S. Sunardi, “K-Means Clustering Method For Customer Segmentation Based On Potential Purchases,” *J. Eltikom*, Vol. 8, No. 1, Pp. 83–90, 2024, Doi: 10.31961/Eltikom.V8i1.1137.
- [16] M. T. Wardani, V. O. Ramadhani, N. Anggreani, Sumanto, And A. D. Kuswanto, “Pengelompokan Provinsi Berdasarkan Kelayakan Ruang Kelas Dan Tenaga Kependidikan Sekolah Dasar Menggunakan Algoritma K-Means: Analisa Data Periode 2023-2024,” *Riau J. Tek. Inform.*, Vol. 4, No. 2, Pp. 163–172, 2025, Doi: 10.30606/Rjti.V4i2.3389.
- [17] P. Studi, S. Data, F. I. Komputer, U. Pembangunan, And N. Veteran, “Statmat (Jurnal Statistika Dan Matematika),” Vol. 7, No. 3, Pp. 482–502, 2025.
- [18] A. Kurnia, “Perbandingan Algoritma K-Means Dan Fuzzy C-Means Untuk Clustering Puskesmas Berdasarkan Gizi Balita Surabaya,” *J. Process.*, Vol. 18, No. 1, Pp. 83–88, 2023, Doi: 10.33998/Processor.2023.18.1.696.
- [19] A. F. Firdaus, A. Fahriza Fitriani, And E. Prasetyo Nugroho3, “Parallel Computing Pada Clustering K-Means Untuk Analisis Keketatan Program Studi Sntb 2023,” *Komputika J. Sist. Komput.*, Vol. 14, No. 1, P. 7, 2025, Doi: 10.34010/Komputika.V14i1.14983.
- [20] A. Obot, A. Nolte, And J. D. Herbsleb, “How To Organize An In-Person , Online Or Hybrid Hackathon – A Revised Planning Kit,” 2025.
- [21] S. H. Hassan And A. M. Abdulazeez, “Indonesian Journal Of Computer Science,” Vol. 13, No. 1, Pp. 4049–4066.
- [22] M. O. Akhter, N. Muhammad, And A. Razia, “Cmos Technology,” No. June, Pp. 553–568, 2022, Doi: 10.1049/Cds2.12125.
- [23] M. R. Zirnbaue, “Color-Flavor Transformation Revisited,” Pp. 1–45, 2021.