

Analisis Daftar Pemilih Tetap Pemilihan Gubernur Dan Wakil Gubernur Menggunakan Algoritma *K-Means*

Syahrul Dwi Hilda^{*1}, Apriade Voutama², Yuyun Umaidah³

^{1,2,3}Sistem Informasi ,Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang,
Karawang

Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361,
Tlp (0267) 641177

e-mail: ¹syahruldwi hilda@gmail.com, ²apriade.voutama@staff.unsika.ac.id,
³yuyun.umaidah@staff.unsika.ac.id

Abstrak

Pada daftar pemilih tetap merupakan sebuah kumpulan data yang sangat banyak dalam sebuah pemilihan. Data daftar pemilih tetap sangat penting bagi sebuah lembaga yang berhubungan dengan hal tersebut salah satunya sebuah Desa. Kurangnya pemahaman mengenai Data Mining dalam tingkat desa merupakan sebuah hal yang kurang baik, karena Data Mining ini sangat dibutuhkan apabila dihubungkan dengan data – data yang sangat banyak karena salah satunya dapat mempermudah dalam melakukan mengelompokan data atau clustering dengan algoritma k-means yang menjadi pilihan yang efektif untuk clustering. Pada penelitian ini menggunakan metode clustering algoritma k-means didukung dengan aplikasi RapidMiner. Pengolahan data daftar pemilih tetap akan dibutuhkan oleh lembaga atau yang berhubungan. Hasil penelitian ini yang dibantu oleh aplikasi rapidminer dengan metode clustering algoritma k-means berhasil mengelompokan atau clustering dari daftar pemilih tetap dengan variabel usia dan alamat di lembaga atau yang membutuhkan khususnya pada penelitian ini di Desa Curug Kecamatan Klari Kabupaten Karawang.

Kata kunci— daftar pemilih tetap, k-means, rapidminer, clustering, data mining

Abstract

Permanent voters list is a very large collection of data in an election. Permanent voters list data is very important for an institution related to this, one of which is a village. Lack of understanding of Data Mining at the village level is a bad thing, because Data Mining is very much needed when associated with a lot of data because one of them can make it easier to group data or clustering with the k-means algorithm which is an effective choice for clustering. In this study using the k-means algorithm clustering method supported by the RapidMiner application. Processing of permanent voter list data will be needed by institutions or related. The results of this research assisted by the rapidminer application with the k-means algorithm clustering method successfully grouped or clustering of the permanent voters list with age and address variables in institutions or those in need, especially in this study in Curug Village, Klari District, Karawang Regency.

Keywords— permanent voter list, k-means, rapidminer, clustering, data mining

1. PENDAHULUAN

Di dalam sila keempat Pancasila dan Pasal 1 ayat (2) Undang-Undang Dasar Negara Republik Indonesia pada Tahun 1945 diwujudkan melalui pemilihan umum, yang juga merupakan sarana untuk menjalankan sistem demokrasi. Pemilihan umum digunakan untuk memilih orang-orang yang akan duduk sebagai wakil rakyat di lembaga eksekutif dan legislatif baik di tingkat nasional maupun daerah. Sejak tahun 1955, pemilihan umum telah diselenggarakan, dan pemilihan yang terbaru adalah pada tahun 2019[1][2].

Kurangnya strategi sosialisasi yang efektif dari lembaga terkait kepada masyarakat, sedangkan apabila sosialisasi harus dilakukan dengan cara menyeluruh dan acak di setiap daerah, maka akan memakan biaya yang cukup besar. Kemajuan teknologi saat ini menghasilkan situasi yang kaya akan data namun minimnya informasi[3].

Setiap pemilihan umum menghasilkan sejumlah besar data dari Daftar Pemilih Tetap (DPT), sehingga para data scientist diharuskan untuk memproses dan menganalisa data yang sangat banyak dalam sistem untuk menemukan pola dan algoritma yang dapat digunakan untuk mendapatkan informasi baru. Teknik-teknik data mining digunakan, bersama dengan alat bantu dari sistem, untuk memproses dan menganalisis banyak data dan informasi dalam jumlah yang sangat besar. Pada penelitian ini menggunakan algoritma K-Means yang adalah salah satu metode dari data mining yang dapat membagi dan mengelompokkan data dalam jumlah yang sangat besar menjadi *record-record* yang lebih kecil. Algoritma ini mengelompokkan data ke dalam sejumlah klaster yang telah ditentukan sebelumnya. Pada awalnya, K-Means memilih jumlah klaster yang diinginkan (k) dan titik awal yang berfungsi sebagai centroid untuk setiap klaster. Data kemudian diklasifikasikan ke dalam klaster yang memiliki centroid terdekat dengan data tersebut. Setelah semua data diklasifikasikan, centroid dari setiap klaster diperbarui dengan menggunakan rata-rata data yang termasuk dalam klaster tersebut. Proses ini diulangi beberapa kali hingga tidak ada perubahan dalam klasifikasi data [3][4].

Beberapa penelitian yang membahas mengenai K-Means diantaranya penelitian [3], dilakukan penelitian dengan menggunakan *K-Means* yang bertujuan untuk menerapkan teknik *data mining*. Berdasarkan penelitian yang telah dilakukan, didapatkan hasil analisis yaitu *cluster* Kecamatan Langkapura di dapat 10167 orang untuk usia muda, untuk usia dewasa didapat 9527 orang, berjumlah 4821 untuk lansia. usia muda pada Kecamatan Rajabasa berjumlah 12557 orang, berjumlah 10930 orang untuk dewasa dan untuk lansia berjumlah 5097orang. Untuk usia muda, lansia dan dewasa pada Kecamatan Kemiling didapat 19442 orang, 19086 orang dan 9394 orang.

Teknik Elbow dan algoritma K-Means digunakan dalam penelitian [5]. Temuan penelitian menunjukkan bahwa siswa dapat dibagi menjadi tiga *cluster*, dengan *cluster* 0 memiliki nilai rata-rata terendah dan anggota *cluster* terbanyak yaitu 57 siswa. *Cluster* 1 memiliki nilai rata-rata tinggi dan anggota *cluster* paling sedikit, yaitu 1 siswa. Setelah itu, *cluster* 2 adalah kelompok nilai rata-rata dengan 42 siswa di dalamnya.

Lalu pada penelitian [6] menggunakan K-Means sebagai algoritma yang diterapkan dalam penelitian ini. Teknik elbow digunakan untuk memilih *cluster* yang ideal. Jumlah *cluster* terbaik, seperti yang ditentukan oleh pendekatan *Silhouette*, adalah menjadi dua *cluster*. Nilai sebesar 0,408 untuk *Davies-Bouldin Index* (DBI) dan nilai sebesar 0,747 untuk *Silhouette Coefficient* (struktur yang kuat) adalah hasil dari studi pengelompokan. *Cluster* 1 memiliki tujuh kecamatan, sedangkan *Cluster* 2 memiliki dua puluh tiga kecamatan. Berdasarkan hasil analisis pengelompokan, *Cluster* 1 memiliki jumlah rata-rata atribut pendidikan yang lebih tinggi dibandingkan *Cluster* 2, termasuk guru, kelas, laboratorium, dan perpustakaan pada semua tingkat pendidikan (SD, SMP, SMA, SMK, dan SLB), serta jumlah rata-rata penduduk yang

lebih tinggi. Hasilnya, *Cluster* 1 masuk dalam kategori tinggi, sedangkan *Cluster* 2 masuk dalam kategori rendah.

Sementara itu pada penelitian [7] menggunakan Algoritma K-Means yang mana dari evaluasi kualitas *cluster* pada dataset Penyakit Diare Kabupaten Karawang tahun 2017-2019. Pengujian kualitas cluster yang dievaluasi yaitu rentang $k=4$ sampai dengan $k=8$, dari tahun 2017 sampai dengan tahun 2019 dengan 50 data uji. Pada tahap evaluasi, digunakan hasil *cluster* terbaik yaitu *cluster* 5 yang memiliki nilai Silhouette Coefficient sebesar 0.4960an.

Kemudian pada penelitian [8] menggunakan metode *clustering* dan algoritma K-Means dengan jumlah sampel data berjumlah 27 Kabupaten/Kota di Jawa Barat, yang menghasilkan 3 *cluster*. Menggunakan silhouette coefficient untuk hasil evaluasi menunjukkan bahwa pembentukan kelompok atau *cluster* sebanyak 3 ($k=3$) adalah jumlah *cluster* yang optimal.

Kali ini pada penelitian [9] menggunakan 3 metode pengujian, pengujian pertama, yang menggunakan pendekatan naive bayes, dengan akurasi 58,62%. Pengujian kedua memiliki tingkat akurasi 65,52% dan menggunakan metode *decision tree*. Pengujian ketiga dengan tingkat akurasi 51,72% dengan menggunakan metode random forest. Mempertimbangkan temuan-temuan dari pengujian yang telah dikumpulkan menunjukkan bahwa pendekatan *decision tree* merupakan metode terbaik untuk meramalkan kemajuan akademik mahasiswa berdasarkan hasil pengujian yang menggunakan menggunakan 3 metode dibantu aplikasi RapidMiner telah dilakukan, Hasil pengujian menunjukkan teknik *Decision tree*, dengan tingkat akurasi sebesar 65,52%, merupakan strategi yang paling akurat untuk memprediksi kemajuan akademik mahasiswa.

Setelah itu pada penelitian [10] Algoritma K-Means dan K-Medoids akan digunakan sebagai cara untuk mengelompokkan data obat. Penelitian ini bertujuan untuk meng-*cluster* informasi obat di Puskesmas Karangsambung ke dalam beberapa kelompok yang dapat digunakan sebagai panduan perencanaan obat. Tiga kategori pengelompokan data adalah lambat, sedang, dan cepat. Temuan dari dua algoritma menunjukkan yaitu Hasil yang diberikan oleh algoritma K-Means lebih baik daripada algoritma K-Medoids. Dibandingkan dengan K-Medoids, Silhouette Coefficient yang didapatkan oleh algoritma K-Means lebih tinggi, yaitu 0.627. K-Medoids adalah 0.536, namun K-Means adalah 0.540, yang menunjukkan bahwa temuan pengelompokan K-Means lebih berkualitas daripada K-Medoids. Berdasarkan temuan pengelompokan, Obat Puskesmas dapat melakukan pemerataan stok obat dengan menggunakan data pengelompokan obat untuk memastikan bahwa tidak ada kelebihan atau kekurangan stok obat.

Berdasarkan latar belakang yang sudah dijelaskan, terdapat beberapa perbedaan yang perlu diatasi dalam penelitian ini. Meskipun penelitian ini menggunakan metode *clustering* algoritma K-Means *Clustering*, belum ada penjelasan yang memadai mengapa algoritma ini dipilih dan bagaimana kinerjanya dibandingkan dengan algoritma clustering lainnya. Kemudian belum ada penjelasan yang jelas tentang bagaimana metode clustering ini dapat diterapkan secara konkret dalam konteks daftar pemilih tetap (DPT) pada pemilihan umum dan jenis informasi baru apa yang diharapkan bisa ditemukan. Selain itu, validitas dan akurasi penggunaan *clustering* dalam mengolah dan menganalisis data DPT juga perlu dievaluasi lebih lanjut. Oleh karena itu, penelitian ini diharapkan dapat mengisi celah pengetahuan ini dan memberikan wawasan baru tentang penggunaan metode clustering algoritma K-Means *Clustering* dalam konteks pemilihan umum dan data DPT.

2. METODE PENELITIAN

Berikut merupakan langkah-langkah yang digunakan untuk menyelesaikan penelitian:

2.1 *Persiapan Internal*

Tahap ini merupakan tahap menentukan objek penelitian, mengidentifikasi masalah yang terdapat pada objek penelitian yang dipilih dan melakukan studi literatur yang dapat menyelesaikan masalah yang terdapat pada objek penelitian.

2.2 *Perencanaan*

Tahap ini merupakan tahap untuk menemukan ruang lingkup yang akan dikerjakan, menemukan metodologi penelitian yang akan digunakan, merancang kerangka kerja penelitian agar tidak keluar dari ruang lingkup yang telah ditetapkan, dan merancang skema analisis untuk memudahkan proses analisis.

2.3 *Pengumpulan Data*

Pada tahap ini pengumpulan data dibutuhkan untuk melakukan penelitian. Teknik pengumpulan data dilakukan secara acak dengan menggunakan cara Studi Literatur yaitu kegiatan mengumpulkan dan mempelajari berbagai literatur, artikel dan jurnal penelitian yang berkaitan dengan penelitian.

2.4 *Pengolahan Data*

Tahap ini merupakan tahap untuk melakukan pengelompokan daftar pemilih tetap yang sesuai dengan *cluster* usia, jumlah pemilih atau pun per alamat.

2.5 *Penyampaian Saran dan Kesimpulan*

Tahap ini berisi saran dan kesimpulan dari seluruh proses penelitian

3. HASIL DAN PEMBAHASAN

3.1 *Jumlah Data*

Data yang digunakan dalam penelitian ini terdiri dari daftar pemilih tetap Desa Curug pada pemilihan Gubernur dan Wakil Gubernur Jawa Barat tahun 2018. Jumlah total data yang digunakan sebanyak 7639.

3.2 *Variabel yang Digunakan*

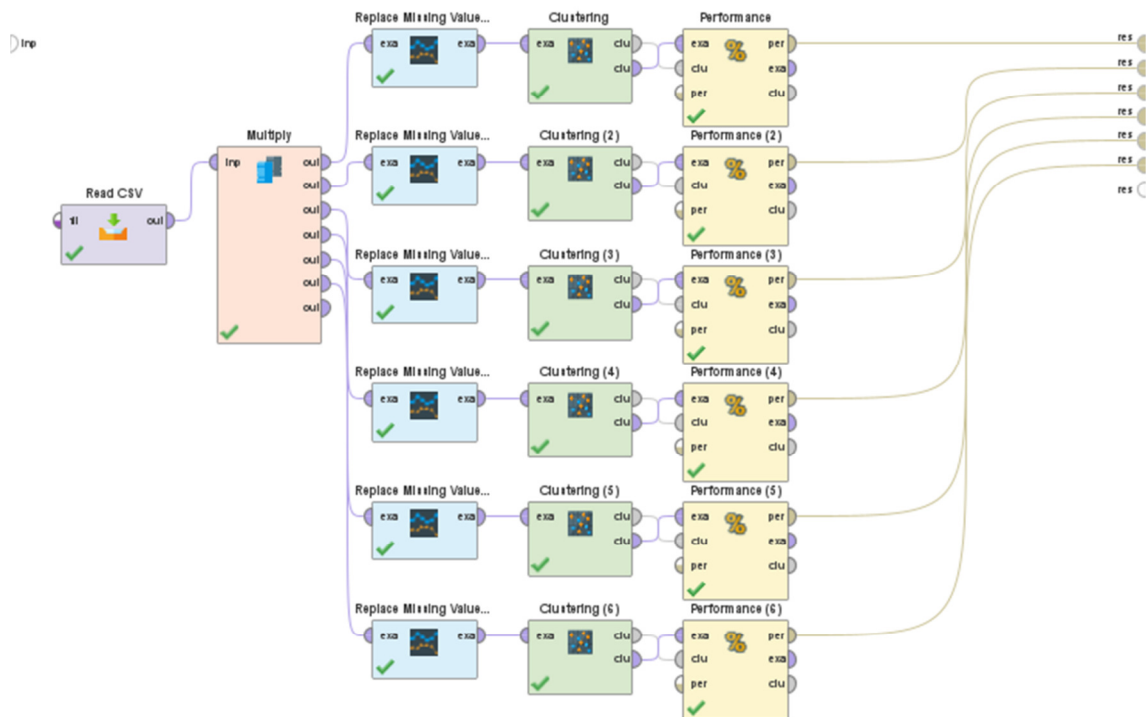
Penelitian menggunakan variabel usia, dan variabel alamat/dusun. Lihat Tabel 1. Variabel yang Digunakan.

Tabel 1. Variabel yang Digunakan

Alamat	Usia
Dusun Cilentah	37
...	...
Dusun Parungkadali	17

3.3 *Penentuan K Algoritma K-Means pada RapidMiner*

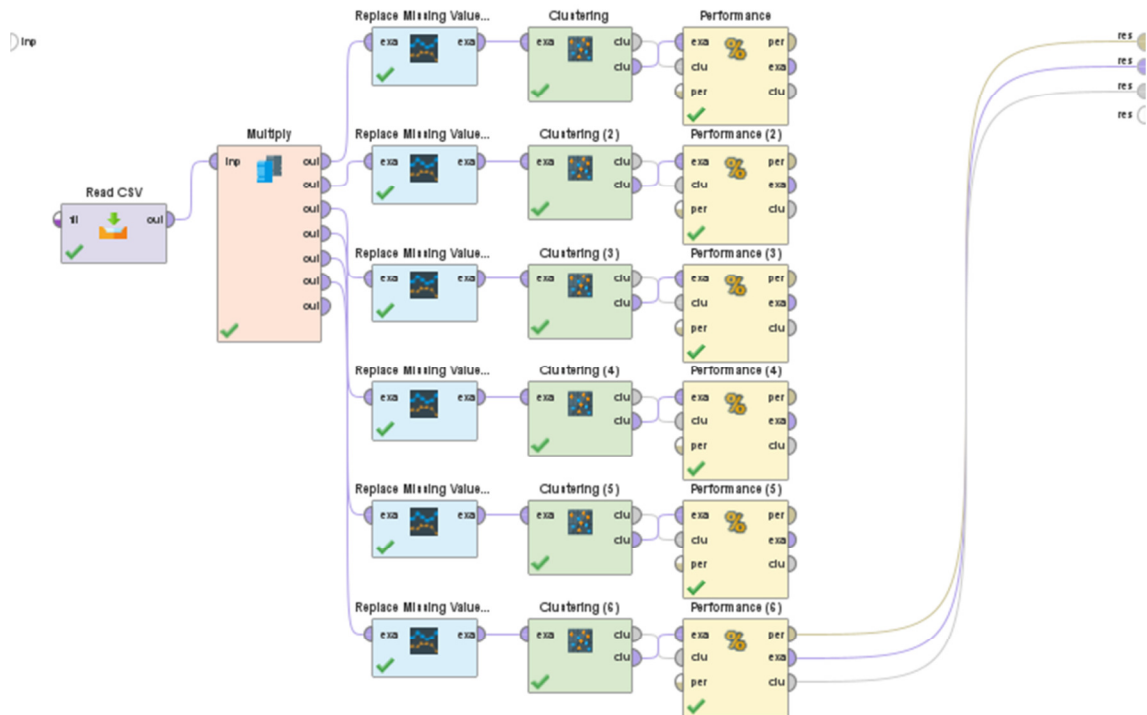
Untuk menentukan k algoritma K-Means nya, langkah pertamanya adalah membuat desain operator untuk melihat nilai dari *Davies Bouldin* terkecil.



Gambar 1. Desain Operator pada Aplikasi RapidMiner

Berikut fungsi dari setiap operator Pada Gambar 1. Desain Operator pada aplikasi RapidMiner:

- Operator *Read CSV* berfungsi sebagai input untuk memasukkan dataset.
- Operator *Multiply* berfungsi untuk mempermudah pengujian ganda agar tidak perlu memasukkan operator *Read csv* kembali.
- Operator *Replace Missing Value* berfungsi untuk mengatasi data yang hilang atau tidak lengkap.
- Operator *Clustering* digunakan untuk mengelompokkan data ke dalam kelompok atau *cluster* yang memiliki karakteristik atau ciri-ciri yang sama.
- Operator *Performance* berfungsi untuk memilih model atau algoritma yang paling sesuai untuk digunakan dalam proses analisis data.



Gambar 2. Desain Operator Jumlah k yang Sudah Ditentukan

Untuk menentukan jumlah k yang akan digunakan untuk menentukan banyaknya jumlah *cluster* yang di tentukan dengan nilai Davies Bouldin Terkecil, Lihat Tabel 2. Nilai Davies Bouldin pada setiap k. Maka, desain operatornya berubah menjadi seperti pada Gambar 2. Desain Operator jumlah k yang sudah ditentukan, hanya saja perbedaannya semua *output* di fokuskan pada operator *Performance*.

Tabel 2. Nilai Davies Bouldin pada Setiap k

Nilai k	Nilai Davies Bouldin
k = 2	Davies Bouldin: 0.552
k = 3	Davies Bouldin: 0.525
k = 4	Davies Bouldin: 0.520
k = 5	Davies Bouldin: 0.516
k = 6	Davies Bouldin: 0.544
k = 7	Davies Bouldin: 0.499

Setelah mengetahui nilai k yang terbaik dari nilai Davies Bouldin terkecil, yang diperoleh k = 7 dengan nilai Davies Bouldin 0.499, selanjutnya hanya perlu tinggal melanjutkan *clustering* pada nilai k yang terbaik.

3. 4 Performance Vector

*Performance*Vector:

Avg. within centroid distance: 10.372

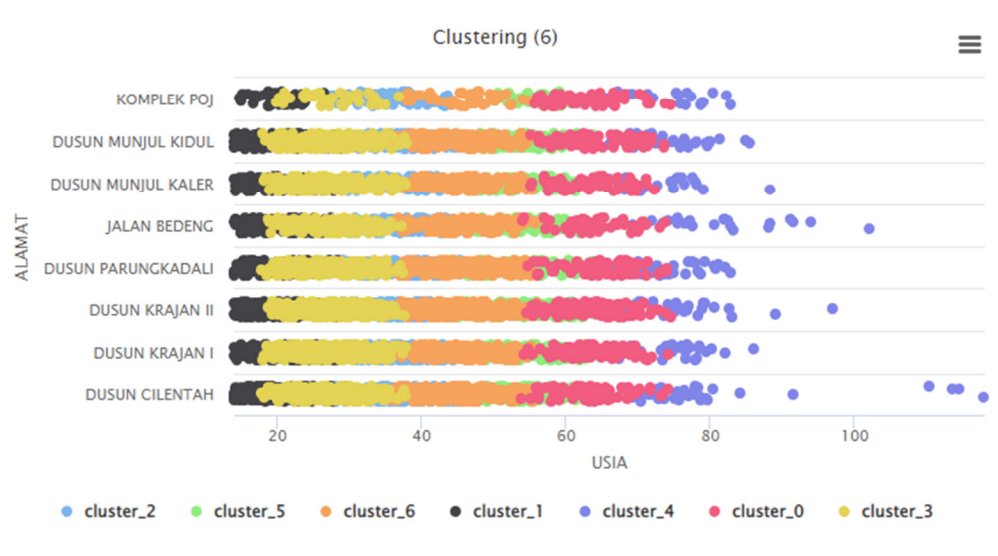
Avg. within centroid distance_cluster_0: 13.033

Avg. within centroid distance_cluster_1: 8.165

Avg. within centroid distance_cluster_2: 102.691
 Avg. within centroid distance_cluster_3: 9.811
 Avg. within centroid distance_cluster_4: 8.117
 Avg. within centroid distance_cluster_5: 7.833
 Avg. within centroid distance_cluster_6: 15.764
 Davies Bouldin: 0.499

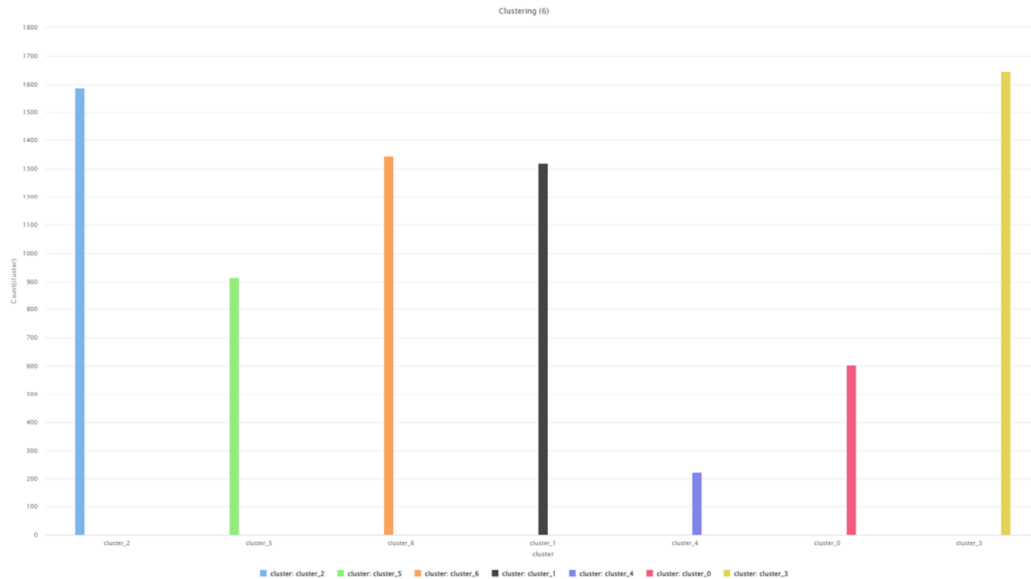
3. 5 Visualisasi Hasil

Berikut adalah hasil yang diperoleh berdasarkan 6 *cluster* pada aplikasi *RapidMiner*.



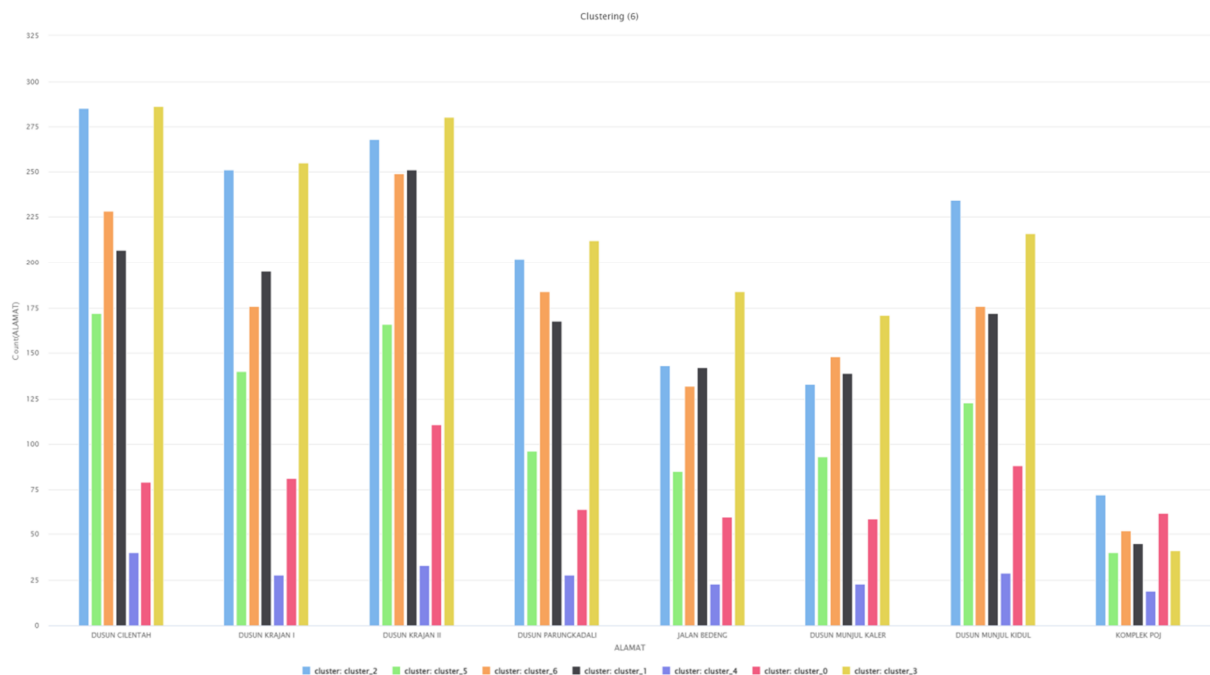
Gambar 3. Clustering Usia Berdasarkan Alamat

Pada Gambar 3. *Clustering* Usia berdasarkan Alamat, untuk melihat rentang usia di setiap alamat yang di tandai oleh warna yang berbeda di setiap *cluster*.



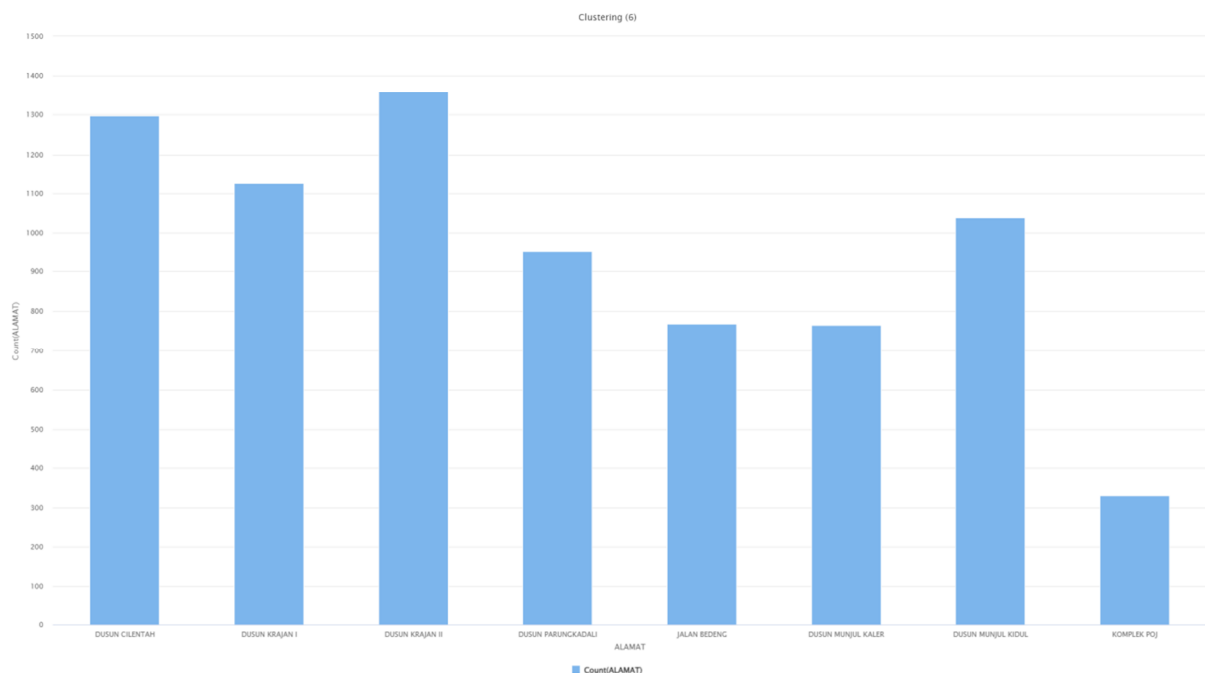
Gambar 4. Clustering Berdasarkan Cluster dan Frekuensi Banyaknya Data

Pada Gambar 4. *Clustering* berdasarkan *cluster* dan frekuensi banyaknya data, untuk melihat banyaknya frekuensi banyaknya data di setiap *cluster* yang di tandai oleh warna yang berbeda di setiap *cluster*.



Gambar 5. Clustering Berdasarkan Alamat dan Frekuensi Banyaknya Data

Pada Gambar 5. *Clustering* berdasarkan Alamat dan frekuensi banyaknya data, untuk melihat frekuensi banyaknya data di setiap alamat yang di bagi berdasarkan tiap *cluster* yang ditandai oleh warna yang berbeda di tiap *cluster*-nya.



Gambar 6. Clustering Berdasarkan Banyaknya Data Pemilih di Setiap Alamat

Pada Gambar 6. *Clustering* berdasarkan banyaknya data pemilih di setiap alamat, untuk melihat banyak nya pemilih pada setiap alamat.

3. 6 Cluster Model

Cluster 0: 604 items
 Cluster 1: 1319 items
 Cluster 2: 1588 items
 Cluster 3: 1645 items
 Cluster 4: 223 items
 Cluster 5: 915 items
 Cluster 6: 1345 items
 Total number of items: 7639

4. KESIMPULAN

Hasil Penelitian

Telah dilakukan pencarian nilai K dengan menggunakan aplikasi RapidMiner sebelum melakukan pengolahan data. Diperoleh nilai K terbaik yaitu 7 dengan nilai Davies Bouldin 0.499, Kemudian juga dilakukan pengolahan data Daftar Pemilih Tetap Pemilihan Gubernur dan Wakil Gubernur di Jawa Barat tahun 2018 di Desa Curug sebanyak 7639 data, dengan menggunakan Data Mining Algoritma K-Means *Clustering* sehingga diperoleh 7 *Cluster* dari 7639 jumlah data.

Kelebihan Penelitian

Dapat menentukan *cluster*/kelompok Daftar Pemilih Tetap Pemilihan Gubernur dan Wakil Gubernur di Jawa Barat tahun 2018 di Desa Curug secara transparan sesuai dengan usia dan alamat pemilih. Dan Algoritma K-Means *Clustering* dapat digunakan untuk menentukan *cluster*/kelompok dalam proses daftar pemilih tetap sesuai usia dan alamat pemilih.

Kekurangan Penelitian

Tidak dilakukannya perhitungan manual, dan hanya mengandalkan RapidMiner untuk menentukan nilai Davies Bouldin.

5. SARAN

Hasil dari penelitian yang diperoleh dapat dipakai oleh Desa Curug sebagai salah satu metode mengelompokkan daftar pemilih tetap pemilihan Gubernur dan Wakil Gubernur di Jawa Barat tahun 2018 dan seterusnya.

Penelitian selanjutnya dapat dilakukan dengan penghitungan manual untuk membuktikan nilai dari Davies Bouldin, juga dapat dilakukan dengan dataset yang berbeda.

DAFTAR PUSTAKA

- [1] Y. Yusrin and S. Salpina, “Partisipasi Generasi Millennial Dalam Mengawasi Tahapan Pemilu 2024,” *J. Educ.*, Vol. 5, No. 3, pp. 9646–9653, 2023, doi: 10.31004/joe.v5i3.1842.
- [2] I. Nasir, “Analisis Hukum Penanganan Pelanggaran Administrasi Pemilu/Pemilihan,” *Khazanah Huk.*, Vol. 2, No. 1, pp. 41–50, 2020, doi: 10.15575/kh.v2i1.7689.
- [3] G. Fajriansyah, G. F. Nama, and Y. Mulyani, “Analisis Daftar Pemilih Tetap pada Hasil Rekapitulasi KPU Berdasarkan Usia Menggunakan Algoritma K-Means (Studi kasus : Kota Bandar Lampung),” *Electrician*, Vol. 15, No. 1, pp. 39–53, 2021, doi: 10.23960/elc.v15n1.2147.
- [4] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, “Analisis K-Means Clustering pada Data Sepeda Motor,” *INFORMAL Informatics J.*, Vol. 5, No. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.
- [5] A. P. Riani, A. Voutama, and T. Ridwan, “Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow,” *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, Vol. 6, No. 1, p. 164, 2023, doi: 10.53513/jsk.v6i1.7351.
- [6] A. Amrullah, I. Purnamasari, B. N. Sari, G. Garono, and A. Voutama, “Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang),” *J. Inform. dan Rekayasa Elektron.*, Vol. 5, No. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.

- [7] V. Ramadhan and A. Voutama, “Clustering Menggunakan Algoritma K-Means pada Penyakit ISPA di Puskesmas Kabupaten Karawang,” *J. Pendidik. dan Konseling*, Vol. 4, No. 5, pp. 462–473, 2022.
- [8] B. Nurseptia, A. Voutama, and N. Haeryana, “Penerapan Algoritma K-Means Untuk Mengelompokkan Kabupaten/Kota Dalam Upaya Pemetaan Lapangan Pekerjaan Baru,” *J. Teknol. Informasi*, Vol. 6, No. 2, pp. 181–186, 2022.
- [9] F. M. Almufqi and A. Voutama, “Perbandingan Metode Data Mining Untuk Memprediksi Prestasi,” *J. Tek.*, Vol. 15, No. 1, pp. 61–66, 2023, doi: 10.30736/jt.v15i1.929.
- [10] R. A. Farissa, R. Mayasari, and Y. Umaidah, “Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat Dengan Silhouette Coefficient di Puskesmas Karangsambung,” *J. Appl. Informatics Comput.*, Vol. 5, No. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.