# ANALISIS PENGELOMPOKKAN DATASET PEMILU 2014 DAN 2019 DPR RI DI KOTA CIREBON MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING*

**Krisna Mulyana [1], Nining Rahaningsih [2], Raditya Danar Dana [3]**
[1] Teknik Informatika, STMIK IKMI CIREBON
[2] Komputerisasi Akuntansi, STMIK IKMI CIREBON
[3] Manajemen Informatika, STMIK IKMI CIREBON
Jalan Perjuangan 10B, Karyamulya Cirebon, Indonesia
*krisnamulyana01@gmail.com*

**ABSTRAK**

Strategi politik merupakan perencanaan matang yang disusun dan dilaksanakan oleh seluruh partai politik dalam menyongsong tahun politik mendatang. Data perolehan suara sangat penting untuk dikaji lebih dalam agar menjadi informasi bermanfaat bagi peluang kemenangan. Namun, belum dilakukan analisis mendalam terhadap data Pemilu yang tersedia di KPU Kota Cirebon, utamanya data suara DPR RI. Oleh karena itu, penelitian ini bertujuan memanfaatkan teknik data mining agar dapat membantu mempercepat proses pengambilan keputusan strategis dengan menerapkan teknik *clustering* dalam mengolah data Pemilu untuk memetakan pola persaingan antar parpol dan koalisi di setiap daerah pemilihan. Metode *Knowledge Discovery in Database* (KDD) dengan Algoritma *K-Means Clustering* digunakan untuk menganalisis data pemilu. Tahapan KDD dimulai dari seleksi, pra-pemrosesan, transformasi data, klusterisasi dan evaluasi kluster. Penelitian juga mencari nilai parameter K, iterasi, dan *measuretype* optimal darip algoritma *K-Means* berdasarkan nilai *DBI*. Hasil penelitian ini memperoleh Nilai DBI dari iterasi 1 dan *numerical measure* sebagai *measure type* terbaik untuk mendapatkan nilai DBI terbaik yaitu 0,334 pada K=3, yang menunjukkan tingkat kompetisi partai yang berbeda di Kota Cirebon, yaitu C1 (tinggi) diperoleh 232 anggota, C2 (menengah) diperoleh 160 anggota, dan C3 (rendah) diperoleh 224 anggota. Hasil ini bermanfaat bagi parpol dalam merumuskan strategi memenangkan pemilu pada basis masa masing-masing.

***Kata kunci:*** Data Pemilu, Data Mining, *Clustering*, *Knowledge in Database* (KDD), *K-Means*, *Davies Bouldin Index*

## 1. PENDAHULUAN

KPU Kota Cirebon merupakan salah satu lembaga pemerintah yang menyelenggarakan pemilihan umum pada tahun 2024 untuk menentukan presiden, gubernur, wakil gubernur, serta anggota DPD, DPR, dan DPRD [1].

Menjelang pemilu tersebut, semua partai politik harus merumuskan strategi politik yang tepat untuk meraih kemenangan khususnya pemilihan legislatif. Diketahui bahwa sangat penting untuk memanfaatkan informasi historis mengenai perolehan suara di setiap daerah untuk mengembangkan strategi yang efektif berdasarkan data pemilu masa lalu [1], [2].

Data hasil Pemilu serentak tahun 2014 dan 2019 lalu dapat menjadi acuan penting dalam perumusan strategi pemilihan berikutnya. Misalnya, dengan menganalisis perolehan suara masing-masing partai di tiap kecamatan dan kelurahan, dapat diketahui wilayah mana yang menjadi basis massa setiap partai. Sebagai big data, kumpulan data pemilu tersebut dapat diolah lebih lanjut untuk menghasilkan pengetahuan strategis bagi partai politik. Dengan demikian, pengelolaan data yang berukuran besar harus dilakukan dengan sangat hati-hati, sebab keseluruhan prosesnya menjadi semakin rumit sejalan dengan meningkatnya volume data [3].

Data mining telah menjadi teknik penting dalam menggali informasi berharga dari sekumpulan data besar yang berisi bahasa alami, dan kemudian menganalisisnya dalam bentuk digital [4].

Data mining bekerja dengan teknik analisis khusus untuk menemukan pola dan aturan penting yang tersembunyi dalam data, sehingga memperoleh informasi dan wawasan baru sesuai dengan analisis yang diperlukan [5].

Salah satu teknik data mining berdasarkan data yang tidak berlabel kelas (*Unsupervised Learning*) adalah *Clustering* [6].

Didalam teknik *Clustering* terdapat algoritma *K-Means* yang hanya mengelompokkan sekumpulan data bersifat numerik [7].

Cara kerjanya berdasarkan kesamaan fitur antar data dan memaksimalkan kesamaan data dalam *cluster* yang sama [8].

*K-Means* adalah algoritma *clusterisasi* paling populer dan umum digunakan karena prosesnya sederhana, cepat, dan fleksibel untuk diterapkan guna mengekstraksi informasi berharga dari big data seperti hasil suara pada data dengan kualitas *cluster* maksimal berdasarkan jarak data [9].

Metode ini memungkinkan untuk menghitung jarak antara titik data individu dengan pusat kelompok, mengidentifikasi kesamaan antar objek data, dan menetapkannya ke kelompok terdekatnya [10].

Pada penelitian terdahulu yang memiliki kaitan dengan teknik *Clustering* telah dilakukan oleh oleh Siti Ramadhani, Dini Azzahra, dan Tomi Z pada tahun 2022 berfokus pada pengelompokan skripsi mahasiswa berdasarkan topik yang kira-kira berkaitan dengan topik skripsi lainnya. Tujuan utamanya adalah

untuk memudahkan menemukan topik yang relevan dengan grup. Beberapa penelitian sebelumnya telah menerapkan teknik seperti *K-Means* dan *K-Medoids* untuk mendapatkan *Indeks Davies-Bouldin* (DBI) dan nilai waktu terbaik. Hasil penelitian menunjukkan bahwa penelitian ini tidak sepenuhnya fokus pada kinerja nilai parameter lain yang mungkin mempengaruhi kinerja algoritma *K-Means*. Namun, terdapat potensi untuk penelitian lebih lanjut dalam konteks metode *Clustering* yang memungkinkan pengguna menentukan nilai parameter dengan lebih jelas untuk meningkatkan kualitas *DBI* [11].

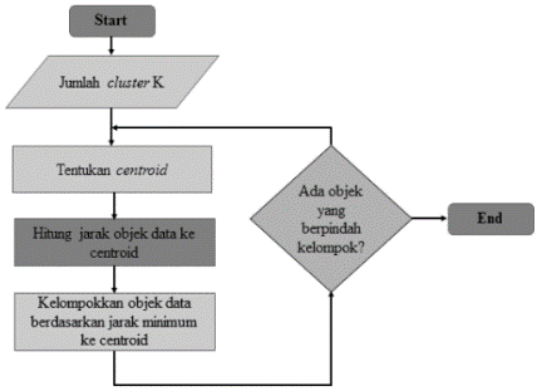
Penelitian selanjutnya dilakukan oleh Hang, STMIK Pekanbaru, Tuah, berjudul "*Implementation of Data Mining for Determining Majors Using K-Means Algorithm in Students of* SMA Negeri 1 Pangkalan Kerinci" persoalaan pada penelitian ini pengelompokkan informasi tentang siswa SMA dalam upaya membantu sekolah menentukan jurusan mana yang memenuhi persyaratan yang telah ditentukan. Algoritma *K-Means* digunakan sebagai metode utama dalam penelitian ini. Tujuan utama penelitian ini adalah menganalisis data siswa SMA untuk mempermudah proses pengelolaan data sekolah. Perlu diperhatikan bahwa dalam penelitian ini tidak ada penekanan pada penentuan jumlah *cluster* (nilai K) yang terbaik maupun pada perbandingan kinerja nilai parameter evaluasi terkait algoritma *K-Means*. Penelitian lebih fokus pada penggunaan algoritma *K-Means* untuk mengelompokkan siswa berdasarkan kriteria yang telah ditentukan, dengan tujuan membantu sekolah mengambil keputusan terkait penentuan jurusan siswa [12].

Penelitian ini memiliki fokus utama dalam pengembangan sistem *Clustering* dengan pemanfaatan dataset pemilu. Tujuannya adalah untuk mengidentifikasi dan memperoleh hasil *Clustering* terbaik dari data pemilu dengan penekanan pada evaluasi karakteristik *cluster* terbaik. Selain itu, penelitian ini juga bertujuan untuk menguji *performance* nilai parameter iterasi, *measure type*, yang secara signifikan dapat meningkatkan nilai *Davies-Bouldin Index* (DBI) dalam algoritma *K-Means*, dan melakukannya pada platform *RapidMiner*. Hasil penelitian ini juga mempunyai kemampuan untuk menentukan jumlah partai berdasarkan *cluster* tertentu, pola persaingan antar partai politik dan koalisi di setiap daerah pemilihan. Hal ini dapat digunakan untuk menyimpulkan informasi seperti dominasi kelompok politik di daerah tertentu. Selain itu, informasi ini bisa menjadi peluang bagi partai untuk lebih aktif mensosialisasikan partainya di daerah yang perolehan suaranya lebih sedikit.

## 2. TINJAUAN PUSTAKA

Peneliti mengkaji jurnal ilmiah sebagai rujukan dalam rangka mendemonstrasikan pemahaman mendalam atas penelitian-penelitian terdahulu terkait topik serupa, serta memberikan pembenaran kuat mengenai kesesuaian dan kebaruan kontribusi dari penelitian ini. Tinjauan pustaka yang komprehensif berperan penting untuk menunjukkan celah penelitian yang hendak diisi oleh penelitian saat ini.

Penelitian serupa yang relevan pernah dilakukan oleh Haris Kurniawan dkk. yang menerapkan metode *Knowledge Discovery in Database* (KDD) meliputi proses *Data Selection*, *Pre-processing*, *Transformation*, Data Mining menggunakan algoritma *K-Means Clustering*, dan Interpretation untuk pengelompokan data. Kasus yang diangkat adalah pengelompokan data besaran Uang Kuliah Tunggal mahasiswa baru ke dalam 5 kluster optimum berdasarkan tingkat kemiripan perhitungan jarak *Euclidean*. Dalam penelitian ini, proses pengelompokan data dilakukan dengan menerapkan algoritma k-means clustering. Tahapan-tahapan dalam proses pengelompokan data menggunakan algoritma k-means ditunjukkan secara visual pada Gambar 1.




Gambar 1. *Flowchart* metode algoritma *K-means*
Sumber : [5]

Hasilnya menunjukkan bahwa implementasi data mining dengan algoritma *K-Means Clustering* telah berhasil dan efektif membantu pengelompokan nilai Uang Kuliah Tunggal calon mahasiswa ke dalam 5 kluster serupa karakteristiknya [5].

Dalam penelitian lain, algoritma *K-Means* telah dilakukan untuk mengelompokkan fasilitas sekolah menggunakan *Davis-Bouldin Index* (DBI) yang dievaluasi menggunakan data yang diperoleh dari situs resmi Badan Pusat Statistik (BPS) RI. Penelitian ini bertujuan untuk mengoptimalkan nilai *DBI* dalam menentukan jumlah *cluster* desa dengan fasilitas sekolah dan tingkat pendidikan. Hasil optimasi menunjukkan bahwa metode *K-Means* dengan jenis pengukuran *mixed gauge* memberikan nilai *DBI* terbaik yaitu 0,168 sehingga membentuk dua *cluster*. Evaluasi dilakukan dengan membandingkan tiga jenis pengukuran: ukuran campuran, jarak *Bregman divergence-Mahalanobis, dan jarak Euclidean kuadrat divergensi Bregman*. Di antara ketiga jenis pengukuran tersebut, nilai *DBI* terbaik diperoleh dengan menggunakan metode pengukuran campuran yang secara efisien menghasilkan dua *cluster*. Hasil ini membantu lebih memahami efektivitas algoritma *K-Means* dalam konteks pengelompokan fasilitas

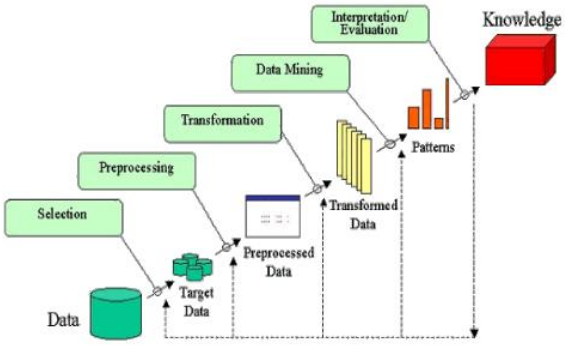
sekolah. Nilai *DBI* yang lebih rendah mencerminkan hasil *Clustering* yang lebih terbaik [13].

Berbeda dalam penelitian sebelumnya, dalam penelitian ini dilakukan dengan memilih jenis metrik/ukuran jarak yang akan digunakan algoritma *Clustering* untuk menemukan data tetangga (*neighbor*) terdekat dari setiap data. Metrik jarak ini yang nantinya dipakai untuk menghitung jarak/ketidaksamaan *(dissimilarity)* antar data. Penelitian ini melakukan percobaan dari 3 *measure type Mixed Measures, Numerical Measures,* dan *Bregman Divergences. Mixed Measures* : Ukuran campuran digunakan untuk menghitung jarak dalam kasus atribut nominal dan numerik .*Numerical Measures* : Untuk kasus hanya atribut numerik, berbagai pengukuran jarak yang berbeda dapat digunakan untuk menghitung jarak pada atribut numerik. *Bregman Divergences* : *Bregman Divergences* adalah jenis ukuran "kedekatan" yang lebih umum yang tidak memenuhi ketidaksamaan segitiga atau simetri. Ketidaksamaan segitiga adalah sifat bahwa jarak langsung antara dua titik haruslah lebih pendek atau sama dengan jarak tidak langsung melewati titik ketiga. Dengan begitu, dapat diketahui perbandingan jenis pengukuran (*measure type*) mana yang paling baik digunakan untuk menganalisis data pemilu.

Selain itu, pada penelitian lain yang memiliki kaitan dengan algoritma *K-Means* telah dilakukan oleh Rubangiya, Tuti Hartati, dan Yudhistira Arie Wijaya dalam analisis *Clustering* pada data lalu lintas jaringan, dengan pendekatan *Knowledge Discovery in Database* (KDD). Penelitian ini menggunakan evaluasi *Davies Bouldin Index (DBI)* yang merupakan metrik yang dapat digunakan untuk menilai kualitas hasil pengelompokan data ke dalam cluster. DBI mengukur perbandingan antara seberapa dekat jarak data dalam satu kelompok dengan seberapa jauh jarak antarkelompok yang berbeda. Semakin kecil nilai DBI, maka semakin baik kualitas pengelompokan yang dilakukan. Kohesi cluster dihitung berdasarkan jarak rata-rata antara anggota cluster dengan pusat cluster. Dengan demikian, DBI dapat menunjukkan tingkat optimalitas jumlah cluster hasil pengelompokan. Dalam pencarian nilai *Davies-Bouldin Index* (DBI) terbaik, dilakukan pengujian dari jumlah *cluster* (K) mulai dari 2 hingga 10. Hasil penelitian menunjukkan bahwa nilai *Davies-Bouldin Index (DBI)* terbaik, yaitu 0,028, didapatkan dari model proses yang tidak menerapkan operator normalisasi dengan jumlah klaster K = 2. Dalam menentukan klaster terbaik di antara klaster 0 dan klaster 1, dilakukan evaluasi dengan menggunakan *Performance Centroid Distance* dari masing-masing klaster dan penerapan *DBI*. Penggunaan nilai rata-rata within centroid distance dapat dijadikan acuan dalam menentukan klaster yang paling optimal [14].

## 3. METODE PENELITIAN

Metode penelitian yang digunakan untuk menganalisa pemilu dalam penelitian ini adalah analisis *Knowledge Discovery in Database* (KDD) dengan menerapkan Algoritma *K-Means Clustering*. KDD *(Knowledge Discovery in Databases)* merupakan metode untuk mengekstraksi wawasan/pengetahuan baru dari dataset yang sudah ada dan mengidentifikasi korelasi/hubungan yang ada antar tabel *database* yang terkait. Tujuan utamanya adalah untuk menghasilkan wawasan baru dari serangkaian proses data mining untuk menginformasikan pengambilan keputusan dan strategi [15].




Gambar 2. Proses metode penelitian KDD
Sumber : [5]

Seperti terlihat pada Gambar 2, Metode penelitian dilakukan dalam enam tahap untuk memperoleh hasil. Tahapan penelitian menggunakan proses KDD *(knowledge discovery in database)* dan proses yang dilakukan dimulai dengan pemilihan data, *preprocessing data*, transformasi data, data mining menggunakan *K-Means Clustering*, evaluasi, dan terakhir melakukan analisis terhadap *cluster* tersebut [5].

### 3.1. Data

Dataset yang digunakan dalam penelitian ini merupakan data perolehan suara calon anggota Dewan Perwakilan Rakyat Republik Indonesia (DPR RI) Kota Cirebon untuk Daerah Pemilihan (Dapil) VIII. Data perolehan suara ini diperoleh dari Kantor Komisi Pemilihan Umum (KPU) Kota Cirebon sebagai instansi yang berwenang. Data awal dari KPU Kota Cirebon berupa dokumen perolehan suara DAA dalam bentuk PDF yang kemudian dikonversi atau diubah ke dalam format Excel agar lebih mudah diolah. Data yang digunakan pada penelitian ini adalah perolehan suara pemilu anggota Dewan Perwakilan Rakyat Republik Indonesia (DPR RI) Kota Cirebon untuk Daerah Pemilihan (Dapil) VIII). Dataset ini berjumlah 616 baris yang merupakan gabungan *record* data perolehan suara Pemilihan Umum (Pemilu) DPR RI Kota Cirebon pada tahun 2014 dan 2019.

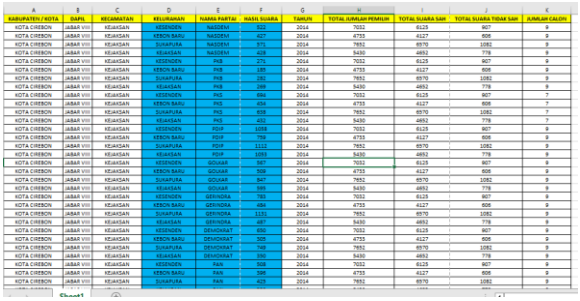
Tabel 1. Atribut *default* dari dataset pemilu DPR RI Kota Cirebon

| No. | Atribut | Tipe Data | Ket |
|---|---|---|---|
| 1. | Kabupaten / Kota | Nominal | |
| 2. | Dapil | Nominal | |
| 3. | Kecamatan | Nominal | |
| 4. | Kelurahan | Nominal | |
| 5. | Nama Partai | Nominal | |
| 6. | Hasil Suara | Integer | |
| 7. | Tahun | Integer | |
| 8. | Total Jumlah Suara | Integer | |
| 9. | Total Suara Sah | Integer | |
| 10. | Total Suara Tidak Sah | Integer | |
| 11. | Jumlah Calon | Integer | |

Pada tabel 1 menunjukkan 11 kolom atribut pada dataset yang merepresentasikan informasi dataset pemilu.

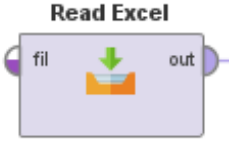
### 3.2. Seleksi Data

Pada tahap ini dilakukan seleksi data dengan tujuan untuk memilih data yang akan digunakan dalam proses pengelompokan (*Clustering*) melalui *Microsoft Excel*. Proses seleksi data dilakukan secara manual dengan melihat *record* dari masing-masing atribut. Atribut yang memiliki nilai yang tetap tidak digunakan karena dapat mempengaruhi hasil secara dominan terhadap atribut tersebut. Pemilihan atribut ini dilakukan berdasarkan tujuan dengan mempertimbangkan beberapa atribut wajib yang harus dipilih yaitu kelurahan, nama partai, dan hasil suara.



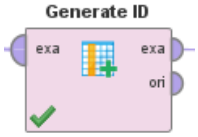
Gambar 3. Seleksi data pemilu berdasarkan kebutuhan

Gambar 3 menampilkan hasil seleksi data yang menunjukkan 3 atribut yang menyesuaikan kebutuhan berdasarkan tujuan, yaitu kelurahan, nama partai, dan hasil Suara. Langkah pertama yang harus dilakukan di *RapidMiner* adalah memasukkan data pemilu yang sudah dilakukan seleksi data dalam format *Excel (.xlsx)* ke dalam operator *read excel*.




Gambar 4. Operator *read excel* pada *rapidminer*

Gambar 4 merupakan operator *read Excel* yang merupsakan operator yang didalamnya memiliki proses untuk membaca data dari file *Excel*. Karena dataset pemilu tidak memiliki atribut yang dapat berperan sebagai *identifier* unik untuk setiap *record*.




Gambar 5. Operator *Generate ID* pada *RapidMiner*

Terlihat bahwa pada gambar 5 merupakan operator *Generate ID* yang didalamnya memiliki proses penambahan kolom baru yang berperan sebagai identifier unik untuk setiap *record* (baris) pada dataset *input*. Dari hasil pembacaan operator *generate id* didapat informasi sebagai berikut.

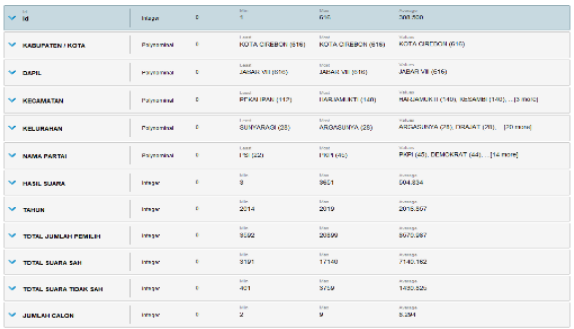
Tabel 2. Statistik dataset

| No. | Uraian | Keterangan |
|---|---|---|
| 1. | *Record* | 616 |
| 2. | *Special Attribute* | 1 |
| 3. | *Reguler Atribute* | 11 |
| 4. | *Attribute :* | |
| | Kabupaten / Kota | Nominal |
| | Dapil | Nominal |
| | Kecamatan | Nominal |
| | Kelurahan | Nominal |
| | Nama Partai | Nominal |
| | Hasil Suara | Integer |
| | Tahun | Integer |
| | Total Jumlah Suara | Integer |
| | Total Suara Sah | Integer |
| | Total Suara Tidak Sah | Integer |
| | Jumlah Calon | Integer |

Statistik data pemilu pada tabel 2 merupakan hasil dari proses seleksi data dengan melakukan penambahan atribut id pada dataset sebagai identitas setiap *record.*

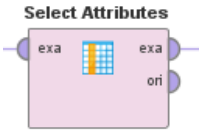
### 3.3. Pra-Pemrosesan Data

Pada tahap *preprocessing* data, telah dilakukan analisis awal terhadap kemungkinan adanya data yang missing atau memiliki nilai yang tidak konsisten.



Gambar 6. *Result* statistik dataset

Berdasarkan statistik dataset pada gambar 6, dilakukan pemeriksaan lebih detail per record secara manual, diketahui bahwa dataset mengandung data yang berulang (duplikasi) atau bernilai tetap. Untuk itu operator *select attribute* dipilih guna menghilangkan data yang tidak akan digunakan.




Gambar 7. Operator *select atribute* pada *rapidminer*

Gambar diatas adalah operator *select attribute* yang digunakan untuk menghilangkan atribut yang mengandung data duplikasi atau bernilai tetap. Parameter operator *select atribut* yang digunakan ditunjukkan pada tabel berikut.

Tabel 3. Parameter dan atribut yang dipilih pada operator *select atribute*

| No. | Parameter | Isi |
|---|---|---|
| 1. | *Attribut Filter Type* | *subset* |
| 2. | *Attributes* | *Select attributes …* |
| | | *Id*, Kelurahan, Nama Partai, dan Hasil Suara |

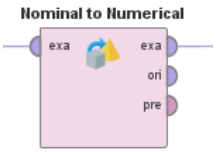
Tabel 3 memperlihatkan parameter yang diterapkan pada operator *Select attributes.* Dari hasil operator *Select Atribute* didapat informasi sebagai berikut.


Gambar 8. *Result* operator *select atribute*

Berdasarkan gambar di atas, dapat dilihat hasil dari proses seleksi atribut (*select attribute*) di mana 4 atribut dipilih sedangkan atribut lainnya dihilangkan.

### 3.4. Transformasi Data

Karena data pada dataset yang digunakan bertipe data nominal, sedangkan algoritma *Clustering K-Means* mensyaratkan tipe data numerik, maka perlu dilakukan transformasi untuk mengonversi data nominal menjadi numeric dengan menerapkan operator *Nominal to Numerical*.



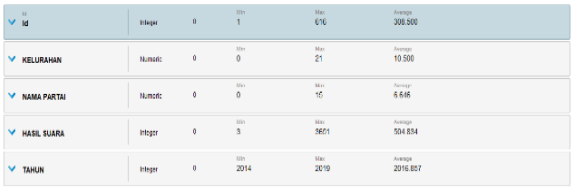

Gambar 9. Operator *nominal to numerical* pada *rapidminer*

Gambar di atas menunjukkan operator *Nominal to Numerical* pada *RapidMiner*. Operator *Nominal to Numerical* memiliki fungsi untuk mengubah atribut dengan jenis data nominal agar menjadi data numerik. Detail parameter yang digunakan pada operator *Nominal to Numerical* ditampilkan melalui tabel berikut.

Tabel 4. Rincian parameter dan atribut yang dipilih pada operator *Nominal to Numerical*

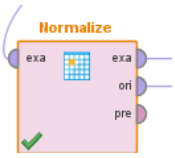
| No. | Parameter | Isi |
|---|---|---|
| 1. | Attribut filter type | *subset* |
| 2. | Attributes | *Select attributes …* |
| | | Kelurahan, Nama Partai |
| 3. | Coding type | Unique Integers |

Tabel 4 memperlihatkan parameter yang diterapkan pada operator *Nominal to Numerical.* Hasil dari implementasi operator *Nominal to Numerical* ditunjukkan pada gambar di bawah.




Gambar 10. Hasil proses operator *Nominal to Numerical*

Berdasarkan gambar diatas, dapat diketahui bahwa telah dilakukan proses transformasi data di mana seluruh data bertipe nominal telah diubah menjadi bentuk numerik. Dalam penelitian ini, normalisasi digunakan untuk mencegah dominasi atribut numerik dengan rentang nilai yang luas (seperti hasil suara) terhadap perhitungan jarak pada algoritma *K-means*. Dengan normalisasi, atribut non-numerik seperti nama partai dan kelurahan dapat turut berpengaruh terhadap pengelompokan data pemilu yang dilakukan *K-means*.




Gambar 11 Operator *Normalize* pada *RapidMiner*

Gambar 11 merupakan operator *normalize* untuk melakukan normalisasi dalam menstandarisasi nilai-nilai ini ke skala yang sama. Parameter yang digunakan pada operator *normalize* ditunjukkan pada tabel di bawah ini.

Tabel 5 Parameter dan atribut pada operator *Normalize*

| *No.* | *Parameter* | *Isi* |
|---|---|---|
| 1. | *Attribut filter type* | *subset* |
| 2. | *Attributes* | *Select attributes …* |
| | | Hasil Suara |
| 3. | *method* | *range transformation* |
| 4. | *Min* dan *Max* | *Default* : *Min* (0.0) *Max* (1.0) |

Tabel 5 memperlihatkan parameter yang diterapkan pada operator *normalize*. Terdapat 1 atribut yaitu hasil suara yang dilakukan normalisasi data. Hasil dari proses operator *normalize* ditunjukkan pada gambar dibawah ini.

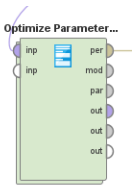
| Row No. | id | HASIL SUARA | KELURAHAN | NAMA PARTAI |
|---|---|---|---|---|
| 1 | 1 | 0.142 | 0 | 0 |
| 2 | 2 | 0.116 | 1 | 0 |
| 3 | 3 | 0.156 | 2 | 0 |
| 4 | 4 | 0.117 | 3 | 0 |
| 5 | 5 | 0.073 | 0 | 1 |
| 6 | 6 | 0.050 | 1 | 1 |
| 7 | 7 | 0.076 | 2 | 1 |
| 8 | 8 | 0.073 | 3 | 1 |
| 9 | 9 | 0.189 | 0 | 2 |
| 10 | 10 | 0.118 | 1 | 2 |
| 11 | 11 | 0.174 | 2 | 2 |
| 12 | 12 | 0.118 | 3 | 2 |
| 13 | 13 | 0.289 | 0 | 3 |

Gambar 12. Hasil proses operator *Nominal to Numerical*

Gambar 12 menunjukkan hasil transformasi normalisasi data menggunakan operator *normalize*. Dengan transformasi normalisasi, atribut hasil suara memiliki skala yang sama dan tidak ada nilai yang mendominasi.

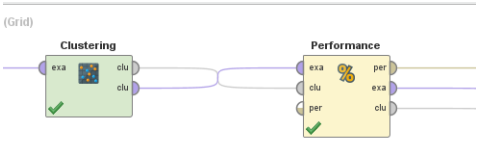
### 3.5. Data Mining

Pada langkah data mining, digunakan operator *Optimize Parameters (Grid)*. Operator ini secara efisien menghitung nilai parameter terbaik berdasarkan semua atribut dengan memanfaatkan struktur data. Dengan demikian, proses pencarian nilai optimal dari algoritma *Clustering* menjadi lebih cepat dan mudah, tidak diperlukan operator *multiply* untuk mencoba satu per satu.




Gambar 13. Operator *optimize parameters (grid)* pada *rapidminer*

Gambar diatas merupakan Operator *Optimize Parameters (Grid)* yang mengotomatisasi pencarian parameter terbaik sehingga analisis *Clustering* bisa dilakukan secara efektif.




Gambar 14. Sub proses pemodelan data mining

Pada gambar 14 merupakan sub proses Operator *Optimize Parameters (Grid)* yang melakukan proses analisis data mining dengan memilih algoritma *K-Means* dan *Cluster Distance Performance* sebagai evaluasi metrik dalam bentuk pencarian nilai *Davies Bouldin*.


Gambar 15. Operator *k-means* pada *rapidminer*

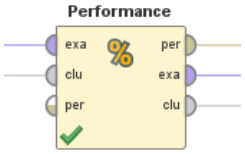
Gambar di atas menunjukkan operator *K-Means Clustering* pada *RapidMiner*. Operator ini melakukan pengelompokan data ke dalam kluster dengan algoritma *K-Means.*

Tabel 7. Parameter operator algoritma *k-means*

| **No.** | **Parameter** | **Isi** |
|---|---|---|
| 1. | *Clustering. K* | *Min* : 2, *Max* : 25, *Steps* : 25, *Scale* : Linear |
| 2. | *Clustering.max _runs* | *Min* : 1, *Max* : 100, *Steps* : 100, *Scale* : Linear |
| 3. | *Clustering.measure_types* | *Mixed Measures, Numerical Measures, Bregman, Divergences* |

Parameter *k-means* pada tabel 7 *clustering* dijalankan dengan beragam kombinasi parameter untuk mengeksplorasi pembentukan kluster optimal pada data, meliputi variasi jumlah kluster 2 hingga 25, iterasi 1 sampai 100, dan tiga metode pengukuran jarak antar data yaitu *Mixed Measures*, *Numerical Measures*, dan *Bregman Divergences* dengan nilai parameter lain menggunakan default *RapidMiner*.

Evaluasi yang dilakukan terhadap hasil eksperimen pengelompokan data (*Clustering*) pada dataset menggunakan operator *Cluster Distance Performance*.



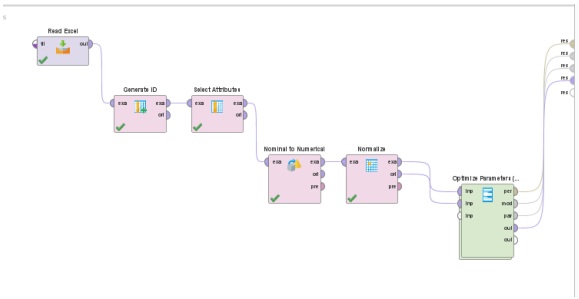

Gambar 16. Operator *cluster distance performance* pada *rapidminer*

Gambar 16 menunjukkan operator cluster distance performance yang dipakai untuk mengukur nilai evaluasi pengelompokan berdasarkan *indeks Davies-Bouldin (DBI).*

Tabel 8. Parameter Operator *Cluster Distance Performance*

| No. | Parameter | Isi |
|---|---|---|
| 1. | *Main Criterion* | *Davies Bouldin* |
| 2. | *Normalisasi* | *True* |
| 3. | *Maximize* | *True* |

Tabel diatas merupakan Parameter pada operator *Cluster Distance Performance* yang digunakan dalam menemukan nilai evaluasi optimal pada algoritma *K-Means.* Gambar model proses sampai dengan proses Data Mining ditunjukkan seperti gambar dibawah ini.




Gambar 17. Model proses sampai dengan langkah data mining

### 3.6. Evaluasi

Hasil evaluasi analisis data pemilu menggunakan clustering algorima sebanyak 24 kali eksperimen pengelompokan data dilakukan dengan nilai K mulai dari 2 hingga 25. Eksperimen bertujuan mencari *kluster* optimal dengan nilai *Davies-Bouldin Index (*DBI*)* terbaik. Eksperimen lain adalah uji coba kombinasi parameter jumlah iterasi (1 hingga 100) dan tipe pengukuran (*Mixed Measures, Numerical Measures, Bregman Divergences*) untuk mendapatkan nilai DBI terkecil. Hasil keseluruhan eksperimen ditampilkan pada gambar 18 , yang menunjukkan nilai DBI untuk setiap kombinasi parameter dan *kluster* yang dihasilkan.

**ParameterSet**

```
Parameter set:

Performance:
PerformanceVector [
-----Avg. within centroid distance: 7.418
-----Avg. within centroid distance_cluster_0: 8.213
-----Avg. within centroid distance_cluster_1: 5.344
-----Avg. within centroid distance_cluster_2: 8.075
*****Davies Bouldin: 0.334
]
Clustering.k    = 3
Clustering.max_runs     = 1
Clustering.measure_types        = NumericalMeasures
```

Gambar 18. Hasil parameter set operator *Optimize Parameters (Grid)*

Gambar 18 menunjukkan beberapa informasi penting terkait performa hasil proses data mining pengelompokan (*clustering*) menggunakan algoritma *K-Means* pada penelitian ini. Diperoleh bahwa jumlah kluster optimal adalah 3 (K=3), ditunjukkan dengan nilai *Davies-Bouldin Index* (DBI) terbaik yaitu 0,334. Oleh karena itu, model klusterisasi dengan 3 kluster tersebut dapat dianggap sebagai model paling optimal. Jenis pengukuran (*measure type*) yang digunakan adalah *NumercalMeasures*, yang membantu algoritma *K-Means* mencapai nilai optimasi terbaik. Iterasi terbaik untuk mencapai nilai K optimal juga ditunjukkan, yaitu pada iterasi ke-1. Secara keseluruhan hasil ini menunjukkan bahwa proses data mining dengan algoritma *K-Means* telah berhasil dilakukan dengan baik sesuai tujuan penelitian.

**Cluster Model**

```
Cluster 0: 232 items
Cluster 1: 160 items
Cluster 2: 224 items
Total number of items: 616
```

Gambar 18. *Cluster Model*

Model klusterisasi pada penelitian ini menghasilkan 3 kluster, yaitu kluster 0 dengan 232 anggota, *cluster* 1 dengan 160 anggota, dan *cluster* 2 dengan 224 anggota. Ketiga kluster ini menunjukkan pola perolehan suara yang berbeda pada pemilu legislatif di Kota Cirebon.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Penerapan *Clustering* Dataset Pemilu 2014 dan 2019 DPR RI Kota Cirebon Menggunakan Algoritma *K-Means*

Pada penelitian ini, telah dilakukan penerapan pengelompokan (*clustering*) dataset perolehan suara di KPU Kota Cirebon dengan menggunakan algoritma *K-Means*. Metode analisis yang digunakan adalah metode *Knowledge Discovery in Database (KDD)* dan perangkat lunak yang digunakan adalah RapidMiner versi 10.3. Tahapan analisis terdiri dari beberapa tahap yaitu seleksi data dengan melakukan penambahan atribut id dengan operator *generate id*, pra-pemrosesan data dengan operator *select attribute*, transformasi data dengan operator *nominal to numerical* dan *normalize*, penerapan operator *K-Means* dan *Cluster Distance Performance* untuk proses data mining, pengukuran evaluasi DBI untuk menentukan kluster optimal, serta evaluasi dengan menganalisis hasil kluster.

### 4.2. Hasil Evaluasi Nilai *Davies Bouldin Index*

Hasil evaluasi *DBI* diperoleh nilai terbaik yaitu pada model dengan 3 *cluster* (K=3) sebesar 0,334. Dengan demikian, berdasarkan nilai *DBI* tersebut, model *K-Means* dengan 3 kluster (K=3) dinilai paling optimal dan mampu melakukan pengelompokkan data secara baik. Hasil ini sejalan dengan penelitian [7], [13] dalam kajiannya menyatakan bahwa Nilai *DBI* yang semakin mendekati 0 menunjukkan pengelompokan data yang baik dengan *cluster* yang kompak dan terpisah satu sama lain. Sehingga nilai *DBI* yang rendah mengindikasikan model *clustering* yang optimal.

### 4.3. Hasil Evaluasi Nilai Iterasi

Dalam proses data mining klusterisasi, dilakukan pemodelan dengan berbagai parameter iterasi mulai dari 1 hingga 100, dengan tujuan mencari nilai iterasi terbaik untuk performa optimal algoritma *K-Means*. Parameter *Max Runs* ditetapkan agar proses *clustering* dilakukan berulang pada setiap nilai K untuk mendapatkan hasil iterasi terbaik. Setelah proses training iteratif tersebut, didapatkan bahwa iterasi ke-1 merupakan nilai terbaik, dimana performa klusterisasi menghasilkan pengelompokan data paling akurat dengan nilai evaluasi teroptimum, yaitu berupa nilai *Davies-Bouldin Index* paling minimum dibanding iterasi lain.

### 4.4. Hasil Evaluasi *Measure Type*

Dalam proses data mining klusterisasi dataset pemilu menggunakan algoritma *K-Means*, telah dilakukan optimasi nilai parameter *measure type* yang menentukan jenis pengukuran untuk menghitung ketidakmiripan antar data. Terdapat 3 jenis *measure type* yang diujicobakan yaitu *Mixed Measures, Numerical Measures, dan Bregman Divergences.* Setelah dilakukan percobaan dengan menerapkan masing-masing jenis *measure type* tersebut untuk proses klusterisasi, didapatkan hasil bahwa *Numerical Measures* memberikan performa paling optimal ditinjau dari nilai evaluasi *Davies Bouldin Index* dan iterasi terbaik yang dihasilkan.

### 4.5. Hasil Analisis Pengelompokkan Dataset Pemilu dengan Algoritma *K-Means*

Berdasarkan proses pengelompokan K-Means yang telah dilaksanakan, didapatkan pengetahuan bahwa terbentuk 3 klaster perolehan suara yang terdiri dari:

a. Kluster 0, yang terletak di Kecamatan Kejaksan, didominasi oleh PDIP dengan urutan partai PDIP, Gerindra, PKS, dan NasDem. Tingkat kompetisi yang tinggi menunjukkan persaingan ketat antar partai di wilayah ini, sementara PDIP mendominasi kluster ini, mencerminkan kekuatan basis pemilihnya di Kecamatan Kejaksan.
b. Kluster 1, yang sebagian besar berasal dari Kecamatan Harjamukti dan Lemahwungkuk, didominasi oleh PDIP dan Gerindra dengan urutan partai PDIP, Gerindra, PKS, dan Golkar. Meskipun tingkat kompetisi dalam kluster ini tergolong menengah, PDIP dan Gerindra tetap mempertahankan dominasi mereka, menunjukkan pengaruh yang kuat di wilayah Kecamatan Harjamukti dan Lemahwungkuk, dengan PDIP memegang posisi unggul.
c. Kluster 2, yang sebagian besar pendukungnya berasal dari Kecamatan Pekalipan dan Kesambi, didominasi oleh PDIP, Gerindra, dan PKS, dengan urutan partai PDIP, Gerindra, PKS, dan NasDem. Tingkat kompetisi dalam kluster ini tergolong rendah, menandakan bahwa persaingan antar partai tidak begitu ketat di wilayah tersebut.

## 5. KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan, dapat disimpulkan bahwa penerapan metode KDD dengan Algoritma *K-Means* untuk klusterisasi data pemilu berhasil dilakukan dan memberikan hasil optimal dengan 3 kluster terbaik ditinjau dari nilai *DBI* 0,334. Iterasi ke-1 juga diperoleh untuk menentukan nilai optimasi kluster. Selain itu, Numerical Measures sebagai *measure type* terbukti menghasilkan nilai *DBI* terbaik. Dengan demikian, dapat disimpulkan bahwa parameter algoritma yang diusulkan telah mampu mengekstraksi pola data pemilu secara optimal ke dalam 3 kelompok yaitu Kluster pertama (0) merupakan kelompok partai politik dengan tingkat kompetisi tinggi (232 Anggota), Kluster kedua (1) merupakan kelompok partai politik dengan tingkat kompetisi menengah (160 anggota), dan Kluster ketiga (2) merupakan kelompok partai politik dengan dengan tingkat kompetisi rendah (224 anggota). Kluster 0, yang terletak di Kecamatan Kejaksan, didominasi oleh PDIP dengan urutan partai PDIP, Gerindra, PKS, dan NasDem. Kluster 1, yang sebagian besar berasal dari Kecamatan Harjamukti dan Lemahwungkuk, didominasi oleh PDIP dan Gerindra dengan urutan partai PDIP, Gerindra, PKS, dan Golkar. Kluster 2, yang sebagian besar pendukungnya berasal dari Kecamatan Pekalipan dan Kesambi, didominasi oleh PDIP, Gerindra, dan PKS, dengan urutan partai PDIP, Gerindra, PKS, dan NasDem. Informasi hasil klusterisasi ini dapat dimanfaatkan untuk mengetahui pola persaingan antar partai politik dan dan koalisi di setiap daerah pemilihan. Untuk penelitian selanjutnya disarankan mengembangkan dengan algoritma lain seperti *K-Medoids*, menambahkan fitur/atribut agar klusterisasi lebih spesifik, serta analisis performa seperti jenis pengukuran parameter lain untuk mendapatkan hasil yang lebih baik.

## DAFTAR PUSTAKA

[1] P. H. Syahda and A. Rafni, "Strategi Calon Legislatif Partai Gerindra dalam Memenangkan Pemilu Legislatif Tahun 2019 di Kota Padang," *Journal of Civic Education*, vol. 4, no. 1, pp. 66–72, 2021, doi: 10.24036/jce.v4i1.419.

[2] D. Alfatah, S. Tinggi, and I. A. Bengkulu, "Application of the K-Means Clustering Algorithm in Mapping the Regional Voter Strategy for the Legislative Candidates for the DPR RI Penerapan Algoritma K-Means Clustering dalam Memetakan Strategi Daerah Pemilih pada Calon Legislatif DPR RI," *Jurnal Komitek*, vol. 1, no. 2, pp. 435–443, 2021, [Online]. Available: https://doi.org/10.53697/jkomitek.v1i2

[3] M. Sholeh, E. K. Nurnawati, and U. Lestari, "Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 8, no.

1, pp. 10–21, 2023, doi: 10.14421/jiska.2023.8.1.10-21.

[4] H. Kim and M. Park, "Discovering fashion industry trends in the online news by applying text mining and time series regression analysis," *Heliyon*, vol. 9, no. 7, p. e18048, 2023, doi: 10.1016/j.heliyon.2023.e18048.

[5] Haris Kurniawan, Sarjon Defit, and Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *Journal of Applied Computer Science and Technology*, vol. 1, no. 2, pp. 80–89, 2020, doi: 10.52158/jacost.v1i2.102.

[6] H. Gunawan and V. Purwayoga, "Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Potensi Penyebaran Virus Corona Di Kota Cirebon," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 1, pp. 1–8, 2022, doi: 10.32736/sisfokom.v11i1.1316.

[7] N. Puspitasari, G. Lempas, H. Hamdani, H. Haviuddin, and A. Septiarini, "Perbandingan Algoritma K-Means dan Algoritma K-Medoids Pada Kasus Covid-19 di Indonesia," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 4, pp. 2015–2027, 2023, doi: 10.47065/bits.v4i4.2994.

[8] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[9] N. Puspitasari, H. Haviluddin, and F. U. J. Helmi Puadi, "Klasterisasi Wilayah Penghasil Tanaman Lada Menggunakan Algoritma K-Means," *Indonesian Journal of Computer Science*, vol. 11, no. 3, pp. 1001–1013, 2022, doi: 10.33022/ijcs.v11i3.3104.

[10] T. M. Ghazal *et al.*, "Performances of k-means clustering algorithm with different distance metrics," *Intelligent Automation and Soft Computing*, vol. 30, no. 2, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[11] S. Ramadhani, D. Azzahra, and T. Z, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis," *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 13, no. 1, pp. 24–33, 2022, doi: 10.31849/digitalzone.v13i1.9292.

[12] Y. Irawan, "Implementation of Data Mining for Determining Majors Using K-Means Algorithm in Students of Sma Negeri 1 Pangkalan Kerinci," *Journal of Applied Engineering and Technological Science*, vol. 1, no. 1, pp. 17–29, 2019, doi: 10.37385/jaets.v1i1.18.

[13] Y. A. Wijaya, D. A. Kurniady, E. Setyanto, W. S. Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM Journal*, vol. 10, no. 3, pp. 1099–1103, 2021, doi: 10.18421/TEM103-13.

[14] T. Hartati and Y. Arie Wijaya, "Analisis Data Lalu Lintas Jaringan Di Kantor Cangehgar Cyber Operation Center Menggunakan Algritma K-Means Network Traffic Data Analysis At Cangehgar Cyber Operation Center Office Using K-Means Algorithm," *Jurnal Ilmiah NERO*, vol. 7, no. 1, pp. 75–84, 2022.

[15] A. Ardhianto, B. Relawanto, and A. Wibowo, "Algoritme K-Means dalam Pengelompokan Kantor Cabang untuk Optimalisasi Manajemen Perbankan," *Jurnal Telematika*, vol. 15, no. 2, pp. 71–76, 2020, [Online]. Available: https://journal.ithb.ac.id/telematika/article/view/351