



# ANALISIS ALGORITMA K-MEANS CLUSTERING DALAM PENGELOMPOKAN DAERAH PENYEBARAN PENYAKIT DEMAM BERDARAH DENGUE

**Ahmad Khusaeri**

Program Studi Sistem Informasi  
Fakultas Ilmu Komputer  
Universitas Singaperbangsa Karawang

[ahmad.khusaeri@cs.unsika.ac.id](mailto:ahmad.khusaeri@cs.unsika.ac.id)

**Received:** 2024-11-18. **Revised:** 2024-12-15. **Accepted:** 2024-12-28.  
**Issue Period:** Vol.8 No.2 (2024), Pp. 434-449

**Abstrak:** Demam Berdarah Dengue merupakan jenis penyakit menular yang disebabkan oleh virus dengue dan disebarluaskan oleh spesies nyamuk *Aedes aegypti*. Demam Berdarah Dengue dapat mengakibatkan suhu tubuh penderita menjadi sangat tinggi dan pada umumnya disertai demam tinggi, mual/muntah, sakit kepala, mimisan, nyeri pada otot serta persendian, bahkan dapat menyebabkan kematian. Objek penelitian ini adalah algoritma K-Means Clustering dengan tempat penelitian di Kabupaten Karawang. Data yang telah diperoleh dari Dinas Kesehatan Kabupaten Karawang. Metodologi penelitian yang digunakan adalah metode CRISP-DM (Cross Industry Standard Process for Data mining), yang terdiri dari 6 fase yaitu pemahaman bisnis (business understanding), pemahaman data (data understanding), pengolahan data (data preparation), pemodelan (modeling), penyebaran (deployment), evaluasi (evaluation). Hasil yang didapat untuk ukuran cluster ( $k$ ) = 8, 9 dan 10, nilai total SSE-nya relatif sama. Untuk  $k$  yang lain diperoleh nilai total SSE yang berbeda. Hal ini berpengaruh pada jumlah iterasi dari algoritma K-Means. Jumlah iterasi pada algoritma K-Means tidak terlalu berpengaruh terhadap waktu eksekusi karena algoritma K-Means memiliki kompleksitas linear terhadap ukuran obyek data. Dari nilai K-2 sampai 10 yang memiliki nilai SSE terkecil adalah terdapat pada nilai K = 10 sebesar 151,48. Selain itu, nilai SSE dari setiap K, memiliki nilai SSE yang semakin menurun dari setiap K. Jumlah iterasi terkecil terdapat pada nilai K = 2 dengan jumlah iterasi sebesar 7. Sedangkan jumlah iterasi terbesar terdapat pada nilai K = 7 dengan jumlah iterasi sebesar 36.

**Kata kunci:** Clustering, K-Means, CRIPS-DM, SSE

**Abstract:** Dengue Hemorrhagic Fever is a type of infectious disease caused by the dengue virus and spread by the *Aedes aegypti* mosquito species. Dengue Hemorrhagic Fever can cause the patient's body temperature to become very high and is generally accompanied by high fever, nausea/vomiting, headaches, nosebleeds, muscle and joint pain, and can even cause death. The object of this research is the K-Means Clustering algorithm with the research location in Karawang Regency. Data obtained from the Karawang Regency Health Office. The research methodology used is the CRISP-DM (Cross Industry Standard Process for



DOI: 10.52362/jisicom.v8i2.1795

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

*Data mining) method, which consists of 6 phases, namely business understanding, data understanding, data preparation, modeling, deployment, evaluation. The results obtained for cluster sizes ( $k$ ) = 8, 9 and 10, the total SSE value is relatively the same. For other  $k$ , different total SSE values are obtained. This affects the number of iterations of the K-Means algorithm. The number of iterations in the K-Means algorithm does not have much effect on execution time because the K-Means algorithm has linear complexity to the size of the data object. From the value of  $K$ -2 to 10, the smallest SSE value is at the value of  $K = 10$  of 151.48. In addition, the SSE value of each  $K$  has a decreasing SSE value for each  $K$ . The smallest number of iterations is at the value of  $K = 2$  with the number of iterations of 7. While the largest number of iterations is at the value of  $K = 7$  with the number of iterations of 36.*

**Keywords:** Clustering, K-Means, CRIPS-DM, SSE

## I. PENDAHULUAN

Demam Berdarah Dengue merupakan jenis penyakit menular yang disebabkan oleh virus dengue dan disebarkan oleh spesies nyamuk *Aedes aegypti*. Demam Berdarah Dengue dapat mengakibatkan suhu tubuh penderita menjadi sangat tinggi dan pada umumnya disertai demam tinggi, mual/muntah, sakit kepala, mimisan, nyeri pada otot serta persendian, bahkan dapat menyebabkan kematian [1]. Mobilitas masyarakat juga merupakan salah satu faktor risiko perpindahan virus Demam Berdarah Dengue pada individu dari satu kota ke kota lain yang dapat memengaruhi penyebaran penyakit Demam Berdarah Dengue. Terdapatnya kebiasaan masyarakat yang menampung air untuk keperluan sehari-hari seperti menampung air hujan atau menampung air sumur sehingga bak mandi atau drum jarang dikuras. Hal ini berpotensi sebagai tempat perkembangbiakan nyamuk *Aedes Aegypti* [2]. Selain itu, Terdapat pula kebiasaan masyarakat yang menyimpan barang-barang yang tidak terpakai/bekas tetapi kurang menyertipatir lingkungan terdapatnya air yang tertampung di dalam tempat penampungan air (TPA) serta kurang melaksanakan kebersihan di lingkungan sendiri. Anjuran 3M Plus (Menguras, Menutup, dan Mengubur Plus menaburkan larvasida, memelihara ikan pemakan jentik nyamuk, serta pemakaian insektisida rumah tangga) untuk mencegah dan memberantas penyakit Demam Berdarah Dengue. Pembinaan serta peran masyarakat juga sangat diperlukan agar lebih efektif [3].

Penyakit Demam Berdarah Dengue (DBD) merupakan salah satu masalah kesehatan masyarakat di Indonesia yang jumlah penderitanya semakin meningkat dan penyebarannya semakin luas, penyakit DBD merupakan penyakit menular yang pada umumnya menyerang pada usia anak-anak umur kurang dari 5 tahun dan juga bisa menyerang pada orang dewasa [4]. Pada tahun 2017 jumlah kasus DBD yang dilaporkan sebanyak 68.407 kasus dengan jumlah kasus meninggal sebanyak 493 orang dan incidence Rate (IR) 26,12 per 100.000 penduduk dibandingkan tahun 2016 dengan kasus sebanyak 204.171 serta IR 78,85 per 100.000 penduduk terjadi penurunan kasus pada tahun 2017 [5]. Selama kurun waktu 10 tahun terakhir mulai tahun 2008 cenderung tinggi sampai tahun 2010 kemudian mengalami penurunan drastis di tahun 2011 sebesar 27,67 per 100.000 penduduk yang dilanjutkan dengan tren kecenderungan meningkat sampai tahun 2016 sebesar 78,85 per 100.000 penduduk namun kembali mengalami penurunan drastis pada tahun 2017 dengan IR 26,12 per 100.000 penduduk. Tren jumlah Kabupaten/Kota terjangkit pada tahun 2010-2017 [6]. Penurunan angka kesakitan DBD pada tahun 2017 juga diringi oleh penurunan jumlah Kabupaten/Kota terjangkit DBD. Pada tahun 2016 terdapat 463 Kabupaten/Kota (90,08 %) menjadi 433 Kabupaten/Kota (84,24 %) pada tahun 2017 [7].

Selama periode tahun 2010 sampai tahun 2016 terlihat jumlah Kabupaten/Kota terjangkit DBD mengalami kenaikan, namun mulai menurun pada tahun 2017 [8]. Luasnya daerah dan jumlah penderita yang bertambah di Kabupaten Karawang memungkinkan perlunya sebuah pengelompokan daerah penyebaran Demam Berdarah Dengue. Pengelompokan daerah penyebaran Demam Berdarah Dengue diproses dengan menggunakan teknik data mining clustering menggunakan algoritma k-means [9]. Algoritma k-means banyak digunakan karena memiliki beberapa kelebihan diantaranya yaitu, mudah untuk diimplementasikan dan dijalankan, waktu yang dibutuhkan untuk menjalankan relatif cepat, umum digunakan dan mampu



DOI: 10.52362/jisicom.v8i2.1795

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).

mengelompokkan dokumen dalam jumlah yang besar dengan waktu komputasi yang cepat. Prinsip dari metode ini yaitu mempartisi suatu koleksi dokumen menjadi beberapa cluster dan menentukan centroid (titik pusat awal) secara acak [10].

Penggunaan algoritma k-means sudah diterapkan dalam berbagai bidang. Beberapa penelitian yang telah dilakukan menggunakan metode tersebut yaitu penelitian yang dilakukan oleh Asroni dan Ronald Adrian (2015), yaitu Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface dengan data yang digunakan yaitu data nilai matakuliah Algoritma dan Pemrograman, Fisika Dasar, Kalukulus dan Indek Prestasi Komuliatif (IPK) di Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Magelang[11]. Penelitian lainnnnya adalah yang dilakukan Adimas Ketut Nalendra (2018), yaitu Pengukuran Keakuratan Metode K-Means untuk Menentukan Status Gizi Pada Balita dengan data yang digunakan yaitu data primer dengan respondennya adalah balita dengan umur 0 – 36 bulan sesuai data posyandu di Desa Badas Kecamatan Badas Kabupaten Kediri[12]. Selain itu, terdapat penelitian lain yang menggunakan metode K-Means Clustering yaitu penelitian yang dilakukan oleh Asroni, Hidayatul Fitri, Eko Prasetyo (2018) yaitu Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calo Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Stufi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik) dengan menggunakan data calon mahasiswa baru di Universitas Muhammadiyah Yogyakarta (UMY) Tahun 2010 sampai dengan tahun 2014[13].

Pada penelitian ini bertujuan menganalisis algoritma k-means dengan memaksimalkan hasil clustering yang akan menghasilkan titik-titik pusat penyebaran penderita penyakit Demam Berdarah Dengue Mengelompokkan daerah penyebaran penyakit Demam Berdarah Dengue di Kabupaten Karawang dengan menggunakan algoritma k-means. Ide maupun gagasan sebelumnya sudah pernah terdapat, namun tidak menutup kemungkinan untuk menganalisa dengan metode yang disebutkan di atas guna mengelompokkan penyebaran penyakit Demam Berdarah Dengue (DBD) di daerah Kabupaten Karawang.

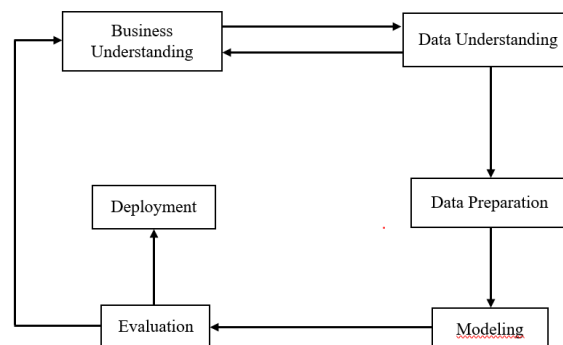
## II. METODE DAN MATERI

### 2.1 Objek Penelitian

Objek penelitian ini adalah algoritma K-Means Clustering dengan tempat penelitian di Kabupaten Karawang. Data yang telah diperoleh dari Dinas Kesehatan Kabupaten Karawang. Hasil akhir yang diperoleh adalah dapat mengetahui hasil clustering daerah penyebaran penyakit Demam Berdarah Dengue di Kabupaten Karawang dan diharapkan dapat menjadi sebuah informasi yang berguna.

### 2.2 Metodologi Penelitian

Dalam penulisan skripsi ini penulis menggunakan metode CRISP-DM (Cross Industry Standard Process for Data mining).

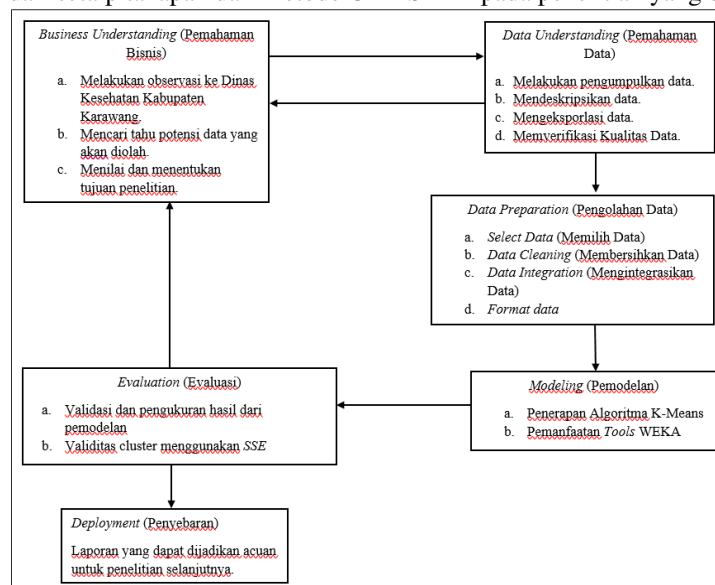


Gambar 2.1 Metodologi Penelitian

Metodologi penelitian yang digunakan adalah metode CRISP-DM (Cross Industry Standard Process for Data mining), yang terdiri dari 6 fase yaitu pemahaman bisnis (business understanding), pemahaman data (data understanding), pengolahan data (data preparation), pemodelan (modeling), penyebaran (deployment), evaluasi (evaluation).

### 2.3 Rancangan Penelitian

Berikut ini penjelasan dari setaip tahapan dari metode CRIPS-DM pada penelitian yang dilakukan adalah:



Gambar 2.2 Rancangan Penelitian

#### 2.3.1 Business Understanding (Pemahaman Bisnis)

Pada tahap ini melakukan observasi ke tempat yang akan dijadikan objek penelitian yaitu Dinas Kesehatan Kabupaten Karawang dan mencari potensi data yang akan diolah maka akan mendapatkan informasi yang akan dibutuhkan untuk menyelesaikan penelitian ini.

Setelah mengetahui potensi data, peneliti dapat menilai situasi yang akan dijadikan objek penelitian tersebut dan peneliti dapat menentukan tujuan dari penelitian. Pada proses observasi ini akan menghasilkan informasi yang dibutuhkan, adapun tujuan dari penelitian ini yaitu mengetahui pengelompokan daerah penyakit Demam Berdarah Dengue di Kabupaten Karawang.

#### 2.3.2 Data Understanding (Pemahaman Data)

Dalam tahap ini akan dilakukan proses pengumpulan data yang telah diperoleh dari Dinas Kesehatan Kabupaten Karawang, kemudian data tersebut dipelajari agar dapat mengenal serta dipahami lebih lanjut. Data yang dibutuhkan dalam penelitian ini yaitu data penderita penyakit Demam Berdarah Dengue di Kabupaten Karawang.

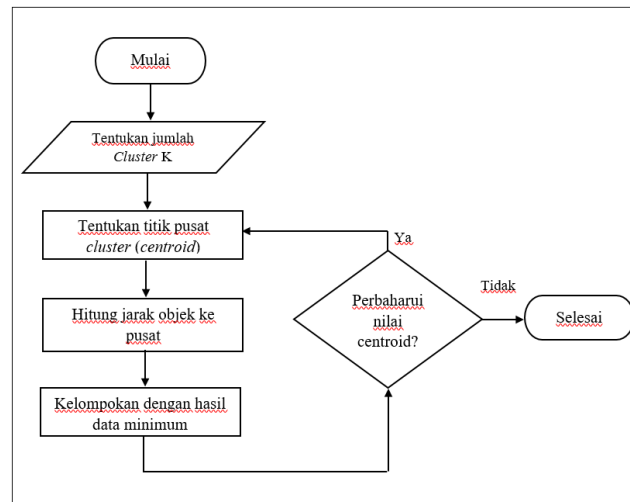
#### 2.3.3 Data Preparation (Pengolahan Data)

Pada tahapan ini dilakukan persiapan pada data yang telah diperoleh dari tahap sebelumnya. Pada tahapan ini perlu dilakukan prapemrosesan data terlebih dahulu agar data nantinya siap digunakan. Langkah-langkah yang akan dilakukan antara lain yaitu select data, cleaning data, integration data, dan format data. Pada tahapan ini juga meliputi semua aktivitas untuk membuat dataset final.

#### 2.3.4 Modeling (Pemodelan)



Tahapan ini merupakan gambaran dari rangkaian kegiatan untuk melakukan dari kegiatan untuk melakukan pelatihan serta pengujian terhadap data yang telah dikumpulkan dan diolah. Di tahap ini dilakukan pemilihan teknik clusteing dengan memilih algoritma K-Means yang akan digunakan, menggunakan aplikasi WEKA 3.9 Aplikasi WEKA 3.9 ini dapat membantu penelitian dalam menyelesaikan permasalahan data mining.



Gambar 2.3 Alur Algoritma K-Means

Tahapan menentukan clustering dengan metode K-Means yang terdiri dari beberapa langkah sebagai berikut:

1. Menentukan jumlah cluster k yang ingin di ingin diolah.
2. Menentukan titik pusat cluster (centroid) awal secara acak atau random.
3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing cluster. Untuk menghitung jarak antara objek dengan centroid dapat menggunakan Euclidian Distance.
4. Mengelompokan masing-masing objek kedalam centroid yang paling dekat.
5. Lakukan langkah ke-3, kemudian tentukan posisi centroid baru dengan melakukan langkah ke-2.
6. Ulangi langkah ke-3 jika posisi centroid baru tidak sama.

#### 2.3.5 Evaluation (Evaluasi)

Pada tahapan ini akan dilakukan validitas dan pengukuran keakuratan hasil yang dicapai oleh model yang sudah digunakan sebelumnya. Pada tahapan penelitian ini yang akan dievaluasi adalah apakah metode clustering dengan algoritma K-Means telah mencapai tujuan. Jika hasilnya belum sesuai dengan tujuan, maka proses akan diulang kembali ke proses data mining. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam tahapan pertama. Validitas cluster dilakukan menggunakan SSE (Sum Squared of Error). Jika semakin kecil nilai SSE, maka semakin bagus hasil clustering yang dihasilkan.

#### 2.3.6 Deployment (Penyebaran)

Tahap terakhir dari metodologi CRIPS-DM adalah tahap Deployment. Pada tahapan ini merupakan tahapan implementasi dari data mining, berupa laporan yang dapat digunakan untuk suatu sumber informasi atau referensi di penelitian selanjutnya.

### III. PEMBAHASA DAN HASIL

Hasil penelitian data mining yang telah dilakukan yaitu bagaimana menganalisis algoritma K-Means dapat pengelompokan daerah penyebaran penyakit Demam Berdarah Dengue di Kabupaten Karawang.





### 3.1 Business Understanding (Pemahaman Bisnis)

Dataset yang telah didapat berasal dari Dinas Kesehatan Kabupaten Karawang. Dari data yang diperoleh akan mendapatkan sebuah informasi yang dibutuhkan untuk menyelesaikan penelitian. Dari informasi yang didapat, diharapkan dapat memberikan informasi untuk pemahaman bisnis.

Tujuan dari penelitian ini yaitu menganalisis algoritma k-means untuk mengelompokkan daerah penyebaran penyakit Demam Berdarah Dengue dengan cepat serta akurat. Kemudian dari hasil pengelompokan tersebut menggunakan algoritma k-means dapat memberikan pengetahuan mengenai algoritma tersebut yang dapat digunakan untuk referensi penelitian selanjutnya. Pada penelitian ini diperlukan pemahaman mengenai latar belakang serta tujuan pada proses bisnis yang berhubungan dengan pengelompokan daerah penyebaran penyakit Demam Berdarah Dengue, tahapan-tahapan tersebut yaitu sebagai berikut:

#### 3.1.1 Determine Business Objective (Menentukan Tujuan Bisnis)

Tujuan dari penelitian ini antara lain yaitu agar data jumlah penderita penyakit Demam Berdarah Dengue dapat dimanfaatkan dengan sebaik mungkin serta dapat memberikan informasi kepada institusi kesehatan mengenai metode k-means untuk mengelompokkan daerah penyebaran penyakit Demam Berdarah Dengue di Kabupaten Karawang dan observasi yang telah dilakukan tersebut di Dinas Kesehatan Kabupaten Karawang.

#### 3.1.2 Assess The Situation (Menilai Situasi)

Adapun kondisi data yang terjadi saat ini pada Dinas Kesehatan Kabupaten Karawang yaitu sebagai berikut:

1. Pada Dinas Kesehatan Kabupaten Karawang terdapat berbagai kumpulan data yang dapat dijadikan informasi maupun pengetahuan. Data penderita Demam Berdarah di kabupaten Karawang tersedia setiap tahunnya. Data yang telah diperoleh yaitu data di tahun 2015, 2016 dan 2017.
2. Berbagai kumpulan data yang banyak serta menumpuk setiap tahunnya belum diolah agar menghasilkan sebuah informasi yang lebih bermanfaat.

#### 3.1.3 Determine Data Mining Goals (Menentukan Tujuan Data Mining)

Tujuan data mining atau tujuan dari penelitian ini yaitu menganalisis data penderita penyakit Demam Berdarah Dengue menggunakan algoritma k-means untuk mengelompokkan jumlah daerah penyebaran penyakit Demam Berdarah Dengue di Kabupaten Karawang agar dapat mengetahui hasil pengelompokan daerah penyebaran penyakit Demam Berdarah Dengue secara cepat serta akurat. Selain itu, dapat digunakan untuk mengasalkan sebuah informasi yang dibutuhkan dan mendapatkan pengetahuan baru dari data yang sudah diolah, sehingga dapat dijadikan acuan di penelitian selanjutnya.

### 3.2 Data Understanding (Pemahaman Data)

Pada data understanding adalah tahapan pengumpulan data awal dan hasil dari observasi yang dilakukan. Selanjutnya data yang diperoleh dipelajari untuk dapat mengenal dan memahami apa saja yang dapat dilakukan pada data-data tersebut. Tahapan ini terdiri dari beberapa tahapan yaitu collect initial the data (pengumpulan data awal), describe the data (mendeskripsikan data), explore the data (mengeksplorasi data), dan verify data quality (memverifikasi kualitas data).

#### 3.2.1 Collect Initial Data (Mengumpulkan Data Awal)

Collect initial data yaitu tahap mengumpulkan data awal. Data-data yang telah diperoleh dalam penelitian ini dikumpulkan dari hasil observasi yang dilakukan di Dinas Kesehatan Kabupaten Karawang. Data yang didapatkan merupakan data penderita Demam Berdarah Dengue di Kabupaten Karawang dari tahun 2015 sampai dengan tahun 2017.

#### 3.2.2 Explore Data (Mengeksplorasi Data)



Tahap ketiga dalam data understanding yaitu eksplorasi terhadap isi data. Dari data yang telah diperoleh berupa data dalam bentuk tabel pada kertas dalam setaip atribut. Tahap eksplorasi data berguna untuk mengetahui karakteristik data yang akan digunakan. Data apa saja yang yang dimasukan pada aplikasi WEKA 3.9 dapat dipelajari dari data yang tersedia. Eksplorasi data kriteria pada dataset dapat dilihat pada Tabel 4.2.

Tabel 4.2 Karakteristik Data

Tabel 3.1 Karakteristik Data

| No  | Atribut             | Tipe Data | Tahun | Keterangan                                                            |
|-----|---------------------|-----------|-------|-----------------------------------------------------------------------|
| 1.  | No                  | Nominal   | 2015  | Nomor urut penderita 1 - 569                                          |
|     |                     |           | 2016  | Nomor urut penderita 1 – 1059                                         |
|     |                     |           | 2017  | Nomor urut penderita 1 – 102                                          |
| 2.  | Penderita           | Nominal   | 2015  | Nama penderita terdapat 569 data                                      |
|     |                     |           | 2016  | Nama penderita terdapat 1059 data                                     |
|     |                     |           | 2017  | Nama penderita terdapat 102 data                                      |
| 3.  | Orang Tua           | Nominal   | 2015  | Nama orang tua terdapat 259 data                                      |
|     |                     |           | 2016  | Nama orang tua terdapat 565 data                                      |
|     |                     |           | 2017  | Nama orang tua terdapat 64 data                                       |
| 4.  | Sex (Jenis Kelamin) | Nominal   | 2015  | Sex (Jenis Kelamin) penderita laki-laki atau perempuan                |
|     |                     |           | 2016  |                                                                       |
|     |                     |           | 2017  |                                                                       |
| 5.  | Umur                | Numerik   | 2015  | Total umur penderita terdapat 569 data                                |
|     |                     |           | 2016  | Total umur penderita terdapat 1059 data                               |
|     |                     |           | 2017  | Total umur penderita terdapat 102 data                                |
| 6.  | Gol umur            | Nominal   | 2015  | Total golongan umur penderita terdapat 569 data                       |
|     |                     |           | 2016  | Total golongan umur penderita terdapat 1059 data                      |
|     |                     |           | 2017  | Total golongan umur penderita terdapat 102 data                       |
| 7.  | Tgl Mulai Sakit     |           | 2015  | Tanggal Dimulainya Penderita Sakit DBD terdapat 569 data              |
|     |                     |           | 2016  | Tanggal Dimulainya Penderita Sakit DBD terdapat 1059 data             |
|     |                     |           | 2017  | Tanggal Dimulainya Penderita Sakit DBD terdapat 102 data              |
| 8.  | Tgl Masuk RS        | Numerik   | 2015  | Tanggal Dimulainya Penderita DBD Masuk Rumah Sakit terdapat 569 data  |
|     |                     |           | 2016  | Tanggal Dimulainya Penderita DBD Masuk Rumah Sakit terdapat 1059 data |
|     |                     |           | 2017  | Tanggal Dimulainya Penderita DBD Masuk Rumah Sakit terdapat 102 data  |
| 9.  | Kampung             | Nominal   | 2015  | Nama Kampung Penderita DBD terdapat 569 data                          |
|     |                     |           | 2016  | Nama Kampung Penderita DBD terdapat 1059 data                         |
|     |                     |           | 2017  | Nama Kampung Penderita DBD terdapat 102 data                          |
| 10. | Rt                  | Numerik   | 2015  | Terdapat 50 puskesmas yang ada di Kabupaten Karawang                  |
|     |                     |           | 2016  |                                                                       |
|     |                     |           | 2017  |                                                                       |

### 3.2.3 Verify Quality Data (Memverifikasi Kualitas Data)



DOI: 10.52362/jisicom.v8i2.1795

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



Tahap evaluasi kualitas data dan kelengkapan data dilakukan dengan mengecek isi data. Hal ini dilakukan karena ada beberapa data berisi kolom yang memuat angka 0 pada tabel, artinya tidak ada hasil atau tidak ada penderita yang terkena penyakit Demam Berdarah Dengue.

### 3.3 Data Preparation (Persiapan Data)

Data preparation mencakup seluruh kegiatan guna membangun dataset yang akan dimasukkan kedalam alat pemodelan dari data mentah awal. Dataset yang diolah dalam pemodelan ini menggunakan algoritma k-means. Pada tahap data preparation terdiri dari beberapa tahapan yaitu, select data, cleaning data, integration data dan format data.

#### 3.3.1 Select Data (Memilih Data)

Data yang diambil mengenai penderita penyakit Demam Berdarah Dengue di Kabupaten Karawang dan disimpan dalam bentuk file excel atau \*.xlsx atau \*.xls. selanjutnya data dibersihkan dengan cara menghapus atribut-atribut yang tidak dibutuhkan pada penelitian yang dilakukan. Data yang dipilih adalah data mengenai penderita penyakit Demam Berdarah Dengue di tahun 2015, 2016 dan 2017.

#### 3.3.2 Cleaning Data (Pembersihan Data)

Pada tahap ini akan dilakukan pembersihan atribut yang tidak diperlukan saat proses perhitungan k-means clustering. Data setelah pembersihan atribut dapat dilihat pada Tabel 4.3, Tabel 4.4 dan Tabel 4.5.

Tabel 3.2 Data Cleaning

| Umur (Bulan->Tahun) | Kampung | Kelurahan | Puskesmas | Kecamatan |
|---------------------|---------|-----------|-----------|-----------|
| 16                  | 684     | 1         | 12        | 9         |
| 14                  | 685     | 2         | 40        | 23        |
| 9                   | 593     | 3         | 30        | 19        |
| 8                   | 686     | 4         | 36        | 21        |
| 10                  | 686     | 4         | 36        | 21        |
| 28                  | 687     | 5         | 32        | 19        |
| 35                  | 688     | 5         | 32        | 19        |
| 25                  | 148     | 6         | 33        | 19        |
| 12                  | 689     | 3         | 30        | 19        |
| 4                   | 192     | 7         | 48        | 28        |
| 24                  | 690     | 6         | 33        | 19        |
| 17                  | 693     | 10        | 1         | 1         |
| ...                 | ...     | ...       | ...       | ...       |
| 4                   | 949     | 12        | 26        | 17        |

Dari Tabel 4.3 atribut No, Penderita, Orang Tua, SEX, Golongan umur, Tanggal mulai sakit, Tanggal masuk RS, Rt, Rw, Diagnosa, S/M, RS, PE dan FOG telah dihapus. Atribut No, Penderita, Orang Tua, SEX, Golongan umur, Tanggal mulai sakit, Tanggal masuk RS, Rt, Rw, Diagnosa, S/M, RS, PE dan FOG adalah atribut identitas penderita DBD sehingga atribut tersebut tidak digunakan pada saat proses perhitungan. Setelah beberapa atribut tersebut dihilangkan, selanjutnya data akan disimpan dengan bentuk ekstensi \*.csv.

#### 3.3.3 Integration Data (Integrasi Data)



DOI: 10.52362/jisicom.v8i2.1795

Ciptaan disebarluaskan di bawah [Lisensi Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



Dataset penderita Demam Berdarah Dengue ini berasal dari Dinas Kesehatan Kabupaten Karawang. Selanjutnya data yang telah diperoleh harus melalui proses integrasi. Integration data merupakan penggabungan data dari berbagai tabel kedalam satu tabel baru. Adapun integrasi data yang dilakukan adalah dengan menggabungkan data tahun 2015 – 2017.

#### 3.3.4 Format Data

Pada tahap ini memproduksi dataset akhir yang siap untuk diolah untuk tools pemodelan data mining. Data yang telah diperoleh dari Dinas Kesehatan Kabupaten Karawang dengan hardcopy selanjutnya diinputkan kedalam Ms. Excel lalu disimpan kedalam bentuk .xlsx dan diubah menjadi format CSV (Comma Separated Values) atau \*.csv. Hal ini dilakukan guna mempermudah dalam proses modeling yang akan dilakukan di Weka 3.9.

#### 3.4 Modeling (Pemodelan)

Pada tahapan modeling data telah tersedia dengan baik dari data preparation dan siap digunakan. Teknik data mining yang digunakan dalam penelitian ini adalah teknik clustering dengan menggunakan algoritma k-means yang akan dilakukan terhadap 569 data di tahun 2015, 1059 data di tahun 2016 dan 102 data di tahun 2016. Hasil algoritma k-means dengan metode clustering akan diperoleh kelompok data yang akan digunakan untuk mengelompokkan daerah penyebaran penyakit Demam Berdarah Dengue dari yang terendah hingga tertinggi.

##### 3.4.1 Select Modeling Techniques (Memilih Teknik Pemodelan)

Dalam langkah awal pemodelan, pilih teknik pemodelan yang akan digunakan. Tujuan awal dalam penelitian ini adalah menggali pengetahuan mengenai pengelompokan daerah penyebaran penyakit Demam Berdarah Dengue sehingga dapat mengetahui tingkat akurasi serta evaluasi untuk model yang terbaik dari data yang telah diolah sebelumnya.

##### 3.4.2 Generate Test Design (Menghasilkan Desain Uji)

Tes pengujian atau tahap pembelajaran. Dalam tahap ini tes pengujian dilakukan dengan menggunakan Use Training Set. Hal ini dilakukan karena data yang diperoleh dalam kasus ini dapat digunakan sebagai training set serta testing set.

##### 3.4.3 Build Model (Membangun Model)

Pengujian jumlah atribut menggunakan software Weka 3.9 selanjutnya diperoleh nilai dari keseluruhan atribut data yang ada. Berikut ini merupakan nilai rank pada fitur seleksi untuk keseluruhan atribut yang diperoleh dari hasil perhitungan Weka 3.9. Berikut ini adalah hasil dari pengujian clustering k-means K=1 pada tahun 2015, 2016 dan 2017 dengan menggunakan kombinasi seed.





```
kMeans
=====

Number of iterations: 7
Within cluster sum of squared errors: 385.10311253290376

Initial starting points (random):

Cluster 0: 15,122,2,40,23
Cluster 1: 22,853,131,18,12

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
              (1703.0)      (936.0)      (767.0)
=====
UMUR           19.239         18.6999      19.897
KAMPUNG        405.8708       418.6741     390.2464
KELURAHAN      87.9225        82.7511      94.2334
PUSKESMAS      22.8103        34.7382       8.2542
KECAMATAN      14.2895        20.8654       6.2647
```

Gambar 3.1 Hasil Perhitungan K=2

Berdasarkan hasil diatas, data penderita penyakit Demam Berdarah Dengue dihasilkan sum of square errors sebesar 385.10 dan iterasinya sebesar 7. Adapun jumlah persentase dari 2 cluster yaitu cluster 0 sebanyak 55% (936 anggota), cluster 1 sebanyak 45% (767 anggota).

```
kMeans
=====

Number of iterations: 19
Within cluster sum of squared errors: 312.01100210168465

Initial starting points (random):

Cluster 0: 15,122,2,40,23
Cluster 1: 22,853,131,18,12
Cluster 2: 11,388,35,5,4

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
              (1703.0)      (531.0)      (444.0)      (728.0)
=====
UMUR           19.239         17.9731      20.0752      19.6525
KAMPUNG        405.8708       204.7872     702.3243     371.7363
KELURAHAN      87.9225        108.6573     55.9617      92.2912
PUSKESMAS      22.8103        37.1469      30.4572       7.6896
KECAMATAN      14.2895        21.9322      18.7995       5.9643
```

Gambar 3.2 Hasil Perhitungan K=3 dan S=20

Berdasarkan hasil diatas, data penderita penyakit Demam Berdarah Dengue dihasilkan sum of square errors sebesar 312,01 dan iterasinya sebesar 19. Adapun jumlah persentase dari 3 cluster yaitu cluster 0 sebanyak 31% (531 anggota), cluster 1 sebanyak 26% (444 anggota) , cluster 2 sebanyak 43% (728 anggota).





```
kMeans
=====
Number of iterations: 21
Within cluster sum of squared errors: 269.2332709097909

Initial starting points (random):

Cluster 0: 15,122,2,40,23
Cluster 1: 22,853,131,18,12
Cluster 2: 11,388,35,5,4
Cluster 3: 10,68,56,27,19

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute      Full Data      Cluster#
              (1703.0)      0          1          2          3
=====
UMUR           19.239        19.6735     19.9819     19.6954     16.7724
KAMPUNG        405.8708      201.1068     758.8927     370.339     251.9867
KELURAHAN      87.9225       34.5663     62.6859     92.1786     164.6711
PUSKESMAS      22.8103       33.0097     30.0995     7.4051      39.4784
KECAMATAN      14.2895       19.6602     18.5838     5.8762      23.1993
```

Gambar 3.3 Hasil Perhitungan K=4

Berdasarkan hasil diatas, data penderita penyakit Demam Berdarah Dengue dihasilkan sum of square errors sebesar 269,23 dan iterasinya sebesar 21. Adapun jumlah persentase dari 4 cluster yaitu cluster 0 sebanyak 18% (309 anggota), cluster 1 sebanyak 22% (382 anggota), cluster 2 sebanyak 42% (711 anggota), cluster 3 sebanyak 18% (301 anggota).

```
kMeans
=====
Number of iterations: 15
Within cluster sum of squared errors: 216.1761904061171

Initial starting points (random):

Cluster 0: 15,122,2,40,23
Cluster 1: 22,853,131,18,12
Cluster 2: 11,388,35,5,4
Cluster 3: 10,68,56,27,19
Cluster 4: 38,166,58,9,6

Missing values globally replaced with mean/mode

Final cluster centroids:
Attribute      Full Data      Cluster#
              (1703.0)      0          1          2          3          4
=====
UMUR           19.239        20.1699     19.1061     21.3748     16.4233     19.2315
KAMPUNG        405.8708      193.9412     747.4777     734.6803     268.3101     185.5507
KELURAHAN      87.9225       36.1013     58.3156     67.0781     164.5993     108.7453
PUSKESMAS      22.8103       32.6765     31.5447     6.8216      40.324      8.5839
KECAMATAN      14.2895       19.5523     19.3492     5.4424      23.5923     6.6046
```

Gambar 3.4 Hasil Perhitungan K=5

Berdasarkan hasil diatas, data penderita penyakit Demam Berdarah Dengue dihasilkan sum of square errors sebesar 216,17 dan iterasinya sebesar 15. Adapun jumlah persentase dari 5 cluster yaitu cluster 0 sebanyak 18% (306 anggota), cluster 1 sebanyak 21% (358 anggota) , cluster 2 sebanyak 16% (269 anggota) , cluster 3 sebanyak 17% (287 anggota) , cluster 4 sebanyak 28% (483 anggota).



```

kMeans
=====

Number of iterations: 27
Within cluster sum of squared errors: 187.30338259960448

Initial starting points (random):

Cluster 0: 15,122,2,40,23
Cluster 1: 22,853,131,18,12
Cluster 2: 11,388,35,5,4
Cluster 3: 10,68,56,27,19
Cluster 4: 38,166,58,9,6
Cluster 5: 30,854,131,18,12

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute    Full Data    Cluster#
              (1703.0)    0          1          2          3          4          5
              (1703.0)    (309.0)    (207.0)    (249.0)    (255.0)    (333.0)    (350.0)
=====
UMUR          19.239      20.2945    18.2554    21.4892    15.7824    19.7538    19.3169
KAMPUNG       405.8708    192.3981    260.715    752.7912    280.8118    171.7357    747.2543
KELURAHAN     87.9225     37.7217    175.7681    64.9799    162.7059    74.1532    55.2257
PUSKESMAS     22.8103     32.4045    15.6039    6.8514     42.4941    5.8649     31.7371
KECAMATAN     14.2895     19.3754    11.1208     5.4538     24.7373    4.7177     19.4543

```

Gambar 3.5 Hasil Perhitungan K=6

Berdasarkan hasil diatas, data penderita penyakit Demam Berdarah Dengue dihasilkan sum of square errors sebesar 187,303 dan iterasinya sebesar 27. Adapun jumlah persentase dari 6 cluster yaitu cluster 0 sebanyak 18% (309 anggota), cluster 1 sebanyak 12% (207 anggota) , cluster 2 sebanyak 15% (249 anggota) , cluster 3 sebanyak 15% (255 anggota) , cluster 4 sebanyak 20% (333 anggota) , cluster 5 sebanyak 21% (350 anggota).

```

kMeans
=====

Number of iterations: 36
Within cluster sum of squared errors: 174.3471086504024

Initial starting points (random):

Cluster 0: 15,122,2,40,23
Cluster 1: 22,853,131,18,12
Cluster 2: 11,388,35,5,4
Cluster 3: 10,68,56,27,19
Cluster 4: 38,166,58,9,6
Cluster 5: 30,854,131,18,12
Cluster 6: 8,3,172,36,21

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute    Full Data    Cluster#
              (1703.0)    0          1          2          3          4          5          6
              (1703.0)    (130.0)    (204.0)    (235.0)    (288.0)    (329.0)    (319.0)    (198.0)
=====
UMUR          19.239      14.2977    18.1611    20.6162    20.2191    19.9146    20.7332    17.004
KAMPUNG       405.8708    427.6308    253.6225    744.3277    191.6979    172.4043    766.3605    265.4091
KELURAHAN     87.9225     78.2077    175.4706    60.1872    37.9688    74.0456    58.4828    180.1667
PUSKESMAS     22.8103     46.5923    15.4902     6.1277    31.0417     5.6869    28.9216    41.1717
KECAMATAN     14.2895     27.1692    11.0637     5.0298    18.559     4.6809    18.0439    23.8535

```

Gambar 3.6 Hasil Perhitungan K=7

Berdasarkan hasil diatas, data penderita penyakit Demam Berdarah Dengue dihasilkan sum of square errors sebesar 174,34 dan iterasinya sebesar 36. Adapun jumlah persentase dari 7 cluster yaitu cluster 0 sebanyak 8% (130 anggota), cluster 1 sebanyak 12% (204 anggota) , cluster 2 sebanyak 14% (235 anggota) , cluster 3 sebanyak 17% (288 anggota) , cluster 4 sebanyak 19% (329 anggota) , cluster 5 sebanyak 19% (319 anggota) , cluster 6 sebanyak 12% (198 anggota).

```

KMeans
=====

Number of iterations: 16
Within cluster sum of squared errors: 159.878593029912

Initial starting points (random):

Cluster 0: 15,122,2,40,23
Cluster 1: 22,853,131,18,12
Cluster 2: 11,388,35,5,4
Cluster 3: 10,68,56,27,19
Cluster 4: 38,166,58,9,6
Cluster 5: 30,854,131,18,12
Cluster 6: 8,3,172,36,21
Cluster 7: 19,285,16,5,4

Missing values globally replaced with mean/mode

Final cluster centroids:

```

| Attribute | Full Data<br>(1703.0) | Cluster#<br>0<br>(122.0) | 1<br>(117.0) | 2<br>(222.0) | 3<br>(281.0) | 4<br>(229.0) | 5<br>(244.0) | 6<br>(192.0) | 7<br>(296.0) |
|-----------|-----------------------|--------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| UMUR      | 19.239                | 14.2926                  | 18.9818      | 20.4586      | 20.2957      | 17.9825      | 20.8078      | 17.0094      | 20.5868      |
| KAMPUNG   | 405.8708              | 407.1475                 | 786.4017     | 751.0045     | 183.9644     | 195.7948     | 732.8033     | 257.5208     | 195.9966     |
| KELURAHAN | 87.9225               | 74.9426                  | 140.9316     | 57.9595      | 39.2633      | 174.2795     | 30.5697      | 180.7031     | 61.2703      |
| PUSKESMAS | 22.8103               | 47.1066                  | 27.8034      | 5.8739       | 31.0498      | 12.8253      | 29.3443      | 41.2448      | 6.0845       |

Gambar 3.7 Hasil Perhitungan K=8

Berdasarkan hasil diatas, data penderita penyakit Demam Berdarah Dengue dihasilkan sum of square errors sebesar 159,87 dan iterasinya sebesar 16. Adapun jumlah persentase dari 8 cluster yaitu cluster 0 sebanyak 7% (122 anggota), cluster 1 sebanyak 7% (117 anggota) , cluster 2 sebanyak 13% (222 anggota) , cluster 3 sebanyak 17% (281 anggota) , cluster 4 sebanyak 13% (229 anggota) , cluster 5 sebanyak 14% (244 anggota) , cluster 6 sebanyak 11% (192 anggota) , cluster 7 sebanyak 17% (296 anggota).

```

KMeans
=====

Number of iterations: 23
Within cluster sum of squared errors: 153.37632416311163

Initial starting points (random):

Cluster 0: 15,122,2,40,23
Cluster 1: 22,853,131,18,12
Cluster 2: 11,388,35,5,4
Cluster 3: 10,68,56,27,19
Cluster 4: 38,166,58,9,6
Cluster 5: 30,854,131,18,12
Cluster 6: 8,3,172,36,21
Cluster 7: 19,285,16,5,4
Cluster 8: 36,301,18,10,7

Missing values globally replaced with mean/mode

Final cluster centroids:

```

| Attribute | Full Data<br>(1703.0) | Cluster#<br>0<br>(123.0) | 1<br>(119.0) | 2<br>(210.0) | 3<br>(282.0) | 4<br>(220.0) | 5<br>(240.0) | 6<br>(190.0) | 7<br>(227.0) | 8<br>(92.0) |
|-----------|-----------------------|--------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|
| UMUR      | 19.239                | 14.2089                  | 19.6039      | 20.0039      | 20.4011      | 16.7636      | 20.4004      | 16.6726      | 13.5581      | 42.3913     |
| KAMPUNG   | 405.8708              | 408.8049                 | 789.4958     | 763.8048     | 186.617      | 197.1182     | 733.2        | 257.1053     | 210.652      | 194.9783    |
| KELURAHAN | 87.9225               | 75.3008                  | 140          | 57.4905      | 39.4433      | 175.8682     | 30.0125      | 181.3316     | 60.2996      | 71.5109     |

Gambar 3.8 Hasil Perhitungan K=9

Berdasarkan hasil diatas, data penderita penyakit Demam Berdarah Dengue pada tahun 2015 dihasilkan sum of square errors sebesar 153,37 dan iterasinya sebesar 23. Adapun jumlah persentase dari 9 cluster yaitu cluster 0 sebanyak 7% (123 anggota), cluster 1 sebanyak 7% (119 anggota) , cluster 2 sebanyak 12% (210 anggota) , cluster 3 sebanyak 17% (282 anggota) , cluster 4 sebanyak 13% (220 anggota) , cluster 5 sebanyak 14% (240 anggota) , cluster 6 sebanyak 11% (190 anggota) , cluster 7 sebanyak 13% (227 anggota) , cluster 8 sebanyak 5% (92 anggota).

```

kMeans
=====

Number of iterations: 34
Within cluster sum of squared errors: 151.46758279958307

Initial starting points (random):

Cluster 0: 15,122,2,40,23
Cluster 1: 22,853,131,18,12
Cluster 2: 11,388,35,5,4
Cluster 3: 10,68,56,27,19
Cluster 4: 38,166,58,9,6
Cluster 5: 30,854,131,18,12
Cluster 6: 8,3,172,36,21
Cluster 7: 19,285,16,5,4
Cluster 8: 36,301,18,10,7
Cluster 9: 15,206,152,17,12

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute   Full Data   Cluster#   0       1       2       3       4       5       6       7       8       9
            (1703.0)   (123.0)   (202.0) (152.0) (262.0) (112.0) (134.0) (195.0) (288.0) (79.0)  (156.0)
-----
UMUR        19.239    14.0301   10.648  13.2751  19.3588  17.9607  37.8433  17.0964  19.1309  38.3038  18.2408
KAMPUNG     405.8708  399.4715  762.2673 783.3553 188.584  200.4464 724.1194 272.2513 204.9236 717.0633 201.0833

```

Gambar 3.9 Hasil Perhitungan K=10

Berdasarkan hasil diatas, data penderita penyakit Demam Berdarah Dengue pada tahun 2015 dihasilkan sum of square errors sebesar 151,48 dan iterasinya sebesar 34. Adapun jumlah persentase dari 10 cluster yaitu cluster 0 sebanyak 7% (123 anggota), cluster 1 sebanyak 12% (202 anggota) , cluster 2 sebanyak 9% (152 anggota) , cluster 3 sebanyak 15% (262 anggota) , cluster 4 sebanyak 7% (112 anggota) , cluster 5 sebanyak 8% (134 anggota) , cluster 6 sebanyak 11% (195 anggota) , cluster 7 sebanyak 17% (288 anggota) , cluster 8 sebanyak 5% (79 anggota) , cluster 9 sebanyak 9% (156 anggota).

#### 4.1.5 Evaluation (Evaluasi)

Hasil clustering dari setiap kombinasi k (ukuran cluster) dievaluasi menggunakan total SSE (sum of square error). Total SSE sudah mencukupi untuk evaluasi cluster menggunakan K-Means karena meminimalkan SSE (cohesion) ekuivalen dengan memaksimalkan SSB (separation). Nilai total SSE untuk clustering dari setiap kombinasi k disajikan berikut.

Tabel 3.3. Nilai Total SSE

| Nilai K | Besar SSE |
|---------|-----------|
| K=2     | 385,10    |
| K=3     | 312,01    |
| K=4     | 269,23    |
| K=5     | 216,17    |
| K=6     | 187,303   |
| K=7     | 174,34    |
| K=8     | 159,87    |
| K=9     | 153,37    |
| K=10    | 151,48    |

Pada tabel di atas terlihat bahwa untuk ukuran cluster (k) = 8, 9 dan 10, nilai total SSE-nya relatif sama. Untuk k yang lain diperoleh nilai total SSE yang berbeda. Hal ini berpengaruh pada jumlah iterasi dari algoritma K-Means. Jumlah iterasi untuk masing-masing ukuran cluster disajikan pada Tabel ... Jumlah iterasi pada algoritma K-Means tidak terlalu berpengaruh terhadap waktu eksekusi karena algoritma K-Means memiliki kompleksitas linear terhadap ukuran obyek data. Dari nilai K-2 sampai 10 yang memiliki nilai SSE terkecil adalah terdapat pada nilai K = 10 sebesar 151,48. Selain itu, nilai SSE dari setiap K, memiliki nilai SSE yang semakin menurun dari setiap K. Jumlah iterasi yang beraneka ragam. Jumlah iterasi terkecil terdapat pada nilai



K = 2 dengan jumlah iterasi sebesar 7. Sedangkan jumlah iterasi terbesar terdapat pada nilai K = 7 dengan jumlah iterasi sebesar 36.

### 3.5 Deployment (Penyebaran)

Tahap terakhir dari metodologi CRIPS-DM adalah tahap Deployment. Pada tahapan ini merupakan tahapan implementasi dari data mining, berupa laporan yang dapat digunakan untuk suatu sumber informasi atau referensi di penelitian selanjutnya.

## IV. KESIMPULAN

Hasil yang didapat untuk ukuran cluster ( $k$ ) = 8, 9 dan 10, nilai total SSE-nya relatif sama. Untuk  $k$  yang lain diperoleh nilai total SSE yang berbeda. Hal ini berpengaruh pada jumlah iterasi dari algoritma K-Means. Jumlah iterasi pada algoritma K-Means tidak terlalu berpengaruh terhadap waktu eksekusi karena algoritma K-Means memiliki kompleksitas linear terhadap ukuran obyek data. Dari nilai K-2 sampai 10 yang memiliki nilai SSE terkecil adalah terdapat pada nilai K = 10 sebesar 151,48. Selain itu, nilai SSE dari setiap K, memiliki nilai SSE yang semakin menurun dari setiap K. Jumlah iterasi terkecil terdapat pada nilai K = 2 dengan jumlah iterasi sebesar 7. Sedangkan jumlah iterasi terbesar terdapat pada nilai K = 7 dengan jumlah iterasi sebesar 36.

## REFERENASI

- [1] S. R. H. Mulya Rahma Karyanti, "Perubahan Epidemiologi Demam Berdarah Dengue Di Indonesia," *Sari Pediatri*, pp. 424-432, 2009.
- [2] K. K. R. Indonesia, "Profil Kesehatan Indonesia Tahun 2013," *Kementrian Kesehatan Republik Indonesia*, 2014.
- [3] W. N. R. R. E. S. F. W. R. , T. L. P. Anggraeni, "Modified Regression Approach for Predicting Number of Dengue Fever Incidents in Malang," *Procedia Computer Science*, pp. 142-150, 2017.
- [4] C. J. M. S. Fina Nasari, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Penyebaran Diare Di Kabupaten Langkat," *Cogito Smart Journal*, pp. 108-119, 2016.
- [5] H. Munawaroh, "Perbandingn Algoritma ID3 dan C5.0 dalam Identifikasi Penjurusan Siswa SMA," *Jurnal Sarjana Teknik Informatika*, 2013.
- [6] E. Prasetyo, "Data Mining - Konsep dan Aplikasi Menggunakan MATLAB," 2012.
- [7] A. C. Dian Wirdasari, "Penerapan Data Mining Untuk Mengolah Data Penempatan Buku Di Perpustakaan SMK TI PAB 7 Lubuk Pakam," *Jurnal SAINTIKOM*, 2011.
- [8] H. S. M. S. Mujib Ridwan, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Jurnal EECCIS*, pp. 59-64, 2013.
- [9] Nango, "Penerapan Algoritma K-Means Untuk Clustering Data Anggaran Pendapatan Belanja Daerah di Kabupaten XYZ," *Skripsi*, pp. 11-12, 2012.
- [10] A. S. S. M. S. Ade Tria Rahmayanti Saragih, "Penerapan Metode Clustering K-Means Untuk Proses Seleksi Calon Peserta Lomba MTQ," *Jurnal Pelita Informatika*, pp. 117-122 , 2018.
- [11] Windha Mega Pradnya Duhita, "CLUSTERING MENGGUNAKAN METODE K-MEANS UNTUK MENENTUKAN STATUS GIZI BALITA," *Jurnal Informatika*, p. 163, 2015.
- [12] M. A. M. K. Aldi Nurzahputra, "Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa," *Techno.COM*, pp. 17-24, 2017.
- [13] S. D. Y. Rio Andika Malik, "PERBANDINGAN ALGORITMA K-MEANS CLUSTERING DENGAN FUZZY C-MEANS DALAM MENGUKUR TINGKAT KEPUASAN TERHADAP TELEVISI DAKWAH SURAU TV," *Jurnal Teknologi dan Sistem Informasi Univrab*, pp. 10-21, 2018.





- [14] R. A. Asroni, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *JURNAL ILMIAH SEMESTA TEKNIKA*, p. 78, 2015.
- [15] I. Budiman, "DATA CLUSTERING MENGGUNAKAN METODOLOGI CRISP-DM UNTUK PENGENALAN POLA PROPORSI PELAKSANAAN TRIDHARMA," *Tesis*, p. 20, 2012.
- [16] ., I. S. A. E. P. Beta Estri Adiana, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering," *JUTEI Edisi Volume.2 No.1*, pp. 24-32, 2018.
- [17] &. d. Nugraha C, "Penerapan Metode Decision Tree(Data Mining) Untuk Memprediksi Tingkat Kelulusan Siswa Smpn1 Kintamani," *Seminar Nasional Vokasi dan Teknologi (SEMNASVOKTEK)*, 2016.
- [18] N. I. T. S. L. K. E. R. E. W. Denisha Intan P., "ANALISIS PERBANDINGAN METODE SINGLE LINKAGE DAN K-MEANS CLUSTERING (STUDI KASUS : KASUS DEMAM BERDARAH DENGUE (DBD) DI PROVINSI JAWA TENGAH TAHUN 2015)," *Seminar Nasional Teknologi Informasi dan Multimedia 2018*, pp. 2.4-1-2.4-6, 2018.
- [19] A. A. H. B. K. Rudi Rosadi, "Aplikasi K-Means Clustering Untuk Mengelompokkan Data Kinerja Akademik Mahasiswa," *SENTER 2016: Seminar Nasional Teknik Elektro 2016*, pp. 92-96, 2016.
- [20] A. K. Nalendra, "Pengukuran Keakuratan Metode K-Means untuk Menentukan Status Gizi Balita," *JURNAL EKONOMI DAN TEKNIK INFORMATIKA VOL. 6 NO. 2* , pp. 48-54, 2018.
- [21] H. F. E. P. Asroni, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *SEMESTA TEKNIKA*, pp. 60-64, 2018.
- [22] Y. S. Nugroho, "Modul Praktikum Data Mining Berdasarkan Kurikulum 2013," *Surakarta: Program Studi Teknik Informatika Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta*, 2014.
- [23] A. P. Windarto, ""Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, pp. 26-33, 2017.
- [24] S. Mulyati, "Penerapan Data Mining Dengan Metode Clustering Untuk Pengelompokan Data Pengiriman Burung," *Senatkom*, 2015.
- [25] A. M. Alfannisa Annurullah Fajrin, "PENERAPAN DATA MINING UNTUK ANALISIS POLA PEMBELIAN KONSUMEN DENGAN ALGORITMA FP-GROWTH PADA DATA TRANSAKSI PENJUALAN SPARE PART MOTOR," *Kumpulan jurnaL Ilmu Komputer (KLIK)*, p. 29, 2018.
- [26] A. W. Setyawati, "Implementasi Algoritma Partitioning Around Medoid (PAM) untuk Pengelompokan Sekolah Menengah Atas di DIY Berdasarkan Nilai Daya Serap Ujian Nasional," *Yogyakarta: Fakultas Sains dan Teknologi Universitas Sanata Dharma*, 2017.

