

**OPTIMASI HASIL EVALUASI *CLUSTERING* MELALUI
KOMBINASI ALGORITMA *DYNAMIC K-MEANS*
DAN *K-MEANS BINARY SEARCH CENTROID***

THESIS

**Oleh :
MUHAMMAD ANDRYAN WAHYU SAPUTRA
NIM. 210605210010**



**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**OPTIMASI HASIL EVALUASI *CLUSTERING* MELALUI KOMBINASI
ALGORITMA *DYNAMIC-K-MEANS* DAN *K-MEANS BINARY
SEARCH CENTROID***

THESIS

Diajukan Kepada :

**Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Magister Komputer (M.Kom)**

Oleh :

**MUHAMMAD ANDRYAN WAHYU SAPUTRA
NIM. 210605210010**

**PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

HALAMAN PERSETUJUAN

**OPTIMASI HASIL EVALUASI *CLUSTERING* MELALUI KOMBINASI
ALGORITMA *DYNAMIC-K-MEANS* DAN *K-MEANS BINARY*
*SEARCH CENTROID***

THESIS

Oleh :

**MUHAMMAD ANDRYAN WAHYU SAPUTRA
NIM. 210605210010**

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal : 15 Mei 2023

Dosen Pembimbing I,



Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

Dosen Pembimbing II,



Dr. Ririen Kusumawati, S.Si, M.Kom
NIP. 19720309 200501 2 002

Mengetahui,

Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Islam Negeri Maulana Malik Ibrahim Malang



Dr. Caryo Crsydian
NIP. 19740424 200901 1 008

HALAMAN PENGESAHAN

**OPTIMASI HASIL EVALUASI *CLUSTERING* MELALUI KOMBINASI
ALGORITMA *DYNAMIC-K-MEANS* DAN *K-MEANS BINARY*
*SEARCH CENTROID***

THESIS

Oleh :

**MUHAMMAD ANDRYAN WAHYU SAPUTRA
NIM. 210605210010**

Telah Dipertahankan di Depan Dewan Penguji Thesis dan
Dinyatakan Diterima Sebagai Salah Satu Persyaratan untuk
Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal : 24 Mei 2023

Susunan Dewan Penguji :

Penguji Utama : Dr. Muhammad Faisal, M.T
NIP. 19740510 200501 1 007

Ketua Penguji : Dr. Ririen Kusumawati, S.Si, M.Kom
NIP. 19720309 200501 2 002

Sekretaris Penguji : Dr. Fresy Nugroho, MT
NIP. 19710722 201101 1 001

Anggota Penguji : Dr. M. Imamudin, Lc., MA
NIP. 19740602 200901 1 010

Tanda Tangan



Mengetahui,
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Maulana Malik Ibrahim Malang



Caryo Crsydian
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini :

Nama : Muhammad Andryan Wahyu Saputra
NIM : 210605210010
Program Studi : Magister Informatika
Fakultas : Sains dan Teknologi
Judul Thesis : Optimasi Hasil Evaluasi *Clustering* Melalui Kombinasi
Algoritma *Dynamic-K-means* Dan *K-means Binary Search Centroid*

Menyatakan dengan sebenarnya bahwa thesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilan daya, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila di kemudian hari terbukti atau dapat dibuktikan thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan saya tersebut

Malang, 15 Mei 2023

Yang membuat pernyataan



Muhammad Andryan Wahyu Saputra

NIM. 210605210010

HALAMAN MOTTO

وَعَسَىٰ أَنْ تَكْرَهُهَا شَيْئًا وَهُوَ خَيْرٌ لَّكُمْ ۚ وَعَسَىٰ أَنْ تُحِبُّوا شَيْئًا وَهُوَ شَرٌّ
لَّكُمْ ۗ وَاللَّهُ يَعْلَمُ وَأَنْتُمْ لَا تَعْلَمُونَ ۚ ٢١٦

“Boleh jadi kamu tidak menyenangi sesuatu, padahal itu baik bagimu, dan boleh jadi kamu menyukai sesuatu, padahal itu tidak baik bagimu. Allah mengetahui, sedang kamu tidak mengetahui.” (QS. Al-Baqarah ayat 216).

Allah سبحانه و تعالى selalu berikan yang terbaik untukmu, mungkin bukan terbaik yang kamu inginkan, tapi pasti terbaik yang kamu butuhkan.

HALAMAN PERSEMBAHAN

الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

**Puji syukur kehadiran Allah *subhanahu wa ta'ala*, shalawat serta salam
kepada Rasulullah *shalallahu 'alaihi wa sallam*.**

Karya ini penulis persembahkan kepada:

Kepada kedua orang tua penulis, Bapak Sigit Handoko Saputro dan Ibu Erna Irawati serta Mas Adhitya Kurniawan Rizky Saputra selaku kakak penulis dan Azzahra Anisa Putri selaku adik penulis yang senantiasa menyayangi, mendidik, mendukung dan mendo'akan penulis. Ribuan kalimat tidak cukup untuk menggambarkan pengorbanan mereka. Semoga mereka senantiasa dalam rahmat Allah سبحانه و تعالی.

Seluruh guru-guru penulis mulai dari kecil hingga sekarang yang tidak dapat disebutkan satu per satu, yang telah membimbing dan memberikan ilmunya hingga penulis bisa seperti ini.

Seluruh dosen Teknik Informatika, terkhusus untuk dosen pembimbing penulis Bapak Dr. Muhammad Faisal, M.T dan Ibu Dr. Ririen Kusumawati, S.Si, M.Kom yang senantiasa membimbing dan memotivasi penulis untuk dapat menyelesaikan thesis ini.

Keluarga Magister Informatika dan Teknik Informatika, terutama Asosiasi S2 Informatika dan Angkatan 4 S2 Informatika serta UFO Unity Of Informatics Force (S1 Teknik Informatika angkatan 2018) yang memberikan semangat dan do'a kepada penulis.

Keluarga besar Pondok Pesantren Al Islam Kidul Pasar Malang yang selalu membantu dan meyemangati hingga penulis bisa menyelesaikan studi perkuliahan di UIN Maulana Malik Ibrahim Malang.

Seluruh orang-orang baik yang pernah hadir dalam kehidupan penulis, yang tidak dapat disebutkan satu per satu. Penulis ucapkan terima kasih, semoga Allah membalas kebaikan kalian dengan kebaikan yang lebih banyak. آمِينَ يَا رَبَّ الْعَالَمِينَ

KATA PENGANTAR



Segala puji bagi Allah تعالى و سبحانه yang Maha Pengasih lagi Maha Penyayang atas Rahmat dan Hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan thesis ini. Sholawat serta Salam tetap tercurahkan kepada junjungan kita, kekasih Allah, Nabi Muhammad ﷺ, sang pemberi syafaat kelak di hari akhir, beserta seluruh keluarga, sahabat, dan para pengikutnya.

Penelitian thesis yang berjudul “**Optimasi Hasil Evaluasi Clustering Melalui Kombinasi Algoritma Dynamic-K-means Dan K-means Binary Search Centroid**” ini ditulis untuk memenuhi salah satu syarat guna memperoleh gelar Magister Strata Kedua (S2) Fakultas Sains dan Teknologi Universitas Maulana Malik Ibrahim Malang. Selama proses penelitian thesis ini tidak lepas dari bimbingan, saran, dan arahan baik moral maupun spiritual dari berbagai pihak yang telah terlibat. Untuk itu dengan segala kerendahan hati, penulis mengucapkan rasa terimakasih yang sebesar-besarnya kepada:

1. Allah تعالى و سبحانه, yang telah memberikan rahmat, hidayah serta kesehatan sehingga penulis dapat menyelesaikan studi dan tugas akhir ini.
2. Ibu, Ayah dan seluruh keluarga yang selalu memberikan doa, kasih sayang dan motivasi untuk terus maju.
3. Bapak Dr. Muhammad Faisal, M.T selaku Dosen Pembimbing I dan dosen wali serta Ibu Dr. Ririen Kusumawati, S.Si, M.Kom, selaku dosen pembimbing 2 yang telah bersedia meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, berbagai pengalaman, arahan, nasihat, motivasi dan pengarahan dalam pembangunan program hingga penyusunan thesis ini.
4. Bapak Dr. Fresy Nugroho, MT dan Bapak Dr. M. Imamudin, Lc., MA selaku Dosen Penguji yang senantiasa memberikan bimbingan dan kritik yang membangun kepada penulis sehingga tergapai hasil thesis yang lebih baik.
5. Bapak Dr. Cahyo Crysdian selaku ketua program studi Magister Informatika yang mendukung dan mengarahkan thesis ini.
6. Dr. Sri Hariani, M.Si selalu dekan Fakultas Sains dan Teknologi Universitas

Islam Negeri (UIN) Maulana Malik Ibrahim.

7. Bapak Prof. Dr. M Zainuddin, M.A selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
8. Segenap sivitas akademika Fakultas Saintek, Universitas Islam Negeri Maulana Malik Ibrahim Malang terutama seluruh dosen, terimakasih atas segala ilmu dan bimbingannya.
9. Admin jurusan yang senantiasa memberikan arahan dalam penyusunan thesis
10. Teman-teman angkatan 4 S2 Informatika yang berjuang bersama-sama untuk meraih mimpi, terimakasih atas kenang-kenangan indah yang dirajut bersama.
11. Keluarga besar Pondok Pesantren Al Islam Kidul Pasar Malang yang selalu membantu dan meyemangati hingga penulis bisa menyelesaikan studi perkuliahan di UIN Maulana Malik Ibrahim Malang.
12. Semua pihak yang tidak dapat penulis sebutkan satu-persatu atas bantuan, masukan, dukungan serta motivasi kepada penulis.

Harapan penulis semoga semua amal kebaikan dan jasa-jasa dari semua pihak yang telah membantu hingga thesis ini selesai diterima oleh Allah سبحانه و تعالی, dan mendapatkan balasan yang lebih baik dan berlipat ganda serta selalu dilimpahkan rahmat dan hidayah-Nya kepada mereka semua.

Penulis juga menyadari bahwa thesis ini masih jauh dari kesempurnaan yang disebabkan keterbatasan Harapan penulis, semoga karya ini bermanfaat dan menambah ilmu pengetahuan bagi kita semua, Aamiin.

Malang, 15 Mei 2023

Penulis

DAFTAR ISI

| | |
|--|------|
| HALAMAN JUDUL | i |
| HALAMAN PERSETUJUAN..... | ii |
| HALAMAN PENGESAHAN | iii |
| PERNYATAAN KEASLIAN TULISAN | iv |
| HALAMAN MOTTO | v |
| HALAMAN PERSEMBAHAN | vi |
| KATA PENGANTAR | viii |
| DAFTAR ISI..... | x |
| DAFTAR GAMBAR | xii |
| DAFTAR TABEL..... | xiii |
| ABSTRAK..... | xiv |
| ABSTRACT..... | xv |
| المخلص..... | xvi |
| BAB I..... | 1 |
| PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Pernyataan Masalah..... | 6 |
| 1.3 Tujuan Penelitian..... | 6 |
| 1.4 Manfaat Penelitian..... | 6 |
| 1.5 Batasan Masalah..... | 6 |
| BAB II..... | 8 |
| STUDI PUSTAKA | 8 |
| 2.1. <i>Clustering</i> Partisi..... | 8 |
| 2.2. Algoritma <i>Dynamic K-means</i> (DK) | 11 |
| 2.3. Algoritma <i>K-means Binary Search Centroid</i> (KBSC)..... | 15 |
| 2.4. Evaluasi Hasil <i>Clustering</i> | 18 |
| 2.5. Penelitian Terkait | 20 |
| 2.6 Kebaruan Penelitian | 22 |
| 2.7. Kerangka Teori..... | 22 |

| | |
|--|----|
| BAB III | 24 |
| METODOLOGI PENELITIAN..... | 24 |
| 3.1 <i>Roadmap</i> Penelitian..... | 24 |
| 3.2 Tahapan Penelitian | 25 |
| 3.3 Sumber Data..... | 26 |
| 3.4 Variabel Penelitian | 27 |
| 3.5 Tahap Pra-proses Data | 28 |
| 3.6 Tahap <i>Clustering</i> | 29 |
| 3.7 Tahap Analisis..... | 35 |
| BAB IV | 38 |
| HASIL DAN PEMBAHASAN..... | 38 |
| 4.1 Implementasi Sistem | 38 |
| 4.2 Evaluation..... | 47 |
| 4.3 Hasil <i>Clustering</i> | 50 |
| 4.4 Hasil Perbandingan Algoritma | 54 |
| 4.5 Integrasi Islam | 58 |
| BAB V | 65 |
| KESIMPULAN DAN SARAN..... | 65 |
| 5.1 Kesimpulan..... | 65 |
| 5.2 Saran..... | 66 |
| DAFTAR PUSTAKA | 67 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1 Contoh Titik <i>Centroid</i> | 8 |
| Gambar 2.2 Hasil 3 <i>Cluster</i> Berdasarkan Kemiripan Data Dengan <i>Centroid</i> | 9 |
| Gambar 2.3 Jarak <i>Intra Cluster</i> | 10 |
| Gambar 2.4 Jarak <i>Inter Cluster</i> | 10 |
| Gambar 2.5 Flowchart Algoritma <i>Dynamic K-means</i> | 12 |
| Gambar 2.6 Flowchart <i>K-means Binary Search Centroid</i> | 16 |
| Gambar 2.7 Kerangka Teori..... | 23 |
| Gambar 3.1 <i>Roadmap</i> Penelitian | 24 |
| Gambar 3.2 Rancangan Alur Penelitian..... | 26 |
| Gambar 3.3 Alur Penggabungan Kedua Algoritma | 30 |
| Gambar 4.1 <i>Source Code Import Library</i> | 38 |
| Gambar 4.2 <i>Source Code Load Dataset</i> | 39 |
| Gambar 4.3 <i>Load Dataset</i> | 39 |
| Gambar 4.4 <i>Source code Rescaling</i> | 40 |
| Gambar 4.5 Hasil <i>Dataframe Rescaling</i> | 40 |
| Gambar 4.6 <i>Source code</i> Pembentukan Titik <i>Centroid</i> | 41 |
| Gambar 4.7 <i>Source code</i> Perhitungan <i>Euclidean distance</i> | 42 |
| Gambar 4.8 <i>Source code</i> Perhitungan <i>Manhattan distance</i> | 42 |
| Gambar 4.9 <i>Source code</i> Perhitungan <i>Minkowski distance</i> | 43 |
| Gambar 4.10 <i>Source code</i> Pembentukan Keanggotaan <i>Cluster</i> | 43 |
| Gambar 4.11 Hasil Pembentukan Keanggotaan <i>Cluster</i> | 44 |
| Gambar 4.12 <i>Source code</i> Menghitung Jarak <i>Intra Cluster (SSW)</i> | 44 |
| Gambar 4.13 <i>Source code</i> Menghitung Jarak <i>Inter Cluster (SSB)</i> | 45 |
| Gambar 4.14 <i>Source code Dynamic Cluster</i> | 46 |
| Gambar 4.15 Hasil Proses <i>Dynamic Cluster</i> | 47 |
| Gambar 4.16 <i>Source code Evaluation</i> Metode <i>Elbow</i> | 48 |
| Gambar 4.17 Hasil Evaluasi Metode <i>Elbow</i> | 48 |
| Gambar 4.18 <i>Source code Evaluation</i> Metode <i>DBI</i> | 49 |
| Gambar 4.19 Hasil Evaluasi Metode <i>DBI</i> | 49 |
| Gambar 4.20 <i>Source code Evaluation</i> Metode <i>CHI</i> | 50 |
| Gambar 4.21 Hasil Evaluasi Metode <i>CHI</i> | 50 |
| Gambar 4.22 Hasil Visualisasi <i>Cluster</i> | 51 |
| Gambar 4.23 Presentase Hasil <i>Clustering Dynamic K-means</i> dengan <i>Binary Search Centroid</i> | 51 |
| Gambar 4.24 Diagram Evaluasi Jarak Metode <i>DBI</i> | 55 |
| Gambar 4.25 Diagram Evaluasi Jarak Metode <i>CHI</i> | 55 |
| Gambar 4.26 Diagram Evaluasi <i>DBI</i> dan <i>CHI</i> | 56 |
| Gambar 4.27 Contoh Dataset Penyakit Stroke | 57 |
| Gambar 4.28 Hasil Visualisasi Persebaran <i>Cluster</i> Dataset Penyakit Stroke | 57 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 3.1 Variabel Penelitian..... | 27 |
| Tabel 3.2 Contoh Data | 32 |
| Tabel 3.3 Alokasi <i>Cluster</i> Iterasi 1 | 33 |
| Tabel 3.4 Alokasi <i>Cluster</i> Iterasi 2 | 34 |
| Tabel 3.5 Alokasi <i>Cluster</i> Iterasi 3 | 34 |
| Tabel 3.6 Iterasi <i>Clustering</i> | 34 |
| Tabel 3.7 Nilai <i>Sum of Square Within Cluster</i> | 36 |
| Tabel 3.8 Nilai <i>Sum of Square Between Cluster</i> | 36 |
| Tabel 4.1 Hasil <i>Clustering</i> | 52 |
| Tabel 4.2 Pengelompokan <i>Euclidean-Manhattan-Minkowski</i> | 54 |
| Tabel 4.3 Hasil Pengujian <i>KDBSC</i> Dengan <i>K-means</i> Tradisional..... | 56 |
| Tabel 4.4 Hasil Uji <i>DBI</i> Dan <i>CHI Cluster</i> Pada Dataset Penyakit Stroke..... | 58 |

ABSTRAK

Saputra, Muhammad Andryan Wahyu. 2023. **Optimasi Hasil Evaluasi *Clustering* Melalui Kombinasi Algoritma *Dynamic K-means* Dan *K-means Binary Search Centroid***. Thesis. Program Studi Magister Informatika, Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Muhammad Faisal, M.T, (II) Dr. Ririen Kusumawati, S.Si, M.Kom

Kata Kunci : *Clustering; Binary Search Centroid; Dynamic K-means; Profil Kesehatan Pulau Jawa*

Penelitian ini difokuskan untuk menentukan derajat kesehatan setiap kabupaten/kota dalam Pulau Jawa menggunakan algoritma *K-means Binary Search Centroid (KBSC)* dan *Dynamic K-means (DK)*. Data penelitian menggunakan data profil kesehatan Pulau Jawa tahun 2020. Perbandingan algoritma terbaik diuji tingkat validitasnya menggunakan metode *Elbow*, *Davies Bound Index (DBI)* dan *Calinski-Harabasz Index (CHI)* pada algoritma *K-means* tradisional dan *K-means Dynamic Binary Search Centroid*. Berdasarkan pengujian diperoleh 5 jumlah *cluster* di wilayah sebaran, di antaranya *cluster 1* dengan kualitas kesehatan sangat tinggi adalah 11 kabupaten/kota, *cluster 2* dengan kualitas kesehatan sangat rendah adalah 24 kabupaten/kota, dan *cluster 3* dengan kualitas kesehatan rendah adalah 28 kabupaten/kota. dan kualitas kesehatan 45 kabupaten/kota untuk *cluster 4* dengan kualitas kesehatan cukup serta *cluster 5* dengan kualitas kesehatan tinggi adalah 11 kabupaten/kota. Secara keseluruhan optimasi algoritma *K-means Dynamic Binary Search Centroid* menghasilkan rata-rata kualitas *cluster* yang lebih baik dibanding algoritma *K-means* tradisional. Hal ini ditunjukkan dengan evaluasi metode *Elbow* menghasilkan jumlah *cluster* terbaik sebanyak 5, perbedaan nilai *DBI* sebesar 0.6825546 dimana metode yang diusulkan memiliki nilai lebih baik dibanding metode *K-means* tradisional. Serta jumlah dan variasi data mempengaruhi kinerja algoritma dan besar nilai evaluasi algoritma. Hasil pengujian dapat digunakan sebagai salah satu metode terbaik dalam evaluasi tingkat kesehatan yang ada di wilayah khususnya Pulau Jawa serta acuan pengambilan keputusan dalam menentukan kebijakan bagi instansi terkait.

ABSTRACT

Saputra, Muhammad Andryan Wahyu. 2023. **Optimizing Clustering Evaluation Results through the Combination of Dynamic K-means and K-means Binary Search Centroid Algorithms.** Thesis. Master of Informatics, Faculty of Science and Technology. Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Dr. Muhammad Faisal, M.T, (II) Dr. Ririen Kusumawati, S.Si, M.Kom

Kata Kunci : *Clustering; Binary Search Centroid; Dynamic K-means; Java Island Health Profile*

This research is focused on determining the health status of each district/city in Java using the K-means Binary Search Centroid (KBSC) and Dynamic *K-means* (DK) algorithms. The research data used health profile data for the island of Java in 2020. Comparison of the best algorithms was tested for validity using the Elbow, Davies Bound Index (DBI) and Calinski-Harabasz Index (CHI) methods on the traditional K-means algorithm and the K-means Dynamic Binary Search Centroid. Based on the test, it obtained 5 clusters in the distribution area, including cluster 1 with very high health quality were 11 districts/cities, cluster 2 with very low health quality were 24 districts/cities, and cluster 3 with low health quality were 28 districts/cities . and the health quality of 45 districts/cities for cluster 4 with adequate health quality and for cluster 5 with high health quality are 11 districts/cities. Overall the optimization of the K-means Dynamic Binary Search Centroid algorithm produces a better average cluster quality than the traditional K-means algorithm. This is shown by the evaluation of the Elbow method which produces the best number of clusters of 5, the difference in the DBI value is 0.6825546 where the proposed method has a better value than the traditional K-means method. As well as the amount and variety of data affect the performance of the algorithm and the value of the algorithm evaluation. The test results can be used as one of the best methods in evaluating the level of health in the region, especially Java Island, as well as a reference for decision making in determining policies for related agencies.

الملخص

سابوترا، محمد أندريان وحي. ٢٠٢٣. ترقية نتائج تقييم التجميع (*Clustering*) باستخدام الخوارزمية التصنيفية وسائل ك الديناميكية ووسائل ك للبحث الثنائي التمركزية. رسالة الماجستير. قسم المعلومات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. محمد فيصل، الماجستير. المشرف الثاني: د. ريرين كوسوماواتي، الماجستير.

الكلمات الرئيسية: تجميع، البحث الثنائي التمركزية، وسائل ك ديناميكية، الملف الصحي لجزيرة جاوى.

ركزت هذا البحث على تحديد الدرجة الصحية لكل منطقة / مدينة في جزيرة جاوى باستخدام خوارزمية تصنيفية وسائل ك للبحث الثنائي التمركزية و وسائل ك الديناميكية. استخدمت بيانات البحث بيانات الملف الصحي لجزيرة جاوى في عام ٢٠٢٠. تم اختبار أفضل مقارنات الخوارزميات للتأكد من صحتها باستخدام طرق *Elbow* و *Davies Bound Index (DBI)* و *Calinski-Harabasz Index (CHI)* على خوارزمية تصنيفية وسائل ك التقليدية ووسائل ك للبحث الثنائي التمركزية. واستنادا إلى الاختبارات، تم الحصول على ٥ مجموعات في منطقة التوزيع، بما في ذلك المجموعة الأولى ذات الجودة الصحية العالية جدا وهي ١١ منطقة/مدينة، والمجموعة الثانية ذات الجودة الصحية المنخفضة جدا هي ٢٤ منطقة/مدينة، والمجموعة الثالثة ذات الجودة الصحية المنخفضة هي ٢٨ منطقة/مدينة. والجودة الصحية ل ٤٥ منطقة / مدينة للمجموعة الرابعة ذات جودة صحية كافية والمجموعة الخامسة ذات الجودة الصحية العالية هي ١١ منطقة / مدينة. بشكل عام، يؤدي ترقية خوارزمية وسائل ك للبحث الثنائي التمركزية إلى متوسط جودة الكتلة أفضل من خوارزمية وسائل ك التقليدية. يظهر ذلك من خلال تقييم طريقة *Elbow* مما أدى إلى أفضل عدد من المجموعات بقدر ٥، وفرق في قيمة *DBI* قدره ٠.٦٨٢٥٥٤٦، حيث يكون للطريقة المقترحة قيمة أفضل من طريقة وسائل ك التقليدية. وكذلك يؤثر مقدار البيانات وتنوعها على أداء الخوارزمية وحجم قيمة تقييم الخوارزمية. يمكن استخدام نتائج الاختبار كواحدة من أفضل الطرق في تقييم مستوى الصحة في المنطقة، وخاصة جاوى. وكذلك كمرجع لاتخاذ القرار في تحديد السياسات للوكالات ذات الصلة.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indikator keberhasilan pembangunan suatu negara dapat dilihat dari tingkat pencapaian negara tersebut dalam memberikan jaminan kesehatan. Di Indonesia, pemerintah telah menetapkan sejumlah indikator melalui Kementerian Kesehatan yang menjadi tolak ukur kemajuan pembangunan kesehatan di semua tingkat unit wilayah Indonesia (dari provinsi, kabupaten, hingga kecamatan).

Setiap tahun, Kementerian Kesehatan mengumpulkan dan mengolah data kesehatan kependudukan melalui dinas kesehatan yang tersebar di berbagai daerah untuk menghasilkan rangking kesehatan provinsi dan kabupaten/kota. Hasil olahan dan pemeringkatan ini menjadi acuan penting, misalnya pemerintah daerah (Pemda) merumuskan rencana *intervensi* yang lebih tepat sasaran sebagai bahan promosi agar pemda lebih terpacu untuk meningkatkan rangking kesehatan daerahnya, sebagai dasar pemerintah pusat merumuskan keseriusan/Khusus daerah bermasalah kesehatan (DBKKBK), sebagai dasar penetapan alokasi pendanaan kesehatan pemerintah pusat ke daerah juga berfungsi untuk mendukung Departemen Negara Daerah Rentan dalam pembangunan daerah/perkotaan (Wahyudin dkk, 2016).

Pulau Jawa masih menjadi sentral perekonomian nasional. Hampir 60 persen kegiatan ekonomi berada di pulau seluas 128 ribu km² tersebut. Badan Pusat Statistik (BPS) mencatat produk domestik bruto (PDB) nasional sebesar Rp 14,8 ribu triliun pada 2018. Kesehatan dan pertumbuhan ekonomi adalah dua hal yang saling bertaut. Lemahnya perekonomian nasional akibat turunnya tingkat

kesehatan di Jawa saat ini menjadi sebab yang bisa diterima mengingat besarnya ketergantungan roda ekonomi nasional terhadap Pulau Jawa. *Clustering* wilayah dilakukan untuk mengetahui tingkat kesehatan yang ada di wilayah khususnya Pulau Jawa. Di kutip dari Badan Pusat Statistik Pulau Jawa terdiri dari 85 Kabupaten dan 34 Kota. Selama ini dalam pengelompokan data indikator penyehatan lingkungan masih berbasis teknik komputasi manual, dimana perhitungannya masih memiliki sejumlah permasalahan khususnya dalam bidang konsistensi data. Dalam hal peningkatan pelayanan kesehatan Dinas Kesehatan, maka dibutuhkan analisa yang dapat melakukan pengelompokan wilayah sehat berdasarkan pada data kesehatan lingkungan, sehingga penyuluhan, pelayanan, serta pemberian bantuan dapat lebih akurat dan tepat sasaran.

Menghitung nikmat yang telah Allah berikan adalah sebuah pekerjaan yang sulit. Bagaimana tidak, Allah mengatakan bahwa jika seorang hamba ingin menghitung nikmat tersebut, maka tidak akan sanggup menghitungnya. Di antara sekian banyak nikmat yang telah Allah berikan, ada 2 nikmat yang manusia lalai darinya. Nikmat tersebut adalah kesehatan dan waktu luang. *Rasulullah shallallahu ‘alaihi wa sallam* bersabda:

نعمتان مغبون فيهما كثير من الناس الصحة والفراغ

“*Dua kenikmatan yang sering dilupakan oleh kebanyakan manusia adalah kesehatan dan waktu luang.*” (HR. Al-Bukhari: 6412, at-Tirmidzi: 2304, Ibnu Majah: 4170)

Setiap nikmat yang Allah berikan kepada kita wajib untuk kita syukuri. Dalam hadis yang mulia di atas Nabi Muhammad *shallallahu ‘alaihi wa sallam* menjelaskan pentingnya nikmat sehat. Salah satu cara mensyukuri nikmat sehat adalah dengan menjaga nikmat sehat itu sendiri. Islam memperhatikan aspek

kesehatan karena kesehatan merupakan hal yang mutlak dalam menjalani aktivitas kehidupan manusia. Bila tubuh manusia dalam keadaan sehat mereka bisa melakukan aktivitas ibadah (hubungan manusia dengan Tuhannya), aktivitas sosial (hubungan manusia dengan manusia), serta aktivitas dunia (hubungan manusia dengan alam) dengan baik terhadap semua aspek tersebut. Hal ini disebabkan karena dalam perintah Allah pada manusia banyak yang berupa aktivitas fisik yang memerlukan kondisi yang baik dan kuat, seperti shalat, wudhu, puasa, dan ibadah lainnya.

Clustering termasuk kedalam *unsupervised learning* (pembelajaran tidak terbimbing), karena ketika pengelompokan data didasarkan atas kemiripan dan ketidakmiripan antar datanya. Analisis *cluster* telah banyak digunakan termasuk riset pasar, pengenalan pola, analisis data, dan pengolahan citra (Arpit, 2017).

Pentingnya *clustering* kesehatan dalam pandangan islam bertujuan dalam proses pengolahan data menjadi suatu informasi yang benar dan tepat untuk dapat digunakan dalam pembuatan perencanaan dan pengambilan keputusan. Al-Qur'an menggambarkan bahwa peranan data dan informasi yang hak atau benar, memiliki karakteristik yang dapat menyenangkan hati penerima informasi, Informasi yang benar dan perlu diketahui oleh orang lain dan tidak mencampuradukkan berita yang benar dengan berita yang salah/batil, Informasi yang adil dengan tidak memihak salah satu pihak, Informasi yang dapat menyelesaikan perbedaan atau pertentangan diantara penerima dan dapat mendamaikan dari perselisihan mereka dan Informasi yang tidak hanya mengikuti kepuasan subjektif.

Pada teknik partisi *clustering* terdapat algoritma *K-means*, algoritma *K-means* merupakan suatu algoritma *clustering* yang mempartisi dataset kedalam beberapa *cluster*, algoritma *K-means* cukup mudah untuk diimplementasi dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan. Algoritma *K-means* digunakan untuk kasus *clustering* wilayah berdasarkan profil kesehatan karena penentuan jumlah *cluster* dapat disesuaikan dengan kebutuhan sehingga jumlah *cluster* dapat diasumsikan diawal proses *cluster*. namun algoritma *K-means* mempunyai kelemahan yaitu harus dapat menduga jumlah *cluster* diawal, karena asumsi jumlah *cluster* yang ditentukan diawal belum tentu menghasilkan *cluster* yang baik. Kemudian algoritma *K-means* lainnya yaitu dalam penentuan titik *centroid* awal yang masih dipilih secara acak yang akan berpengaruh pada kondisi local minimal (kondisi hasil *cluster* belum tentu yang terbaik). Penentuan titik *centroid* akan berpengaruh pada iterasi proses untuk menghasilkan *cluster* convergence yang artinya hasil *cluster* memiliki anggota *cluster* yang tidak berubah (Gumilar, 2018).

Pada saat kondisi dimana data pengelompokan wilayah tidak ingin ditentukan sesuai kebutuhan, namun ingin menghasilkan segmentasi dari sebaran data, maka untuk penentuan asumsi jumlah *cluster* tidak diduga diawal, namun harus dicari asumsi jumlah *cluster* yang akan terbentuk. Untuk menangani kondisi tersebut, terdapat pengembangan dari algoritma *K-means*, yaitu algoritma *Dynamic K-means* (DK) yang memiliki proses untuk mencari jumlah *cluster* tanpa harus menduga asumsi jumlah *cluster* (Hossain, 2019). Pada penelitian Widiarina dan Romi mengusulkan algoritma *cluster* dinamik pada algoritma *K-means* dalam menetapkan jumlah *cluster* (k) agar dapat menghasilkan kualitas

cluster yang optimal sehingga memberikan hasil segmentasi potensial lebih baik dan tepat (Ariasa, 2020). Akan tetapi dalam algoritma tersebut memiliki beberapa kekurangan yang sama dengan algoritma *K-means* yaitu masalah pada penentuan titik *centroid* pada proses *cluster* yang masih dipilih secara acak. Berdasarkan penelitian dari Hadi Santoso untuk masalah penentuan titik *centroid* dilakukan dengan cara mengembangkan *K-means* menjadi *K-means Binary Search Centroid* (KBSC) yang memiliki proses dalam menentukan titik *centroid* menggunakan pendekatan teknik Binary Search (Santoso, 2022). Hasil dari penelitian tersebut membuktikan bahwa Algoritma *K-means Binary Search Centroid* (KBSC) memiliki nilai *intra* dan *inter cluster* yang lebih baik dari algoritma *K-means*. Akan tetapi Algoritma *K-means Binary Search Centroid* (KBSC) memiliki keterbatasan dalam menentukan dugaan jumlah *cluster* yang akan dibentuk.

Berdasarkan penjelasan masing-masing Algoritma *Dynamic K-means* dan *K-means Binary Search Centroid* dapat dilihat beberapa kelebihan dan kekurangannya. Algoritma *K-means Binary Search Centroid* memiliki kelebihan dalam penentuan pusat cluster awal dan memiliki kekurangan dalam penentuan jumlah cluster sebaliknya algoritma *Dynamic K-means* memiliki kekurangan dalam penentuan pusat cluster awal dan memiliki kelebihan dalam penentuan jumlah cluster.

Oleh karena itu maka diusulkan penggabungan algoritma *Dynamic K-means* dengan algoritma *K-means Binary Search Centroid* dapat saling melengkapi proses *cluster* pada area asumsi jumlah *cluster* dan penentuan titik *centroid cluster*, sehingga dapat menghasilkan nilai Davies-Bouldin Index yang terbaik untuk studi kasus *clustering* wilayah berdasarkan profil kesehatan.

1.2 Pernyataan Masalah

Berdasarkan dari latar belakang masalah yang telah dipaparkan, maka peneliti memberikan pernyataan masalah sebagai berikut :

Dalam penelitian ini, pengujian dilakukan terhadap hasil pengelompokan menggunakan gabungan algoritma *Dynamic K-means* dan algoritma *K-means Binary Search Centroid*. Dalam hasil pengujian ini, akan diperoleh informasi mengenai bagaimana *clustering* yang dihasilkan oleh kombinasi kedua algoritma tersebut.

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

Mengkombinasikan algoritma *Dynamic K-means* dengan algoritma *K-means Binary Search Centroid* sehingga dapat menghasilkan nilai evaluasi *cluster* yang lebih baik dari algoritma *K-means* konvensional yang diuji pada penelitian ini.

1.4 Manfaat Penelitian

Adapun manfaat penelitian yang bisa diambil dari penelitian ini adalah:

1. Memudahkan instansi dalam melakukan proses *clustering* wilayah berdasarkan data profil kesehatan secara efektif dan obyektif.
2. Penelitian ini dapat menghasilkan algoritma yang dapat menentukan jumlah *cluster* dan pusat *cluster* awal secara otomatis sehingga hasil *cluster* memiliki kualitas hasil *cluster* yang baik.

1.5 Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah:

1. Algoritma yang akan digunakan yaitu algoritma *Dynamic K-means* dan algoritma *K-means Binary Search Centroid*
2. Algoritma diuji pada data testing studi kasus mengenai pengelompokan wilayah berdasarkan data profil kesehatan Pulau Jawa pada tahun 2020.
3. Penelitian ini juga akan diuji menggunakan data dalam jumlah besar yaitu dataset Penyakit Stroke.

BAB II

STUDI PUSTAKA

2.1. *Clustering* Partisi

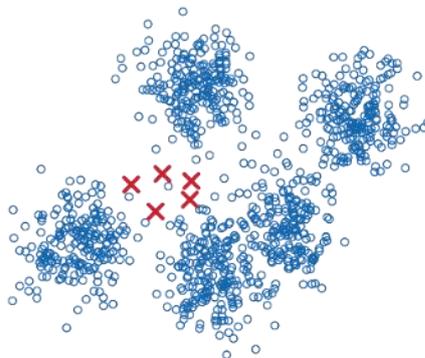
Berikut ini yang termasuk teknik *clustering* partisi salah satunya adalah algoritma *K-means*, dan *K-medoid*. *Dynamic K-means*, dan *K-means Binary Search Centroid* (KBSC) merupakan pengembangan dari algoritma *K-means* (Gumilar & Yusrila, 2017). Dalam *clustering* partisi ada beberapa hal yang perlu diperhatikan untuk proses *cluster* nya yaitu :

1. **Penentuan Jumlah Cluster**

Jumlah *cluster* atau disebut juga jumlah k , ditentukan diawal pada proses *cluster*. Penentuan jumlah *cluster* ini sangat memiliki pengaruh terhadap hasil akhir pembentukan pada proses *cluster*.

2. **Titik Centroid Data**

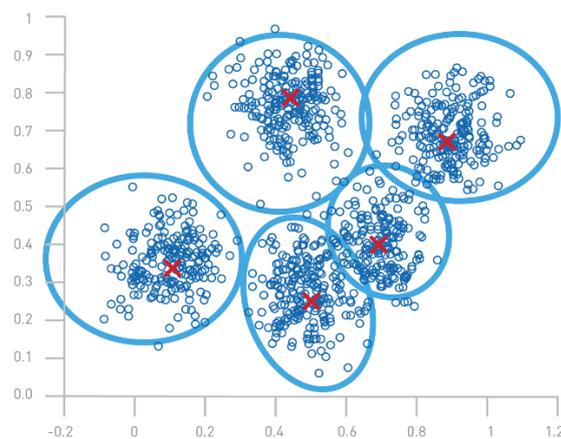
Merupakan titik pusat data yang mewakili dari pembentukan *cluster* yang dihasilkan. Titik *centroid* data memiliki jumlah sesuai dengan jumlah *cluster* yang ditentukan. Di Gambar 2.1 titik *centroid* dijelaskan pada tanda silang berwarna merah sebagai berikut ini :



Gambar 2.1 Contoh Titik *Centroid*

(Sumber : <https://www.nvidia.com/en-us/glossary/data-science/K-means/>)

Pada proses *cluster* maka setiap data akan dikelompokkan berdasarkan kemiripan yang diukur menggunakan perhitungan jarak masing-masing data terhadap titik *centroid*, kemudian diambil jarak yang terkecil. Proses pengelompokan dilakukan secara berulang sehingga menghasilkan kelompok yang tetap. Berikut ini hasil *cluster* yang terbentuk berdasarkan kedekatan data terhadap *centroid* :



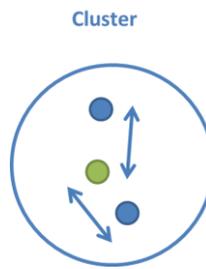
Gambar 2.2 Hasil 3 *Cluster* Berdasarkan Kemiripan Data Dengan *Centroid*
(Sumber : <https://www.nvidia.com/en-us/glossary/data-science/K-means/>)

Berdasarkan Gambar 2.2 menjelaskan setiap data pada masing-masing *cluster* memiliki *centroid* yang merupakan rata-rata dari data yang masuk pada *cluster* (Gumilar & Yusrila, 2017).

3. Jarak *Intra Cluster*

Istilah *intra* digunakan untuk mengukur kekompakan dari suatu kelompok (Widiarina & Romi, 2015). Dengan kata lain jarak *intra* adalah jarak antar titik data dengan titik pusat (*centroid*) dalam suatu kelompok atau *cluster*, sehingga semakin kecil jarak *intra* maka akan semakin baik hasil *cluster* yang dihasilkan.

Berikut ini gambaran mengenai *intra* yang diperlihatkan pada Gambar 2.6.

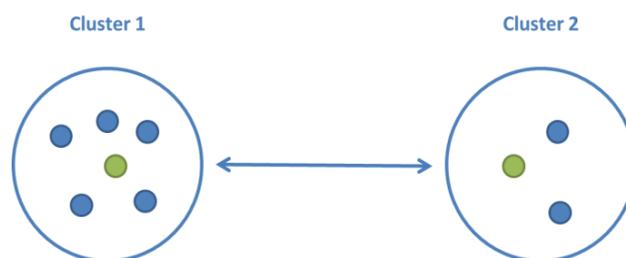


Gambar 2.3 Jarak *Intra Cluster*
(Sumber : Neural Computing and Applications)

Berdasarkan Gambar 2.3, jarak titik data berwarna biru ke titik pusat berwarna hijau diistilahkan sebagai *intra*.

4. Jarak *Inter Cluster*

Istilah *inter* adalah minimum jarak antar pusat *cluster*, *inter* digunakan untuk mengukur pemisahan antar *cluster* (Widiarina & Romi, 2015). Dari keterangan tersebut maka disimpulkan bahwa jarak *inter* adalah jarak antar pusat *cluster* dengan pusat *cluster* lainnya, sehingga semakin besar jarak *inter* maka akan semakin baik hasil *cluster* yang dihasilkan. Berikut ini gambaran mengenai *intra* yang diperlihatkan pada Gambar 2.4.



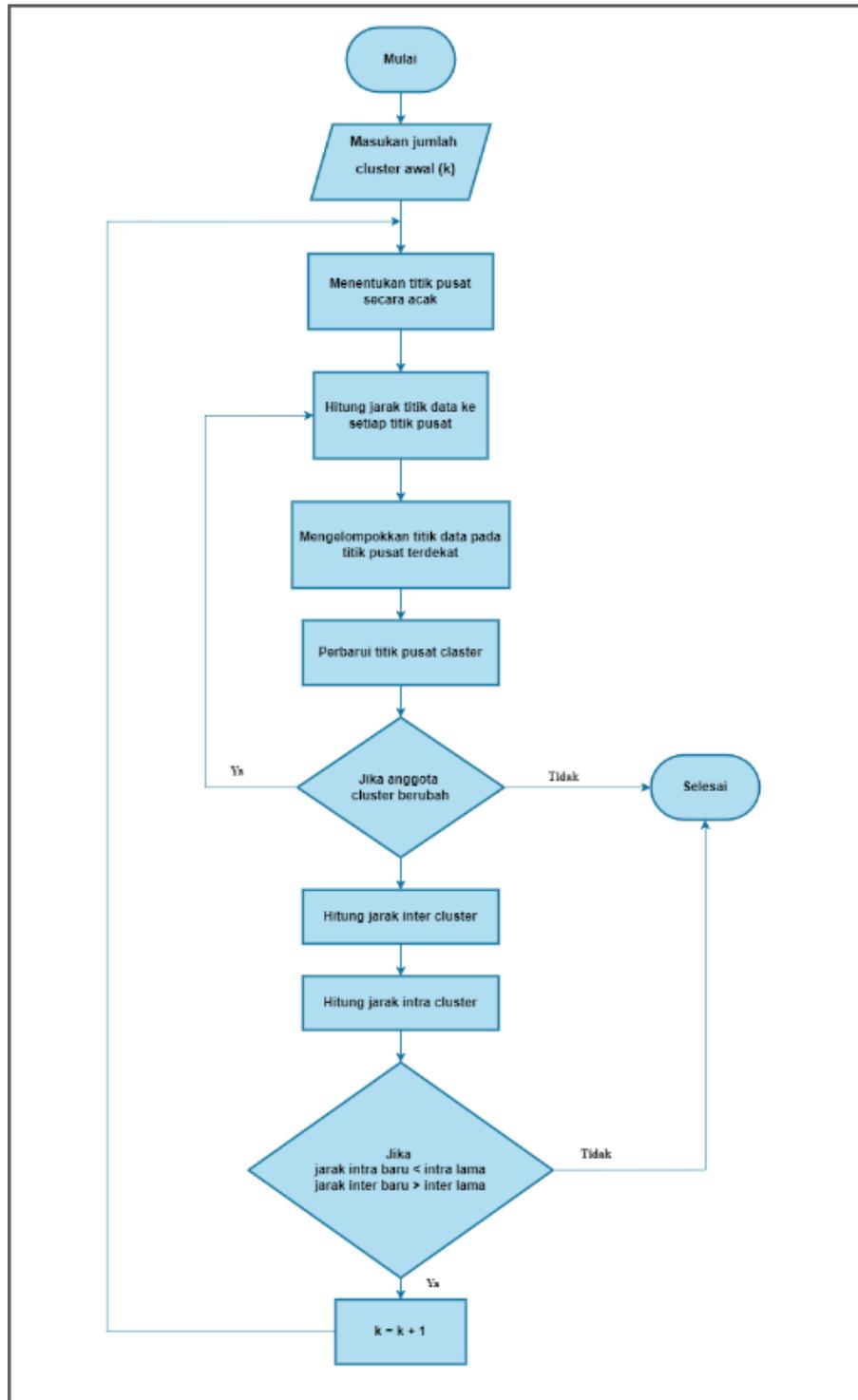
Gambar 2.4 Jarak *Inter Cluster*
(Sumber : Neural Computing and Applications)

Berdasarkan Gambar 2.4, jarak antar titik pusat berwarna hijau diistilahkan sebagai *inter*.

2.2. Algoritma *Dynamic K-means* (DK)

Dynamic K-means (DK) adalah algoritma yang dikembangkan dari algoritma *K-means*. Cara kerjanya sama dengan algoritma *K-means*, namun ketika anggota pada satu *cluster* sudah tidak berubah atau tidak berpindah ke *cluster* lain dilakukan perhitungan jarak *intra* dan *inter cluster*. Jika jarak *intra* semakin kecil dan jika jarak *intra* semakin besar, maka algoritma akan menghitung *cluster* baru dengan menambahkan jumlah *cluster* dengan satu atau $k=k+1$ disetiap iterasi (Widiarina & Romi, 2015).

Hasil *clustering* untuk algoritma *Dynamic K-means* sangat dipengaruhi oleh titik *centroid* data yang dibuat pada awal proses *clustering* karena pada proses pembentukan *cluster* berdasarkan titik *centroid* yang dibuat, akan berpengaruh pada jarak *intra* dan *inter cluster*. Berikut ini merupakan langkah dari algoritma *Dynamic K-means* yang dijelaskan di Gambar 2.5.



Gambar 2.5 Flowchart Algoritma *Dynamic K-means*
(Sumber : Widiarina & Romi, 2015)

Penjelasan flowchart di Gambar 2.5 adalah sebagai berikut :

1. Masukan jumlah *cluster* (k).
2. Menentukan titik *centroid* data.

Penentuan awal titik *centroid* data *cluster* ini bisa dilakukan dengan berbagai cara, namun yang paling sering dilakukan adalah dengan cara acak oleh proses algoritma.

3. Kelompokkan semua data atau objek yang memiliki kemiripan dengan cara menghitung jarak ke titik *centroid* terdekat. Jarak ini diartikan sebagai kesamaan suatu objek data dengan titik *centroid* yang telah ditentukan. Pada proses *clustering*, tahapan menentukan atau mendeskripsikan nilai kuantitatif dari tingkat kemiripan atau ketidakmiripan data (*proximity measure*) memiliki peranan sangat penting, sehingga perlu dilakukannya perbandingan beberapa metode yang sering digunakan, yaitu jarak *Euclidean*, *Manhattan*, dan *Minkowski*.

a. *Euclidean distance*

Kedekatan dua objek ditentukan berdasarkan perhitungan jarak objek tersebut. Untuk menghitung jarak semua data ke setiap titik *centroid cluster* menggunakan teori jarak *Euclidean distance* koordinat (Irani, Pise, & Phatak, 2016) yang dirumuskan sebagai berikut:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots (2.1)$$

b. *Manhattan distance*

Manhattan distance digunakan untuk menghitung perbedaan absolut (mutlak) antara koordinat sepasang objek. Rumus yang digunakan adalah sebagai berikut:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \dots\dots\dots (2.2)$$

c. *Minkowski distance*

Minkowski distance merupakan sebuah metrik dalam ruang vektor di mana suatu norma didefinisikan (*normed vector space*) sekaligus dianggap sebagai generalisasi dari *Euclidean distance* dan *Manhattan distance*. Dalam pengukuran jarak objek menggunakan *Minkowski distance* biasanya digunakan nilai p adalah 1 atau 2. Berikut rumus yang digunakan menghitung jarak dalam metode ini.

$$d(x, y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p} \dots\dots\dots (2.3)$$

Keterangan :

d = jarak antara x dan y
 x = data pusat *cluster*
 y = data pada atribut
 i = setiap data
 n = jumlah data,
 x_i = data pada pusat *cluster* ke i
 y_i = data pada setiap data ke i
 p = *power*

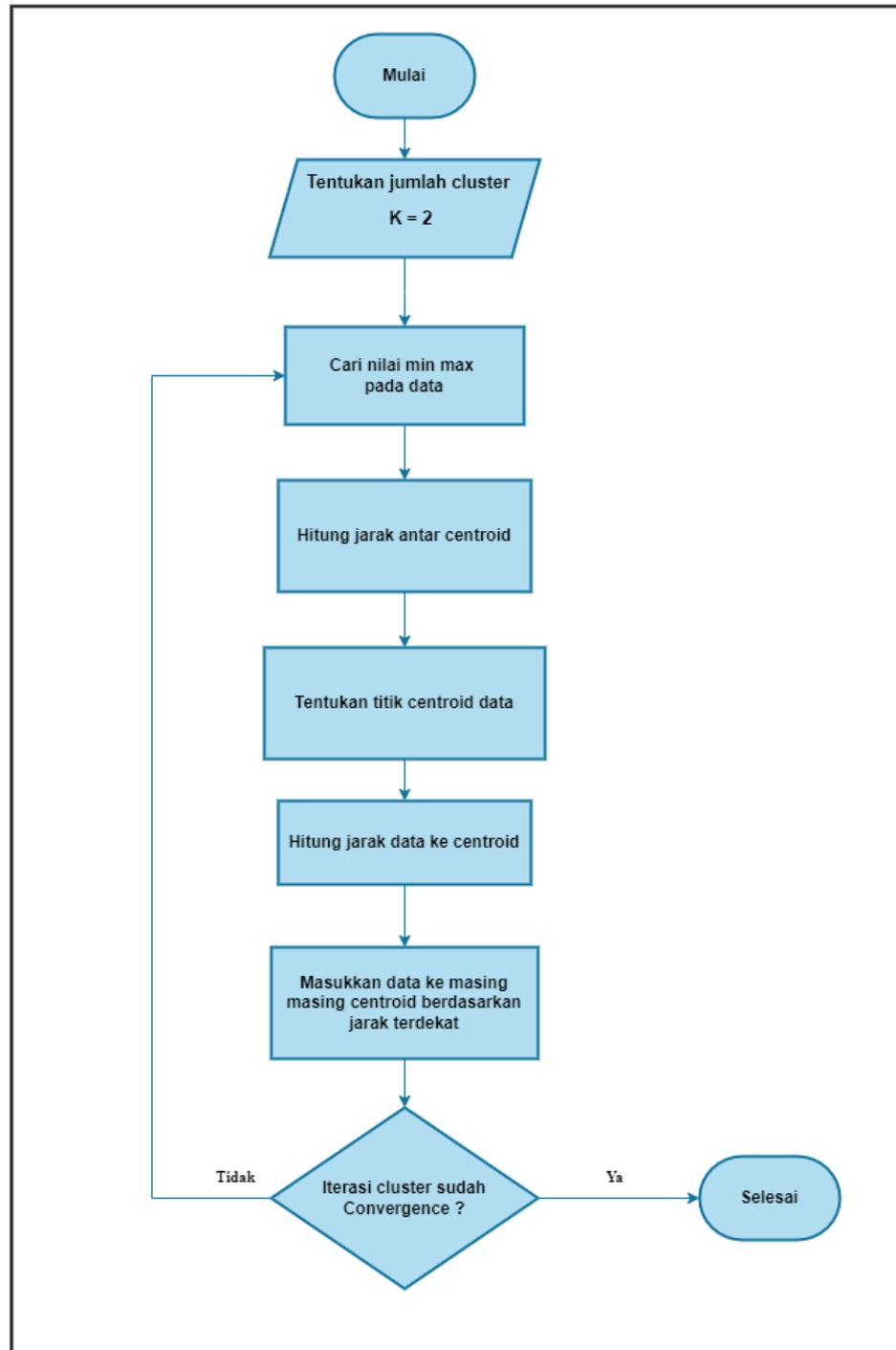
4. Hitung kembali titik *centroid* baru berdasarkan keanggotaan yang terbentuk. Titik *centroid* baru adalah rata-rata dari semua data atau objek dalam *cluster* tertentu.
5. Jika anggota *cluster convergence* tidak berubah maka lakukan pada proses hitung *intra cluster* dan juga *inter cluster*.
6. Jika *intra* baru < *intra* lama dan *inter* baru > *inter* lama, maka jumlah *cluster* akan ditambahkan dengan $k=k+1$ dan kembali pada proses kedua menentukan titik *centroid* data dan mengalokasikan data. Jika tidak maka proses *cluster* akan berhenti pada $k=k$.

Kelebihan dari algoritma *Dynamic K-means* yaitu dalam penentuan jumlah *cluster*, jumlah *cluster* yang dihasilkan berdasarkan perbandingan nilai *intra* dan

inter cluster. Namun kelemahan dari *Dynamic K-means* adalah ketika dalam proses pembentukan titik *centroid*, jika titik *centroid* dilakukan secara acak maka akan berpengaruh pada nilai *intra* dan *inter cluster* yang menyebabkan pembentukan untuk jumlah *cluster* akan berubah-ubah tidak konsisten, dan proses *cluster* memungkinkan akan terjadi kondisi yang membutuhkan iterasi yang banyak untuk membentuk *cluster convergence* (Gumilar & Yusrila, 2017).

2.3. Algoritma *K-means Binary Search Centroid* (KBSC)

K-means Binary Search Centroid (KBSC) merupakan pengembangan dari algoritma *K-means* dengan cara penggabungan antara algoritma *K-means* dengan algoritma tambahan pada area proses pembentukan titik *centroid* data menggunakan pendekatan *Binary Search* (Kumar, 2014). Algoritma *K-means Binary Search Centroid* (KBSC) ini dibuat oleh Hatamlao (Hatamlou, 2012). Berikut ini merupakan gambar dari *flowchart* mengenai langkah-langkah pada algoritma ini.



Gambar 2.6 Flowchart *K-means Binary Search Centroid*
(Sumber : Santoso, 2022)

Penjelasan Gambar 2.6 adalah sebagai berikut ini :

1. Tentukan jumlah *cluster* (k) nilai $k=1, 2, 3 \dots m$.
2. Hitung nilai maximum dan minimum pada data untuk masing-masing atribut data.

3. Hitung range antara titik *centroid*, perhitungan range antar titik *centroid* pada suatu data dengan cara berikut ini :

$$M = \frac{\max(a_i) - \min(a_i)}{k} \dots\dots\dots(2.4)$$

Berdasarkan rumus 2.2 digunakan untuk menghitung suatu nilai pada variabel M yang merupakan jarak spesifik antar titik *centroid data* sebelum memberikan hasil titik *centroid data*. $\min(a_i)$ nilai minimal dari masing masing atribut data, $\max(a_i)$ merupakan nilai maksimum dari masing-masing atribut data, k adalah jumlah *cluster* yang akan dibentuk.

4. Tentukan titik *centroid* data, Untuk menghasilkan titik *centroid* menggunakan rumus sebagai berikut :

$$C_k = \min(a_i) + (k - 1)M \dots\dots\dots (2.5)$$

Berdasarkan rumus 2.3, C_k merupakan *centroid* untuk *cluster* k , $\min(a_i)$ nilai minimum data, dan M range antara titik *centroid* data.

5. Lakukan pengelompokan data dengan cara hitung jarak data terhadap titik *centroid* yang terbentuk C_k . Perhitungan jarak menggunakan *Euclidian distance*.
6. Hitung kembali titik *centroid* berdasarkan *cluster* yang terbentuk, Kemudian lakukan pengelompokan ulang sampai menemukan *cluster* tidak berubah kembali (*convergence*).

Kelebihan dari algoritma KBSC adalah *centroid* yang dihasilkan memiliki nilai dari jarak *intra cluster* yang lebih kecil dibandingkan algoritma *K-means* penelitian yang dilakukan oleh Yugar Kumar dan G.Sahoo (Kumar, 2014). Sementara

kekurangannya adalah tidak dapat menentukan jumlah *cluster* terbaik dari suatu kumpulan data.

2.4.Evaluasi Hasil *Clustering*

Untuk mengetahui kualitas hasil *cluster* pada data *clustering* maka perlu dievaluasi menggunakan metode evaluasi yang sudah ada. Evaluasi *clustering* dalam penelitian ini menggunakan *Elbow*, *Davies Bouldin Index (DBI)* dan *Calinski Harabasz Index (CHI)*.

Metode *elbow* biasanya metode yang sering digunakan untuk menentukan jumlah banyaknya *cluster* yang akan terbentuk di suatu titik. Serta memberikan pengetahuan dalam memilih nilai *cluster*, dan menambahkan nilai *cluster* yang dijadikan sebagai model data untuk menentukan *cluster* terbaik. selain itu melakukan perhitungan persentase menerapkan visualisasi berupa grafik sebagai informasi. Apabila hasil perhitungan *cluster* pertama dengan hasil *cluster* kedua akan mengalami penurunan nilai, sehingga *cluster* tersebut termasuk *cluster* yang baik (Satriawan, 2021).

Setelah itu metode *elbow* perlu melakukan perhitungan dengan nilai SSE setiap *cluster* sesuai dengan titik pusatnya (Nurdiana, Nilogiri, & Rahman, 2022).

Pada metode *elbow* juga memiliki tujuan sebagai penentuan jumlah *cluster* terbaik. Adapun tahapan perhitungan metode *elbow* sebagai berikut (Nurdiana, Nilogiri, & Rahman, 2022):

1. Menetapkan nilai kluster terlebih dahulu.
2. Kemudian menaikkan nilai kluster sesuai yang sudah ditentukan.
3. Selanjutnya menghitung nilai SSE untuk tiap *cluster*.

4. Lalu menghitung nilai SSE hingga kluster ditetapkan.
5. Menganalisa hasil nilai SSE kluster yang turun secara teliti.
6. Dan terakhir kluster akan membentuk kurva

DBI memiliki fokus pada kemiripan antar *cluster* (Prasetyo, 2014), nilai kemiripan didasarkan pada nilai penyebaran dalam *cluster* dan nilai perbedaan antar *cluster*.

Nilai penyebaran dalam satu *cluster* diformulasikan oleh persamaan:

$$SSW_i = \frac{1}{m} \sum_{j=1}^{m_j} d(x_j, c_i) \dots \dots \dots (2.6)$$

Nilai perbedaan antar *cluster* diformulasikan oleh persamaan:

$$SSB_{i,j} = d(c_j, c_i) \dots \dots \dots (2.7)$$

Didefinisikan $R_{i,j}$ adalah ukuran rasio seberapa baik nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*. Nilainya didapatkan dari komponen kerapatan antar data dan separasi antar data. $R_{i,j}$ diformulasikan oleh persamaan berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \dots \dots \dots (2.8)$$

$$DBI = \frac{1}{K} \sum_{i=1}^K \max(R_{i,j}) \dots \dots \dots (2.9)$$

Keterangan:

m_j : adalah jumlah data yang berada dalam *cluster* ke-*i*

c_i : adalah *centroid cluster* ke-*i*

k : adalah jumlah *cluster*

Hasil dari rumus tersebut merupakan nilai yang menjadi ukuran validitas dari *cluster* yang diuji. Semakin kecil nilai *DBI* yang diperoleh (non-negatif ≥ 0), maka

semakin baik *cluster* yang diperoleh dari metode *clustering* yang digunakan (Prasetyo, 2014).

Indeks validitas Calinski-Harabasz (CH) menghitung perbandingan antara nilai *Sum of Square Between cluster (SSB)* sebagai separation dan nilai *Sum of Square within-cluster (SSW)* sebagai compactness yang dikalikan dengan faktor normalisasi, yaitu selisih jumlah data dengan jumlah *cluster* dibagi dengan jumlah *cluster* dikurang satu. Jumlah *cluster* terbaik ditunjukkan dengan semakin besar nilai CH (Baarsch & Celebi, 2012). Misalkan terdapat suatu himpunan data dengan k buah *cluster* dan N buah titik data, misal C_i adalah *cluster* ke- l dengan x_i adalah titik ke- i pada *cluster* ke- l , N_i adalah jumlah titik pada *cluster* ke- l , dan \bar{x}_i adalah titik pusat *cluster* ke- l , maka perhitungan indeks validitas CH dapat dilihat pada rumus berikut.

$$SSW = \sum_{i=1}^k \sum_{x_i \in C_i} (x_i - \bar{x}_i)(x_i - \bar{x}_i)^T \dots\dots\dots (2.10)$$

$$SSB = \sum_{i=1}^k N_i (x_i - \bar{x}_i)(x_i - \bar{x}_i)^T \dots\dots\dots (2.11)$$

$$CH = \frac{\text{trace}(SSB)}{\text{trace}(SSW)} \times \frac{N-k}{k-1} \dots\dots\dots (2.12)$$

2.5. Penelitian Terkait

Penelitian yang dilakukan (Gumilar Akbari, 2018) melakukan pengelompokan pelanggan berdasarkan nilai *Recency*, *Frequency*, dan *Monetary* dengan memiliki jumlah data 500. Pada penelitian ini menggabungkan algoritma antara algoritma *Dynamic K-means* dan *K-means Binary Search Centroid* yang akan diujikan pada data model buatan yang bertujuan untuk melihat karakteristik dari algoritma, dan diujikan pada data studi studi kasus yang bertujuan untuk mengetahui kemampuan algoritma dalam menyelesaikan kasus pengelompokan pelanggan. Berdasarkan

pengukuran Davies Bouldin Index (*DBI*) algoritma gabungan ini menghasilkan nilai *DBI* lebih baik dibandingkan algoritma lainnya. Saat diimplementasikan pada data kasus segmentasi pelanggan.

Penelitian yang telah dilakukan oleh (Zakir Hossain et al, 2019) yaitu mengusulkan algoritma pengelompokan *K-means* baru. Metode yang diusulkan adalah algoritma *clustering K-means* dinamis. Metode yang diusulkan awalnya menghitung nilai ambang sebagai *centroid* dari *KMeans* dan berdasarkan nilai ini jumlah *cluster* yang terbentuk. Di setiap iterasi *K-means*, jika jarak Euclidian antara dua titik lebih kecil dari atau sama dengan nilai ambang, maka kedua titik data ini akan berada di kelompok yang sama. Jika tidak, metode yang diusulkan akan membuat *cluster* baru dengan titik data yang berbeda. Hasil penelitian menunjukkan bahwa metode yang diusulkan mengungguli metode *K-means* yang asli.

Penelitian yang telah dilakukan oleh (Jiankun Jing et al, 2021) melakukan identifikasi litologi untuk memproses dan menafsirkan data rekaman geofisika. Metode *clustering* dinamis *K-means* digunakan untuk membagi sampel batuan menjadi tiga kategori, yang sesuai dengan hasil klasifikasi metode parameter sifat mekanik. Dengan 49 kelompok data, tingkat kesalahan penilaian masing-masing adalah 0,021, 0,021 dan 0,102, yang semuanya rendah. Oleh karena itu, metode analisis *K-means Dynamic clustering* layak dan efektif untuk mengidentifikasi litologi.

Penelitian yang dilakukan (Eric U. Oti, 2021) mengusulkan metode *clustering K-means* baru yang membahas masalah pusat *cluster* awal dalam algoritma *K-*

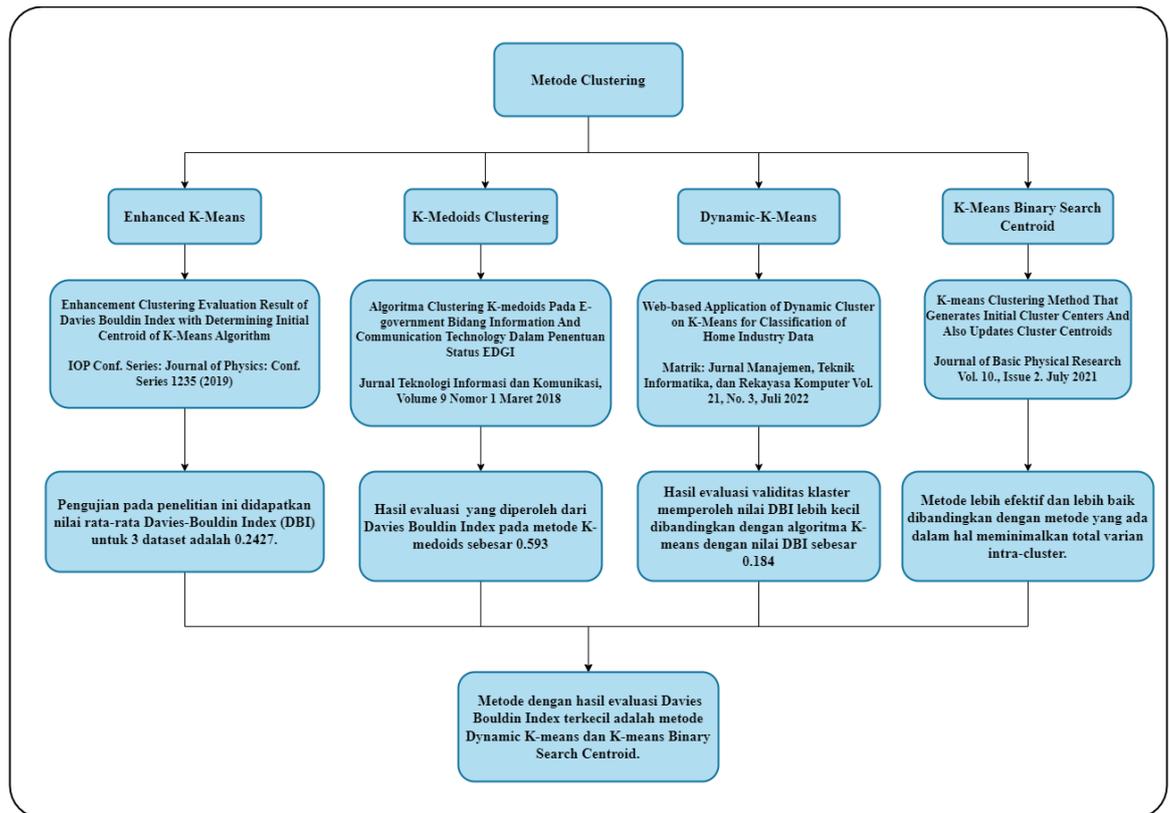
means berdasarkan teknik pencarian biner, dan juga memperbarui pusat *cluster* (*centroid*). Eksperimen dilakukan pada dua set data yang terdiri dari set data iris, dan kanker payudara. Metode yang diusulkan berkinerja baik dibandingkan dengan metode yang ada dalam hal meminimalkan total varians *intra-cluster*.

2.6 Kebaruan Penelitian

Penelitian ini memiliki keterbaruan dibandingkan penelitian sebelumnya yaitu pada penggabungan dua algoritma dengan studi kasus profil kesehatan Pulau Jawa dalam pengujian algoritma mengolah data dengan menggunakan evaluasi *Davies Bouldin Index* dan *Calinski Harabasz Index* disertai penambahan perhitungan jarak yg berbeda yaitu *Euclidean distance*, *Manhattan distance*, dan *Minkowski distance* juga visualisasi persebaran hasil *cluster* serta pengujian dengan jumlah data besar.

2.7. Kerangka Teori

Kerangka teori yang disusun dalam tugas akhir ini akan menggambarkan konsep pemecahan masalah yang dihadapi, kerangka teori dapat dilihat pada Gambar 2.7.



Gambar 2.7 Kerangka Teori

Pada Gambar 2.7 menunjukkan bahwa metode dengan hasil evaluasi *Davies Bouldin Index* terkecil adalah metode *Dynamic K-means* dan *K-means Binary Search Centroid*. Eksperimen akan dilakukan dengan mengelompokkan dengan menggabungkan dua algoritma dan membandingkan nya dengan algoritma *K-means* konvensional.

BAB III

METODOLOGI PENELITIAN

Dalam melaksanakan sebuah penelitian, dibutuhkan sebuah metode penelitian sehingga dapat meningkatkan kemungkinan berjalannya sebuah kegiatan penelitian dengan lancar. Metode penelitian berperan sebagai kerangka kerja dalam melaksanakan penelitian, sehingga dengan mengikuti kerangka kerja tersebut maka penelitian dapat berjalan secara sistematis dan diharapkan penelitian bisa lebih tepat guna dan dalam jangka waktu yang baik.

3.1 Roadmap Penelitian



Gambar 3.1 Roadmap Penelitian

Berdasarkan roadmap penelitian gambar 3.1 maka didapat premis sebagai berikut :

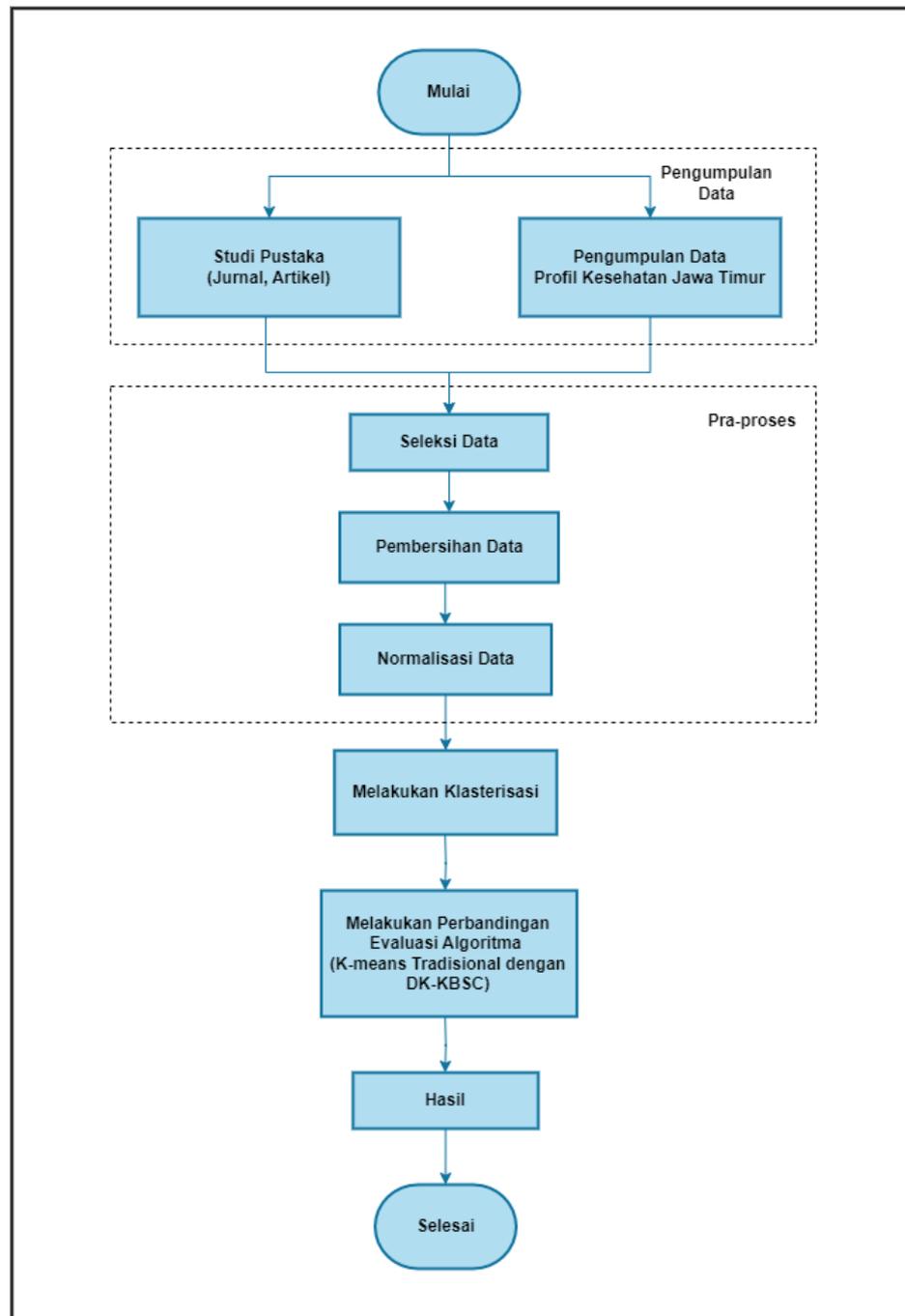
1. Algoritma *Dynamic K-means* memiliki kemampuan dalam mencari jumlah *cluster* dibandingkan, namun algoritma *Dynamic K-means* memiliki kekurangan dalam penentuana asumsi titik awal *centroid* yang berpengaruh terhadap jumlah iterasi proses untuk membentuk *cluster convergence* dan akan terjadi kondisi *local minimal* (hasil *cluster* belum tentu yang terbaik).

2. Algoritma *K-means Binary Search Centroid (KBSC)* memiliki kemampuan dalam menentukan asumsi titik *centroid* dan memiliki hasil *cluster* dengan nilai *intra cluster* (kemiripan dalam satu *cluster*) dan *inter cluster* (perbedaan data antar *cluster*) yang baik, namun masih memiliki kekurangan harus bisa menduga asumsi jumlah *cluster* yang akan dibentuk.

Sehingga muncul hipotesis bahwa pengabungan algoritma *K-means Binary Search Centroid (KBSC)* dan *Dynamic K-means (DK)* dapat menghasilkan *DBI cluster* yang lebih baik daripada algoritma lainnya ketika digunakan pada segmentasi wilayah data profil kesehatan pulau Jawa.

3.2 Tahapan Penelitian

Berikut merupakan kerangka kerja penelitian ini:



Gambar 3.2 Rancangan Alur Penelitian

3.3 Sumber Data

Pengumpulan data untuk penelitian ini mengutip publikasi dari Badan Pusat Statistik dan Dinas Kesehatan Pulau Jawa. Penelitian ini menggunakan data tahun 2020. Penelitian ini akan menggunakan *dataset* dari 85 kabupaten dan 34 Kota di Pulau Jawa yang terdapat pada lampiran.

3.4 Variabel Penelitian

Variabel penelitian yang digunakan merupakan variabel pembentuk Indikator Kesehatan Masyarakat, yang meliputi variabel kesehatan Umum, Fasilitas Kesehatan, dan epidemiologi dengan rincian sebagai berikut:

Tabel 3.1 Variabel Penelitian
(Sumber : Data Profil Kesehatan)

| Variabel | Nama Variabel |
|-----------------|---|
| X ₁ | Angka Harapan Hidup |
| X ₂ | Rasio Puskesmas |
| X ₃ | Rasio Rumah Sakit |
| X ₄ | Rasio Klinik |
| X ₅ | Persentase RT Sanitasi Layak |
| X ₆ | Persentase Sarana Air Minum |
| X ₇ | Persentase Bayi dengan BBLR |
| X ₈ | Persentase Imunisasi Dasar Lengkap |
| X ₉ | Persentase Bayi yang Mendapat Asi Eksklusif |
| X ₁₀ | Persentase Balita dengan Gizi Kurang |
| X ₁₁ | Persentase Balita dengan Gizi Kurang |
| X ₁₂ | Angka Kesakitan Diare (per 1rb penduduk) |
| X ₁₃ | Angka Kesembuhan Covid-19 |
| X ₁₄ | Angka Kematian Covid-19 |
| X ₁₅ | Persentase Penduduk Yang Merokok |
| X ₁₆ | Persentase Pelayanan Penderita Diabetes Melitus |

| Variabel | Nama Variabel |
|----------|--|
| X_{17} | Persentase Tempat Umum Memenuhi Syarat Kesehatan |

3.5 Tahap Pra-proses Data

3.5.1 Data Cleaning

Merupakan tahapan untuk membersihkan data yang tidak diperlukan dalam penelitian. Membersihkan data bertujuan untuk memeriksa apakah ada data yang tidak konsisten, duplikasi data, ataupun terdapat kesalahan pada data. Apabila terdapat data yang tidak sesuai maka akan langsung dihapuskan sehingga menyisakan data yang benar dan sesuai dengan data yang diperlukan dalam penelitian.

3.5.2 Data Transformation

Merupakan tahapan untuk menormalisasikan data yang bertujuan agar data yang digunakan berada pada range 0 sampai 1, selain itu normalisasi data juga bertujuan untuk mencegah agar tidak terjadinya perbedaan range yang sangat jauh antar masing-masing data dengan cara menghitung memakai *standard scaling*.

Suatu nilai dibakukan sebagai berikut:

$$y = (x - \text{rata-rata}) / \text{standar deviasi}$$

Di mana rata-rata dihitung sebagai:

$$\text{rata-rata} = \text{jumlah}(x) / \text{hitung}(x)$$

Dan standar deviasi dihitung sebagai:

$$\text{standar deviasi} = \sqrt{\text{sum}((x - \text{rata-rata})^2) / \text{hitung}(x)}$$

Berikut ini adalah contoh perhitungan dari *standard scaling* untuk data angka harapan hidup dan rasio puskesmas.

- Angka harapan hidup

$$y = (0.409 - 0.29) / 0.187$$

$$y = 0.655$$

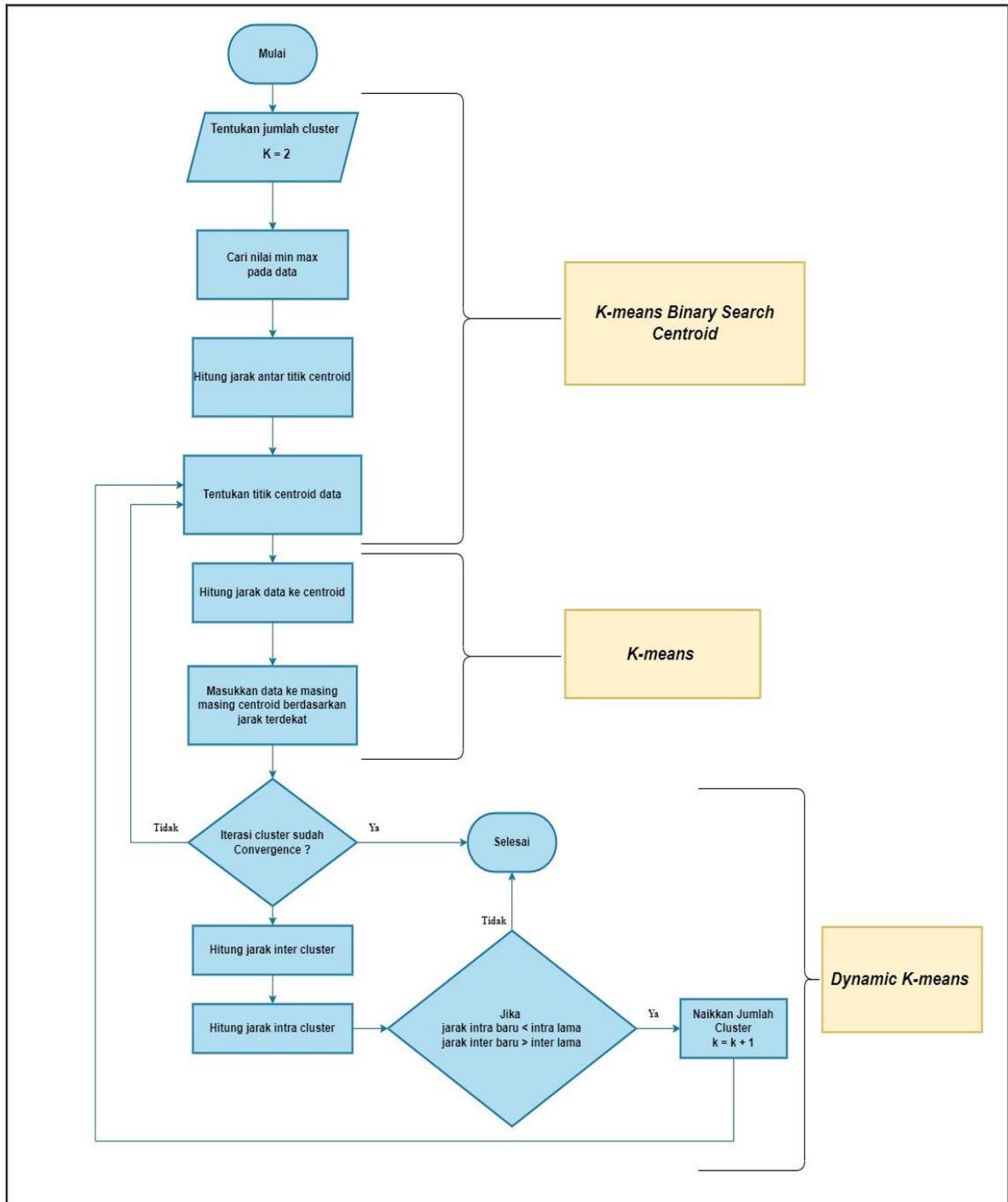
- Rasio puskesmas

$$y = (71,94 - 73) / 2.2371$$

$$y = -0.472$$

3.6 Tahap *Clustering*

Pada penelitian ini akan menggabungkan algoritma *K-means Binary Search Centroid (KBSC)* dan *Dynamic K-means (DK)* dengan kelebihan masing-masing algoritma untuk mencari jumlah *cluster* terbaik, dan menentukan titik *centroid* awal pada setiap pembentukan jumlah *cluster*. Berikut ini gambar mengenai penggabungan kedua algoritma:



Gambar 3.3 Alur Penggabungan Kedua Algoritma

Pada Gambar 3.3 menjelaskan alur penggabungan dari kedua algoritma. *Flowchart*. Titik *centroid* ini akan menggunakan teknik pencarian data dengan pendekatan *Binary Search* yaitu Pencarian bagi dua terhadap data (Made Sudarma,

2017). Proses tersebut akan membagi area data berdasarkan jumlah *cluster* yang dimasukan. Berikut ini penjelasan pada alur *flowchart*:

1. Masukan jumlah *cluster* sebanyak k

$$k = 2$$

2. Setelah itu proses dilanjutkan ke proses penentuan titik *centroid* menggunakan pendekatan *Binary Search Centroid*

Berikut ini merupakan contoh perhitungan titik *centroid* menggunakan algoritma *KBSC* :

Tabel 3.1 Contoh Data Perhitungan Pada Algoritma *KBSC*

| Data ke | Variabel 1 | Variabel 2 |
|---------|------------|------------|
| 1 | 1.1 | 4.3 |
| 2 | 1.3 | 3.9 |
| 3 | 1 | 3.8 |
| 4 | 2 | 4.8 |
| 5 | 2.1 | 3.9 |
| 6 | 2.2 | 3.7 |
| 7 | 2 | 3.6 |
| 8 | 1.9 | 6.9 |
| 9 | 2.2 | 3.2 |
| 10 | 9 | 3.2 |

Alur perhitungan titik *centroid* data sebagai berikut :

- a) Pada contoh data akan dibentuk jumlah *cluster* sebanyak $k=3$
- b) Hitung nilai maximum dan minimum pada masing-masing variabel pada data.

Berikut ini nilai maximum pada variable 1 dan 2 :

Di variabel 1 = 9

Di variabel 2 = 6.9

Berikut ini nilai minimum pada variable 1 dan 2 :

Di variabel 1 = 1.1

Di variabel 2 = 3.2

c) Hitung Nilai dari range antar titik *centroid*

$$M = \left\{ \frac{(9 - 1.1)}{3}, \frac{(6.9 - 3.2)}{3} \right\}$$

$$M = (2.63, 1.23)$$

d) Hitung titik titik *centroid* data berdasarkan jumlah *cluster* k yang dibentuk, hasil titik *centroid* data sebagai berikut :

$$\begin{aligned} C_1 &= (1.1 + ((1 - 1) * 2.63), 3.2 + ((1 - 1) * 1.23)) \\ &= (1.1, 3.2) \end{aligned}$$

$$\begin{aligned} C_2 &= (1.1 + ((2 - 1) * 2.63), 3.2 + ((2 - 1) * 1.23)) \\ &= (3.73, 4.43) \end{aligned}$$

$$\begin{aligned} C_3 &= (1.1 + ((3 - 1) * 2.63), 3.2 + ((3 - 1) * 1.23)) \\ &= (6.36, 5.66) \end{aligned}$$

Tabel 3.2 Contoh Data

| Data | Angka Harapan Hidup | Rasio Puskesmas |
|----------------------|---------------------|-----------------|
| Kabupaten Pacitan | -0.47289 | 0.65501 |
| Kabupaten Ponorogo | -0.10188 | 0.325176 |
| Kabupaten Trenggalek | 0.336171 | 0.149043 |

| | | |
|-----------------------|----------|----------|
| Kabupaten Tulungagung | 0.483678 | 0.135219 |
| Kabupaten Blitar | 0.233362 | -0.48259 |

$$M = \left\{ \frac{(0.483678 - (-0.47289))}{2}, \frac{(0.65501 - (-0.48259))}{2} \right\}$$

$$M = (0.747, 1.049)$$

$$\begin{aligned} C_1 &= (-0.47 + ((1 - 1) * 0.747), -0.48 + ((1 - 1) * 1.049)) \\ &= (-2.45, -1.84) \end{aligned}$$

$$\begin{aligned} C_2 &= (-0.47 + ((2 - 1) * 0.747), -0.48 + ((2 - 1) * 1.049)) \\ &= (-1.71, -0.79) \end{aligned}$$

3. Kemudian lakukan pengelompokan dengan cara menghitung jarak *minimum* antar data dengan *titik centroid* menggunakan *Euclidean distance*, *Manhattan distance*, dan *Minkowski distance* hingga tidak ada lagi titik data yang berpindah *cluster*. Di rumus 2.1 sampai 2.3 menjelaskan perhitungan jarak *Euclidean distance*, *Manhattan distance*, dan *Minkowski distance*.

Euclidean distance :

$$dC1 = \sqrt{(-0.47289 - (-2.45))^2 + (0.65501 - (-1.84))^2} = 11.245$$

$$dC2 = \sqrt{(-0.47289 - (-1.71))^2 + (0.65501 - (-0.79))^2} = 7.2608$$

Tabel 3.3 Alokasi *Cluster* Iterasi 1

| Data | C1 | C2 | Jarak Terdekat | <i>Cluster</i> |
|------------------|----------|----------|----------------|----------------|
| Kab. Pacitan | 11.24515 | 7.260843 | 7.260843 | 2 |
| Kab. Ponorogo | 12.0992 | 7.687441 | 7.687441 | 2 |
| Kab. Trenggalek | 12.01467 | 7.993445 | 7.993445 | 2 |
| Kab. Tulungagung | 12.4825 | 7.949861 | 7.949861 | 2 |
| Kab. Blitar | 13.2883 | 8.592614 | 8.592614 | 2 |

Tabel 3.4 Alokasi *Cluster* Iterasi 2

| Data | C1 | C2 | Jarak Terdekat | Cluster |
|------------------|----------|----------|----------------|---------|
| Kab. Pacitan | 2.321661 | 2.272538 | 2.272538 | 2 |
| Kab. Ponorogo | 1.315457 | 1.348878 | 1.315457 | 1 |
| Kab. Trenggalek | 1.854859 | 1.836203 | 1.836203 | 2 |
| Kab. Tulungagung | 2.71128 | 2.723493 | 2.71128 | 1 |
| Kab. Blitar | 2.236061 | 2.230485 | 2.230485 | 2 |

Tabel 3.5 Alokasi *Cluster* Iterasi 3

| Data | C1 | C2 | Jarak Terdekat | Cluster |
|------------------|----------|----------|----------------|---------|
| Kab. Pacitan | 2.321661 | 2.251585 | 2.251585 | 2 |
| Kab. Ponorogo | 1.315457 | 1.770563 | 1.315457 | 1 |
| Kab. Trenggalek | 1.854859 | 1.95961 | 1.854859 | 1 |
| Kab. Tulungagung | 2.71128 | 3.074395 | 2.71128 | 1 |
| Kab. Blitar | 2.236061 | 2.168235 | 2.168235 | 2 |

Tabel 3.6 Iterasi *Clustering*

| Data | AHH | RP | Iterasi 1 | Iterasi 2 | Iterasi 3 |
|------------------|---------|---------|-----------|-----------|-----------|
| Kab. Pacitan | -0.4728 | 0.6550 | 2 | 2 | 2 |
| Kab. Ponorogo | -0.1018 | 0.3251 | 2 | 1 | 1 |
| Kab. Trenggalek | 0.33617 | 0.1490 | 2 | 2 | 1 |
| Kab. Tulungagung | 0.48367 | 0.1352 | 2 | 1 | 1 |
| Kab. Blitar | 0.23336 | -0.4825 | 2 | 2 | 2 |

4. Setelah *cluster* tidak berubah, tahap selanjutnya masuk pada tahap proses *Dynamic K-means* yang mana prosesnya menghitung nilai *inter* dan nilai *intra*. Perhitungan *inter* dilakukan menggunakan perhitungan nilai *Sum of Square Within Cluster (SSW)* dengan persamaan 2.6 sedangkan perhitungan *intra* menggunakan perhitungan nilai *Sum of Square Between Cluster (SSB)* menggunakan Persamaan 2.7. Dan akan ditambahkan jumlah $k=k+1$ jika nilai *intra* baru $<$ *intra* lama, dan *inter* baru $>$ dari *inter* lama.

Dan akan kembali ke proses penentuan titik *centroid* kembali berdasarkan *cluster* yang terbentuk sebelumnya.

3.7 Tahap Analisis

Analisis dilakukan sejak penentuan data, proses *pre-processing*, dan implementasi perhitungan algoritma. Evaluasi *clustering* diuji tingkat validitasnya menggunakan beberapa metode yaitu *Davies Bouldin Index (DBI)*, dan *Calinski Harabasz Index (CHI)*. Semakin kecil nilai *Davies Bouldin Index* menunjukkan skema *cluster* yang paling optimal dan Semakin besar nilai *Calinski Harabasz Index* menunjukkan kualitas *cluster* yang semakin baik. Untuk mengetahui evaluasi dari model yang digunakan, dilakukan dengan beberapa skenario percobaan untuk menentukan algoritma yang paling sesuai dan akurat dalam evaluasi kinerja algoritma. Setelah semua hasil *cluster* terbentuk maka algoritma dibandingkan dan ditarik kesimpulan algoritma yang bekerja optimal, akurasi algoritma terbaik, dan jumlah *cluster* yang paling optimal berdasarkan kriteria data uji.

3.7.1 *Davies Bound Index (DBI)*

Untuk mendapatkan nilai *Davies Bouldin Index*, terlebih dahulu dihitung nilai *Sum of Square within-cluster*, *Sum of Square Between-cluster* dan *ratio*. Untuk memulai perhitungan manual *Davies Bouldin Index* diperlukan data yang sudah terbentuk pada *clustering* akhir dan titik pusat (*centroid*) *cluster* yang terakhir. Setelah mengetahui jarak setiap data dengan *centroid cluster* yang didapat dari iterasi terakhir pada *Clustering*, selanjutnya akan dilakukan perhitungan nilai *Sum of Square Within Cluster (SSW)* dengan persamaan 2.6.

$$SSW_1 = \frac{10.685 + 4.251 + 4.553 + 3.518 + 5.050 + 3.655 + 2.938 + 3.027 + 2.719 + 3.352 + 9.509}{11} = 4.8422$$

Perhitungan ini dilakukan hingga seluruh *centroid cluster* terpenuhi.

Tabel 3.7 Nilai *Sum of Square Within Cluster*

| | |
|---------|-------------|
| SSW_1 | 4.842294458 |
| SSW_2 | 3.351635504 |
| SSW_3 | 2.707551891 |
| SSW_4 | 2.430425955 |
| SSW_5 | 3.711048951 |

Setelah mengetahui nilai SSW maka dilakukan perhitungan *Sum of Square Between Cluster (SSB)*. Untuk menghitung nilai SSB maka diperlukan *centroid* terakhir pada iterasi terakhir. Untuk menghitung nilai *Sum of Square Between Cluster (SSB)* menggunakan Persamaan 2.7.

$$SSB_{1,2} = \sqrt{\begin{aligned} &(-0.37 - -0.80)^2 + (1.15 - -0.26)^2 + (0.99 - -0.40)^2 + (1.53 - 0.23)^2 + \\ &(0.31 - -0.94)^2 + (-2.12 - 0.18)^2 + (-0.97 - -0.57)^2 + (-0.03 - -0.54)^2 + \\ &(-0.15 - 0.03)^2 + (-0.69 - -0.76)^2 + (-0.13 - 0.14)^2 + (-0.01 - 0.30)^2 + \\ &(0.66 - 0.10)^2 + (-0.75 - -0.90)^2 + (0 - -1.03)^2 + (0.57 - -0.23)^2 + \\ &(0.83 - -0.27)^2 \end{aligned}}$$

$$= 4.063552$$

Perhitungan ini dilakukan hingga seluruh *centroid* terpenuhi.

Tabel 3.8 Nilai *Sum of Square Between Cluster*

| | |
|-------------|-------------|
| $SSB_{1,2}$ | 4.063552187 |
| $SSB_{1,3}$ | 5.016587768 |
| $SSB_{1,4}$ | 4.334535103 |
| $SSB_{1,5}$ | 4.755883503 |
| $SSB_{2,3}$ | 3.667454472 |
| $SSB_{2,4}$ | 3.255543042 |
| $SSB_{2,5}$ | 4.631184051 |
| $SSB_{3,4}$ | 2.488888428 |
| $SSB_{3,5}$ | 4.394454818 |
| $SSB_{4,5}$ | 3.752282955 |

Setelah dihitung nilai SSW dan SSB , akan dilakukan perhitungan rasio antar $cluster$ ke- i dan $cluster$ ke- j ($R_{i,j}$) menggunakan Persamaan 2.6.

$$R_{1,2} = \frac{(SSW_1 + SSW_2)}{SSB_{1,2}}$$

$$R_{1,2} = \frac{(4.842294458 + 3.351635504)}{4.063552187} = 2.016445116$$

Langkah terakhir dari metode *Davies Bouldin Index* yaitu menentukan rasio antar $cluster$ maksimum $R(max)$.

Setelah diketahui nilai rasio antar $cluster$ maksimum, selanjutnya dihitung nilai DBI . Untuk menghitung nilai DBI menggunakan Persamaan 2.7.

$$DBI = \frac{R_1 + R_2 + R_3 + R_4 + R_5}{k}$$

$$DBI = \frac{(2.0164 + 2.0164 + 1.5782 + 1.6778 + 1.7984)}{5} = 1.81748915$$

3.7.2 *Calinski Harabasz Index (CHI)*

Untuk mendapatkan nilai *Calinski Harabasz Index*, terlebih dahulu dihitung nilai jumlah *square within-cluster* dan jumlah *square Between-cluster* menggunakan Persamaan 2.8 dan Persamaan 2.9.

$$SSW = SSW_1 + SSW_2 + SSW_3 + SSW_4 + SSW_5$$

$$= 4.84229 + 3.35163 + 2.70755 + 2.43042 + 3.71104$$

$$= 17.04295676$$

$$SSB = SSB_{1,2} + SSB_{1,3} + SSB_{1,4} + SSB_{1,5} + SSB_{2,3} + SSB_{2,4} + SSB_{2,5} + SSB_{3,4} + SSB_{3,5} + SSB_{4,5}$$

$$SSB = 4.064 + 5.017 + 4.335 + 4.756 + 3.667 + 3.256 + 4.631 + 2.489 + 4.394 + 3.752$$

$$SSB = 40.36036633$$

Setelah dihitung nilai jumlah total seluruh SSW dan SSB , akan dilakukan perhitungan *Calinski Harabasz Index* menggunakan Persamaan 2.10.

$$CH = \frac{40.36036633}{17.04295676} \times \frac{119-5}{5-1} \quad CH = 67.49242262$$

BAB IV

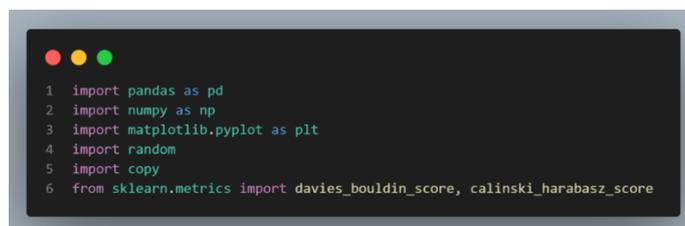
HASIL DAN PEMBAHASAN

4.1 Implementasi Sistem

Hasil penelitian dalam implementasi penggabungan algoritma *K-means Binary Search Centroid (KBSC)* dan *Dynamic K-means (DK)* dilakukan menggunakan data profil kesehatan pulau Jawa 2020 sebagai *dataset* penelitian. Penelitian ini untuk melakukan *clustering* diimplementasikan menggunakan bahasa pemrograman *Python*, sehingga dihasilkan penentuan tingkat kesehatan wilayah di Pulau Jawa berdasarkan hasil implementasi metode *K-means Binary Search Centroid (KBSC)* dan *Dynamic K-means (DK)*.

4.1.1 Import Library

Library merupakan kumpulan modul yang berisi kumpulan kode yang digunakan berulang kali dalam program yang berbeda. Dalam penelitian ini untuk mengimplementasikan metode *K-means clustering* dengan memanggil *library numpy, pandas, matplotlib, seaborn, datetime* yang digunakan untuk *dataframe* dan visualisasi. Kemudian, untuk melakukan *clustering* menggunakan *library sklearn, MinMaxScaler* dan *K-Means*. Berikut merupakan *Source code library* yang digunakan dapat dilihat pada Gambar 4.1.



```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import random
5 import copy
6 from sklearn.metrics import davies_bouldin_score, calinski_harabasz_score
```

Gambar 4.1 *Source Code Import Library*

Setelah melakukan *import library*, data set yang telah disiapkan dimasukan dengan cara memilih *dataset* tersebut. Berikut merupakan *Source code* untuk melakukan *load dataset* dapat dilihat pada Gambar 4.2



```

1 df = pd.read_csv('data.csv')
2 df

```

Gambar 4.2 *Source Code Load Dataset*

Maka dataset yang telah dimasukan terdiri dari beberapa atribut data dan variasi nilai yang akan dilakukan normalisasi. Hasil *load dataset* dapat dilihat pada Gambar 4.3.

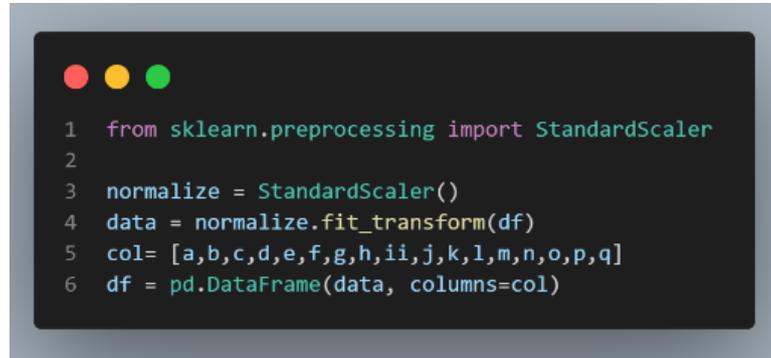
| wilayah | angka_harapan_hidup | rasio_puskesmas | rasio_rumah_sakit | rasio_klinik | sanitasi_layak | air_minum | berat_badan_lahir_rendah | imunisasi_dasar | pemberian_asj_eksklusif |
|-----------------------|---------------------|-----------------|-------------------|--------------|----------------|-----------|--------------------------|-----------------|-------------------------|
| Kabupaten Pacitan | 71.94 | 0.40948 | 0.03412 | 0.0290 | 69.90 | 84.43 | 5.5 | 67.3 | 64.80 |
| Kabupaten Ponorogo | 72.77 | 0.34762 | 0.06320 | 0.0432 | 84.97 | 91.15 | 4.0 | 76.5 | 78.80 |
| Kabupaten Trenggalek | 73.75 | 0.31458 | 0.02736 | 0.0191 | 76.81 | 80.64 | 3.9 | 92.4 | 50.60 |
| Kabupaten Tulungagung | 74.08 | 0.31199 | 0.08259 | 0.0303 | 89.45 | 96.75 | 3.1 | 91.5 | 61.40 |
| Kabupaten Blitar | 73.52 | 0.19612 | 0.06537 | 0.0172 | 80.36 | 95.35 | 3.8 | 91.5 | 92.30 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

Gambar 4.3 *Load Dataset*

4.1.2 Transformasi Data

Rescaling atribut yang bertujuan untuk menormalisasi atribut agar memiliki skala yang sebanding. Dalam penelitian ini menggunakan *StandardScaler*. *StandardScaler* adalah class dari *sklearn* untuk melakukan normalisasi data agar data yang digunakan tidak memiliki penyimpangan yang besar. Standarisasi dataset melibatkan penskalaan ulang distribusi nilai sehingga rata-rata nilai yang diamati adalah nilai bilangan bulat. Ini dapat dianggap sebagai

mengurangkan nilai rata-rata atau memusatkan data. Berikut merupakan *Source code Rescaling* yang dapat dilihat pada Gambar 4.4.



```

1 from sklearn.preprocessing import StandardScaler
2
3 normalize = StandardScaler()
4 data = normalize.fit_transform(df)
5 col= [a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q]
6 df = pd.DataFrame(data, columns=col)

```

Gambar 4.4 *Source code Rescaling*

Sehingga didapatkan hasil *Rescaling* yang dapat dilihat pada Gambar 4.5.

| angka_harapan_hidup | rasio_puskesmas | rasio_rumah_sakit | rasio_klinik | sanitasi_layak | air_minum | berat_badan_lahir_rendah | imunisasi_dasar |
|---------------------|-----------------|-------------------|--------------|----------------|-----------|--------------------------|-----------------|
| -0.284333 | 0.657780 | -5.942704 | -0.477144 | -0.537425 | -0.298819 | 0.849733 | -1.464524 |
| 0.038281 | 0.326560 | 0.201136 | -0.213068 | 0.275069 | 0.057502 | 0.018988 | -0.846122 |
| 0.419199 | 0.149653 | -7.370914 | -0.661254 | -0.164875 | -0.499780 | -0.036395 | 0.222639 |
| 0.547468 | 0.135785 | 4.297734 | -0.452968 | 0.516607 | 0.354436 | -0.479458 | 0.162143 |
| 0.329800 | -0.484623 | 0.659600 | -0.696588 | 0.026522 | 0.280203 | -0.091777 | 0.162143 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 0.236514 | -0.156562 | 0.061695 | 1.672661 | 0.523616 | -4.036480 | -1.753266 | 0.733493 |
| 0.368669 | 0.041870 | 0.061695 | 0.655409 | 0.762458 | -3.964368 | -1.974797 | 0.733493 |
| 0.563015 | 0.217867 | 0.061695 | 1.616870 | 0.550034 | -4.158966 | -1.974797 | 0.410848 |
| 0.699057 | -0.110033 | 0.061695 | 0.292769 | 0.967334 | -4.004666 | -1.974797 | 0.733493 |
| -1.520373 | 9.505192 | 0.061695 | -1.016455 | 0.129500 | -4.125030 | 1.015881 | 0.733493 |

Gambar 4.5 Hasil *Dataframe Rescaling*

4.1.3 Penentuan Titik *Centroid* Data

Proses pembentukan titik *centroid* data menggunakan pendekatan *Binary Search* dengan *centroid* yang dihasilkan memiliki nilai dari jarak *intra cluster* yang lebih kecil proses pembentukan titik *centroid* data secara acak. Berikut merupakan *Source code* pembentukan titik *centroid* yang dapat dilihat pada Gambar 4.6.

```

1 import numpy as np
2
3 def Initiate_Centroid_Binary_Search(df, x, y, k):
4     centroids = {}
5     for i in range(k):
6
7         df_x_min = df[x][i].min()
8         df_x_max = df[x][i].max()
9         df_y_min = df[y][i].min()
10        df_y_max = df[y][i].max()
11
12        #Rumus M=(max(a_i )-min(a_i))/k
13        mx = (df_x_max - df_x_min) / k
14        my = (df_y_max - df_y_min) / k
15
16        #Rumus C_k= min(a_i )+(k-1)M
17        cx = df_x_min + (k - 1) * mx
18        cy = df_y_min + (k - 1) * my
19
20        centroids [i+1] = [cx, cy]
21
22    return centroids

```

Gambar 4.6 Source code Pembentukan Titik *Centroid*

4.1.4 Perhitungan Jarak Ke Titik *Centroid* Terdekat

Jarak ini diartikan sebagai kesamaan suatu objek data dengan titik *centroid* yang telah ditentukan. Pada proses *Clustering*, tahapan menentukan atau mendeskripsikan nilai kuantitatif dari tingkat kemiripan atau ketidakmiripan data (*proximity measure*) memiliki peranan sangat penting, sehingga perlu dilakukannya perbandingan beberapa metode yang sering digunakan, yaitu jarak *Euclidean*, *Manhattan*, dan *Minkowski*.

4.1.5 *Euclidean distance*

Euclidean distance merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari 2 (dua) buah titik dalam *Euclidean space* (meliputi bidang *Euclidean* dua dimensi, tiga dimensi, atau bahkan lebih).

Berikut merupakan *Source code* perhitungan *Euclidean distance* yang dapat dilihat pada Gambar 4.7.

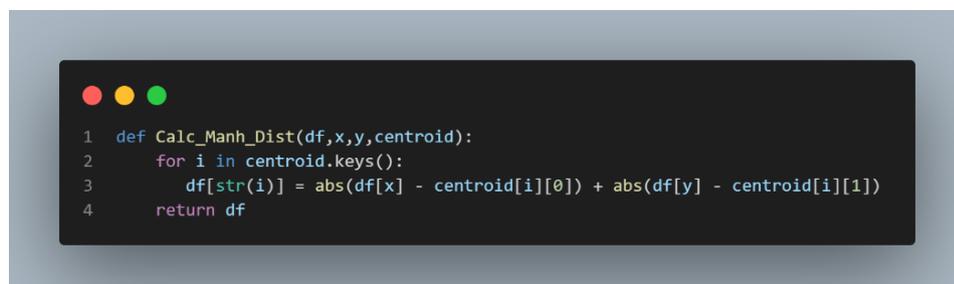


```
1 def Calc_Dist(df,x,y,centroid):
2     for i in centroid.keys():
3         df[str(i)] = np.sqrt((df[x] - centroid[i][0]) ** 2 + (df[y] - centroid[i][1]) ** 2)
4     return df
```

Gambar 4.7 *Source code* Perhitungan *Euclidean distance*

4.1.6 *Manhattan distance*

Manhattan distance digunakan untuk menghitung perbedaan absolut (mutlak) antara koordinat sepasang objek. Berikut merupakan *Source code* perhitungan *Manhattan distance* yang dapat dilihat pada Gambar 4.8.



```
1 def Calc_Manh_Dist(df,x,y,centroid):
2     for i in centroid.keys():
3         df[str(i)] = abs(df[x] - centroid[i][0]) + abs(df[y] - centroid[i][1])
4     return df
```

Gambar 4.8 *Source code* Perhitungan *Manhattan distance*

4.1.7 *Minkowski distance*

Minkowski distance merupakan sebuah metrik dalam ruang vektor di mana suatu norma didefinisikan (*normed vector space*) sekaligus dianggap sebagai generalisasi dari *Euclidean distance* dan *Manhattan distance*. Dalam pengukuran jarak objek menggunakan *Minkowski distance* biasanya digunakan

nilai p adalah 1 atau 2. Berikut merupakan *Source code* perhitungan *Minkowski distance* yang dapat dilihat pada Gambar 4.9.



```

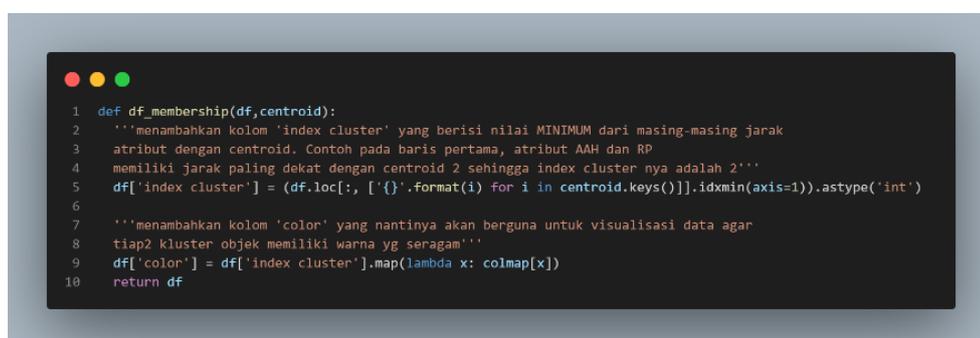
1 def Calc_Mink_Dist(df,x,y,centroid,p):
2     for i in centroid.keys():
3         df[str(i)] = (abs(df[x] - centroid[i][0])**p) + (abs(df[y] - centroid[i][1])**p)**(1/p)
4     return df

```

Gambar 4.9 *Source code* Perhitungan *Minkowski distance*

4.1.8 Membentuk Keanggotaan *Cluster*

Mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain. Tujuannya adalah meminimalisasikan fungsi obyektif yang diset dalam proses pengelompokan, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*. Berikut merupakan *Source code* dalam membentuk keanggotaan *cluster* yang dapat dilihat pada Gambar 4.10.



```

1 def df_membership(df,centroid):
2     '''menambahkan kolom 'index cluster' yang berisi nilai MINIMUM dari masing-masing jarak
3     atribut dengan centroid. Contoh pada baris pertama, atribut AAH dan RP
4     memiliki jarak paling dekat dengan centroid 2 sehingga index cluster nya adalah 2'''
5     df['index cluster'] = (df.loc[:, ['{}'].format(i) for i in centroid.keys()]).idxmin(axis=1).astype('int')
6
7     '''menambahkan kolom 'color' yang nantinya akan berguna untuk visualisasi data agar
8     tiap2 kluster objek memiliki warna yg seragam'''
9     df['color'] = df['index cluster'].map(lambda x: colormap[x])
10    return df

```

Gambar 4.10 *Source code* Pembentukan Keanggotaan *Cluster*

Sehingga didapatkan hasil yang dapat dilihat pada Gambar 4.11.

| persentase_tempat_umum_memenuhi_syarat_kesehatan | index | 1 | 2 | index cluster | color |
|--|-------|-----------|-----------|---------------|--------|
| -0.182265 | 1 | 51.174964 | 48.297457 | 2 | green |
| -0.220541 | 2 | 7.523796 | 10.051032 | 1 | orange |
| 0.037082 | 3 | 62.897891 | 63.902986 | 1 | orange |
| 0.104799 | 4 | 29.709937 | 31.528926 | 1 | orange |
| 0.065052 | 5 | 9.031624 | 13.410620 | 1 | orange |

Gambar 4.11 Hasil Pembentukan Keanggotaan *Cluster*

4.1.9 *Intra dan Inter Cluster*

Istilah *intra* digunakan untuk mengukur kekompakan dari suatu kelompok. Istilah *inter* adalah minimum jarak antar pusat *cluster*, *inter* digunakan untuk mengukur pemisahan antar *cluster* (Widiarina & Romi, 2015).

1. Menghitung Jarak *Intra Cluster* (*SSW*)

Jarak *intra* adalah jarak antar titik data dengan titik pusat (*centroid*) dalam suatu kelompok atau *cluster*, sehingga semakin kecil jarak *intra* maka akan semakin baik hasil *cluster* yang dihasilkan. Berikut merupakan *Source code* dalam menghitung jarak *intra cluster* (*SSW*) dapat dilihat pada Gambar 4.12.

```

1 def intra_c(df, x, y, centroid):
2     for i in centroid.keys():
3         dx = df[x] - centroid[i][0]
4         dy = df[y] - centroid[i][1]
5         dist = np.sqrt(dx ** 2 + dy ** 2)
6         total_dist = sum(dist)/len(centroid)
7     return total_dist

```

Gambar 4.12 *Source code* Menghitung Jarak *Intra Cluster* (*SSW*)

2. Menghitung Jarak *Inter Cluster* (SSB)

Jarak *inter* adalah jarak antar pusat *cluster* dengan pusat *cluster* lainnya, sehingga semakin besar jarak *inter* maka akan semakin baik hasil *cluster* yang dihasilkan. Berikut merupakan *Source code* dalam menghitung jarak *inter cluster* (SSB) dapat dilihat pada Gambar 4.13.



```
1 def inter_c(centroid):
2     a=0
3     for i in centroid.keys():
4         # print(centroid[i])
5         cx = centroid[i][0]
6         cy = centroid[i][1]
7         dist = np.sqrt(cx ** 2 + cy ** 2)
8         a += dist
9     total_dist = a/len(centroid)
10    return total_dist
```

Gambar 4.13 *Source code* Menghitung Jarak *Inter Cluster* (SSB)

4.1.10 *Dynamic Cluster*

Dynamic Cluster adalah algoritma yang dikembangkan untuk melakukan optimasi pada *K-means*. *Dynamic Cluster* menambahkan algoritmanya untuk menutupi sensitifitas penentuan dugaan jumlah *cluster* pada *K-means*, sehingga algoritma ini dapat menemukan jumlah *cluster* yang optimal. Berikut merupakan *Source code* dalam proses *Dynamic cluster* dapat dilihat pada Gambar 4.14.

```

1 def clusterisasi_dinamis_binary_search(df,x,y,k):
2     k=k
3     centroid = Initiate_Centroid_Binary_Search(df,x,y,k)
4     df2 = copy.deepcopy(df)
5     df2 = Calc_Dist(df2,x,y,centroid)
6     df2 = df_membership(df2,centroid)
7     centroid = rearrange_centroid(df2,centroid)
8     while (True):
9         oldcentroid = copy.deepcopy(centroid)
10        df2 = Calc_Dist(df2,x,y,centroid)
11        df2 = df_membership(df2,centroid)
12        centroid = rearrange_centroid(df2,centroid)
13        if (oldcentroid == centroid):
14            # lanjut dengan menghitung jarak intra dan inter cluster
15            inter_dist = inter_c(centroid)
16            intra_dist = intra_c(df2,x,y,centroid)
17            while (True):
18                oldinter = copy.deepcopy(inter_dist)
19                oldintra = copy.deepcopy(intra_dist)
20                centroid = Initiate_Centroid_Binary_Search(df,x,y,k)
21                df3 = copy.deepcopy(df)
22                df3 = Calc_Dist(df3,x,y,centroid)
23                df3 = df_membership(df3,centroid)
24                centroid = rearrange_centroid(df3,centroid)
25                inter_dist = inter_c(centroid)
26                intra_dist = intra_c(df3,x,y,centroid)
27                print("Intra K"+str(k)+ " - " + "new : " + str(intra_dist)+ " - old : " + str(oldintra))
28                print("Inter K"+str(k)+ " - " + "new : " + str(inter_dist)+ " - old : ' + str(oldinter))
29                # Jika intra baru < intra lama dan inter baru > inter lama, maka jumlah cluster akan
30                # ditambahkan dengan k=k+1 dan kembali pada proses kedua menentukan titik centroid
31                print("intra status ",intra_dist > oldintra)
32                print("inter status ",inter_dist < oldinter)
33                if intra_dist > oldintra and inter_dist < oldinter:
34                    print("selesai")
35                    return df3, centroid
36                    break
37            else:
38                print("ulang")
39                newK = k+1
40                return clusterisasi_dinamis_binary_search(df,x,y,newK)

```

Gambar 4.14 Source code Dynamic Cluster

Algoritma *Dynamic Cluster* akan menghitung jarak *inter* dan jarak *intra* dari proses sebelumnya, jika jarak *intra* baru lebih kecil dari jarak *intra* lama dan jarak *inter* baru lebih besar dari jarak *inter* lama maka algoritma *Dynamic Cluster* akan menambahkan satu pada jumlah *cluster* sebelumnya, proses ini akan dilakukan berulang hingga algoritma *Dynamic Cluster* menemukan jarak *intra* yang sekecil mungkin dan jarak *inter* yang sebesar mungkin. Sehingga didapatkan hasil yang dapat dilihat pada Gambar 4.15.

```
Intra K4 - new : 169.23528999557604 - old : 169.23528999557604
Inter K4 - new : 6.031083168516582 - old : 6.031083168516582
intra status False
inter status False
ulang
Intra K5 - new : 93.77877640456629 - old : 90.36388073624538
Inter K5 - new : 5.249192727420942 - old : 5.40030891043085
intra status True
inter status True
selesai
```

Gambar 4.15 Hasil Proses *Dynamic Cluster*

4.2 Evaluation

Pada tahap ini dilakukan evaluasi metode *K-means clustering* untuk menguji *cluster* yang dihasilkan menggunakan *Elbow*, Davies Bouldin Index (*DBI*) dan *Calinski Harabasz Index (CHI)*.

1. Pengujian *ELBOW*

Metode *Elbow* merupakan suatu metode yang dapat digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik tertentu. Metode ini memberikan ide/gagasan dengan cara memilih nilai *cluster* dan kemudian menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan *cluster* terbaik. *Source code* untuk pengujian *Elbow* dapat dilihat pada Gambar 4.16.

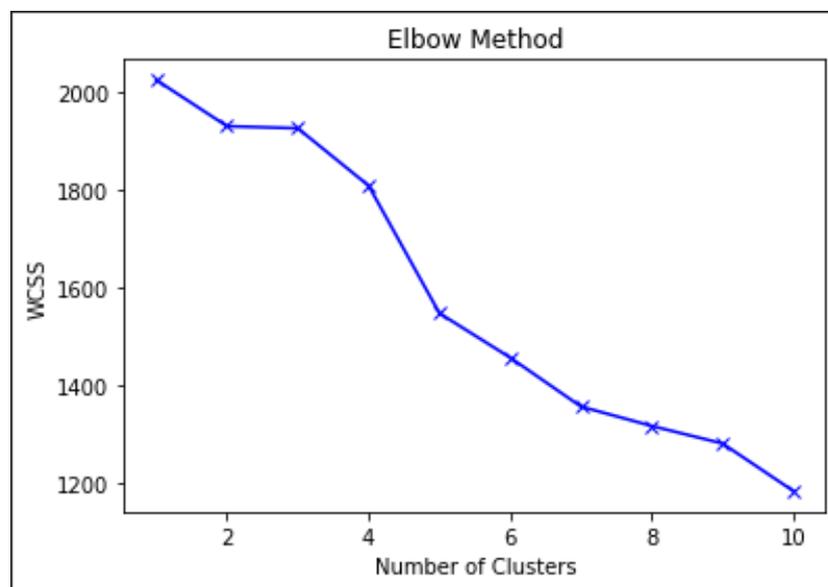
```

1 def sse(df,centroid):
2     sum = 0
3     for i in range(len(centroid)):
4         x = i+1
5         y = df[(df['index cluster'] == x)][['{}'.format(x)]**2
6         sum += y.sum()
7     return sum
8
9     wcss = []
10    for i in range(10):
11        df2, centroid = clusterisasi_binary_search(df,a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,i+1)
12        wcss.append(sse(df2, centroid))
13        print('K=',i+1,' | Distortion: ',wcss[i])
14
15    plt.plot(range(1, 11), wcss, '-bx')
16    plt.title('Elbow Method')
17    plt.xlabel('Number of Clusters')
18    plt.ylabel('WCSS')
19    plt.show()

```

Gambar 4.16 Source code Evaluation Metode Elbow

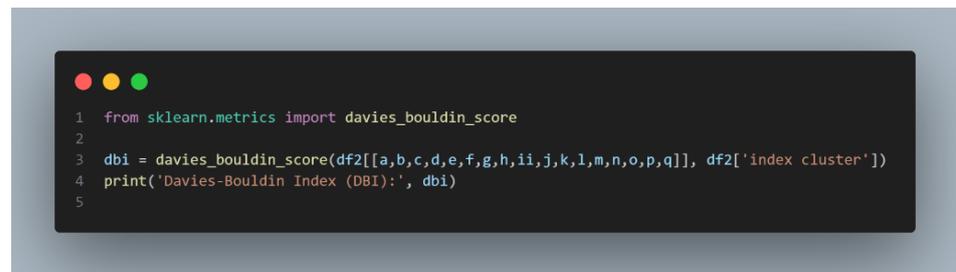
Pada Gambar 4.17 merupakan hasil *Evaluation* metode *Elbow* didapatkan hasil *cluster* terbaik yaitu pada *cluster* 5. Karena *cluster* terbaik merupakan *cluster* yang membentuk siku sehingga sesuai dengan hasil jumlah *cluster* pada proses *Dynamic Cluster*.



Gambar 4.17 Hasil Evaluasi Metode Elbow

2. Pengujian Davies Bouldin Index (*DBI*)

Pengujian *DBI* dilakukan untuk mengukur evaluasi *cluster* yang didasarkan pada nilai separasi dan kohesi. *Source code* yang digunakan dapat dilihat pada Gambar 4.18.



```

1 from sklearn.metrics import davies_bouldin_score
2
3 dbi = davies_bouldin_score(df2[['a,b,c,d,e,f,g,h,ii,j,k,l,m,n,o,p,q]], df2['index cluster'])
4 print('Davies-Bouldin Index (DBI):', dbi)
5

```

Gambar 4.18 *Source code Evaluation Metode DBI*

Hasil dari pengujian metode *DBI* dapat dilihat pada Gambar 4.19. dari gambar tersebut dapat dijelaskan bahwa terdapat nilai *DBI* terbaik dengan nilai *DBI* 1.5153, karena semakin kecil nilai *DBI* maka *cluster* semakin optimal.



```

Davies-Bouldin Index (DBI): 1.5153309017377155

```

Gambar 4.19 Hasil Evaluasi Metode *DBI*

3. Pengujian *Calinski Harabasz Index*

Calinski Harabasz Index (CHI) menghitung perbandingan antara nilai *Sum of Square Between cluster (SSB)* sebagai *separation* dan nilai *Sum of Square within cluster (SSW)* sebagai *compactness* yang dikalikan dengan faktor normalisasi, yaitu selisih jumlah data dengan jumlah *cluster* dibagi dengan jumlah *cluster* dikurang satu.

```
1 from sklearn.metrics import calinski_harabasz_score
2
3 chi = calinski_harabasz_score(df2[[x,y]], df2['index cluster'])
4 print('Calinski-Harabasz Index (CHI):', chi)
```

Gambar 4.20 *Source code Evaluation Metode CHI*

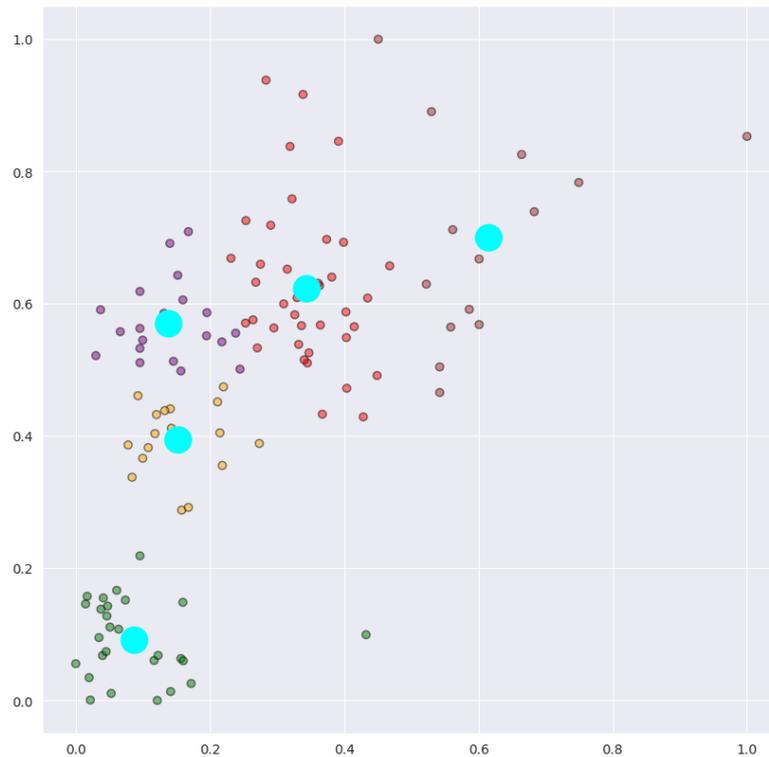
Hasil dari pengujian metode *CHI* dapat dilihat pada Gambar 4.21. dari gambar tersebut dapat dijelaskan bahwa terdapat nilai *CHI* terbaik dengan nilai 4.464317, karena *cluster* semakin optimal ditunjukkan dengan semakin besar nilai *CHI*.

```
Calinski-Harabasz Index (CHI): 4.464317428854567
```

Gambar 4.21 Hasil Evaluasi Metode *CHI*

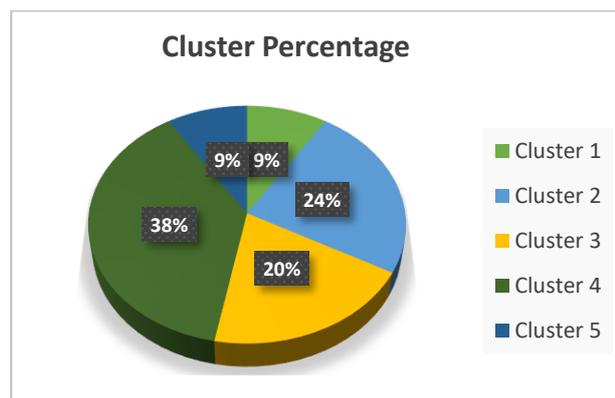
4.3 Hasil *Clustering*

Berdasarkan hasil *clustering* data profil kesehatan Pulau Jawa tahun 2020 menggunakan algoritma *K-means* dinamis dengan inisialisasi *cluster* awal *binary search centroid*, didapatkan visualisasi persebaran *cluster* pada Gambar 4.22.



Gambar 4.22 Hasil Visualisasi *Cluster*

Titik warna biru menunjukkan pusat *centroid* atau pusat *cluster* yang dikelilingi anggota *cluster* dimana setiap *cluster* memiliki warna berbeda. Untuk hasil presentase *cluster* dapat dilihat pada Gambar 4.33.



Gambar 4.23 Presentase Hasil *Clustering Dynamic K-means* dengan *Binary Search Centroid*

Berdasarkan presentase diatas didapatkan hasil 5 jumlah *cluster* di wilayah sebaran, kemudian diambil nilai rata rata setiap *cluster* untuk

mendefinisikan setiap *cluster* sehingga didapatkan hasil di antaranya *cluster* 1 dengan kualitas kesehatan sangat tinggi adalah 11 kabupaten/kota, *cluster* 2 dengan kualitas kesehatan sangat rendah adalah 24 kabupaten/kota, dan *cluster* 3 dengan kualitas kesehatan rendah adalah 28 kabupaten/kota. dan kualitas kesehatan 45 kabupaten/kota untuk *cluster* 4 dengan kualitas kesehatan cukup serta *cluster* 5 dengan kualitas kesehatan tinggi adalah 11 kabupaten/kota.

Tabel 4.1 Hasil *Clustering*

| | Kabupaten/Kota | |
|------------------|--|--|
| <i>Cluster 1</i> | Kota Sukabumi Kota Cirebon Kota Tangerang Kota Cilegon Kota Tangerang Selatan Kota Jakarta Pusat | Kota Jakarta Utara Kota Jakarta Barat Kota Jakarta Selatan Kota Jakarta Timur Kabupaten Kepulauan Seribu |
| <i>Cluster 2</i> | Kabupaten Pacitan Kabupaten Bogor Kabupaten Sukabumi Kabupaten Cianjur Kabupaten Bandung Kabupaten Garut Kabupaten Tasikmalaya Kabupaten Ciamis Kabupaten Kuningan Kabupaten Cirebon Kabupaten Majalengka Kabupaten Sumedang Kabupaten Indramayu Kabupaten Subang | Kabupaten Purwakarta Kabupaten Karawang Kabupaten Pangandaran Kota Bogor Kota Bandung Kota Bekasi Kota Depok Kota Cimahi Kota Tasikmalaya Kabupaten Lebak Kabupaten Pandeglang Kabupaten Serang Kabupaten Tangerang Kota Serang |
| <i>Cluster 3</i> | Kabupaten Kediri Kabupaten Lumajang Kabupaten Jember Kabupaten Banyuwangi Kabupaten Bondowoso Kabupaten Situbondo Kabupaten Probolinggo | Kabupaten Bangkalan Kabupaten Sampang Kabupaten Pamekasan Kota Kediri Kota Malang Kota Probolinggo Kota Pasuruan |

| | Kabupaten/Kota | |
|------------------|---|---|
| | Kabupaten Pasuruan Kabupaten Jombang Kabupaten Nganjuk Kabupaten Madiun Kabupaten Tuban | Kabupaten Grobogan Kabupaten Rembang Kabupaten Pekalongan Kabupaten Brebes Kota Banjar |
| <i>Cluster 4</i> | Kabupaten Ponorogo Kabupaten Trenggalek Kabupaten Tulungagung Kabupaten Blitar Kabupaten Malang Kabupaten Sidoarjo Kabupaten Mojokerto Kabupaten Magetan Kabupaten Ngawi Kabupaten Bojonegoro Kabupaten Lamongan Kabupaten Gresik Kabupaten Sumenep Kota Surabaya Kota Batu Kabupaten Cilacap Kabupaten Banyumas Kabupaten Purbalingga Kabupaten Banjarnegara Kabupaten Kebumen Kabupaten Purworejo Kabupaten Wonosobo Kabupaten Magelang | Kabupaten Boyolali Kabupaten Klaten Kabupaten Sukoharjo Kabupaten Wonogiri Kabupaten Karanganyar Kabupaten Sragen Kabupaten Blora Kabupaten Pati Kabupaten Kudus Kabupaten Jepara Kabupaten Demak Kabupaten Semarang Kabupaten Temanggung Kabupaten Kendal Kabupaten Batang Kabupaten Pemalang Kabupaten Tegal Kota Surakarta Kota Salatiga Kota Semarang Kota Pekalongan Kota Tegal |
| <i>Cluster 5</i> | Kota Blitar Kota Mojokerto Kota Madiun Kota Magelang Kabupaten Kuningan Kabupaten Bekasi | Kabupaten Kulon Progo Kabupaten Bantul Kabupaten Gunung Kidul Kabupaten Sleman Kota Yogyakarta |

Tabel 4.1 merupakan hasil sebaran wilayah *clustering Dynamic K-means* dengan *Binary Search Centroid* dari data profil kesehatan. Hasil dari *clustering* kemudian dapat dijadikan referensi untuk mengetahui persebaran

tingkat kesehatan wilayah, untuk kemudian dijadikan bahan evaluasi oleh dinas terkait

4.4 Hasil Perbandingan Algoritma

Analisa terhadap model sistem dilakukan dengan data yang sama dan tahapan setiap pengujian algoritma sesuai dengan proses perhitungan algoritma.

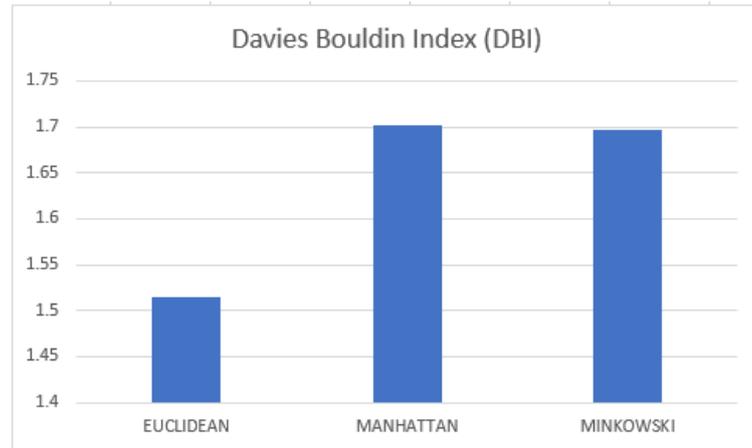
4.4.1 Perbandingan Perhitungan Jarak Ke Titik *Centroid* Terdekat

Berdasarkan hasil pengolahan 119 data kesehatan kabupaten/kota di Pulau Jawa, *cluster* optimal diperoleh dari perhitungan jarak *Euclidean*, *Manhattan*, dan *Minkowski* dengan menggunakan formula (2.1), (2.2) dan (2.3). Hasil perhitungan membentuk pengelompokkan sebagai berikut:

Tabel 4.2 Pengelompokkan *Euclidean-Manhattan-Minkowski*

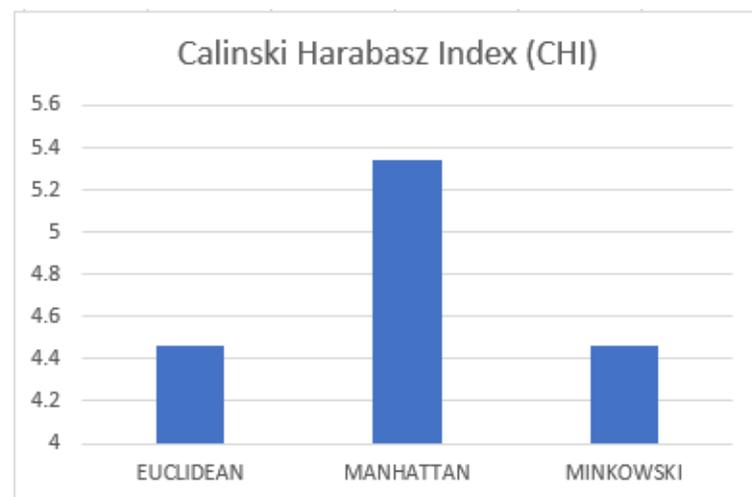
| | <i>DBI</i> | <i>CHI</i> |
|------------------|------------|------------|
| <i>EUCLIDEAN</i> | 1.51533 | 4.464317 |
| <i>MANHATTAN</i> | 1.70152 | 5.33885 |
| <i>MINKOWSKI</i> | 1.69687 | 4.46374 |

Hasil diperoleh dengan cara membandingkan perhitungan jarak menggunakan evaluasi metode *DBI* dan *CHI*. Hasil diagram *DBI* dapat dilihat pada Gambar 4.24.



Gambar 4.24 Diagram Evaluasi Jarak Metode *DBI*

Dapat dilihat bahwa nilai paling kecil terdapat pada metode perhitungan jarak *Euclidean* karena semakin kecil nilai *DBI* semakin baik. Kemudian Hasil diagram *DBI* dapat dilihat pada Gambar 4.25 dimana nilai terbaik terdapat pada metode perhitungan jarak *Manhattan* karena semakin besar nilai *CHI* semakin baik.



Gambar 4.25 Diagram Evaluasi Jarak Metode *CHI*

4.4.2 Perbandingan *K-means Dynamic Binary Search* Dan *K-means Tradisional*

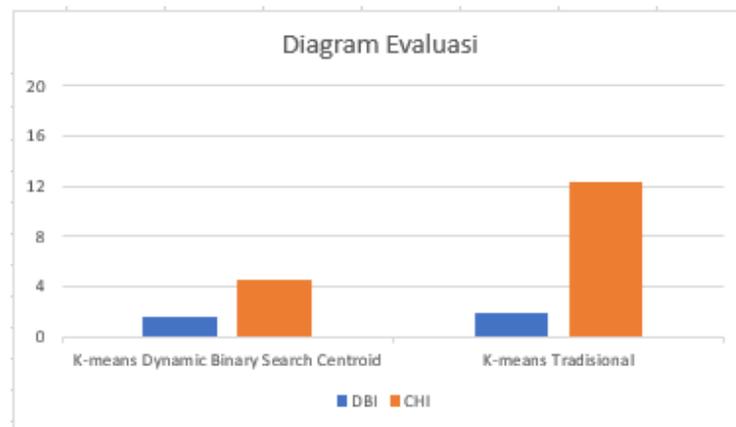
Tujuan dari pengujian ini untuk mengetahui kinerja algoritma dalam melihat karakteristik dengan data berdasarkan hasil *DBI cluster* dan *CHI*

cluster yang dihasilkan. Hasil pengujian membentuk pengelompokan sebagai berikut:

Tabel 4.3 Hasil Pengujian *KDBSC* Dengan *K-means* Tradisional

| | <i>DBI</i> | <i>CHI</i> |
|---|------------|-------------|
| <i>K-means Dynamic Binary Search Centroid</i> | 1.5153309 | 4.464317428 |
| <i>K-means Tradisional</i> | 2.1978855 | 12.4133754 |

Dapat dilihat bahwa nilai paling kecil terdapat pada metode *K-means Dynamic Binary Search Centroid* dibanding metode *K-means* tradisional, karena semakin kecil nilai *DBI* semakin baik dan Jumlah cluster terbaik ditunjukkan dengan semakin besar nilai *CH*. Kemudian Hasil diagram dapat dilihat pada Gambar 4.26.



Gambar 4.26 Diagram Evaluasi *DBI* dan *CHI*

4.4.3 Pengujian Menggunakan Dataset

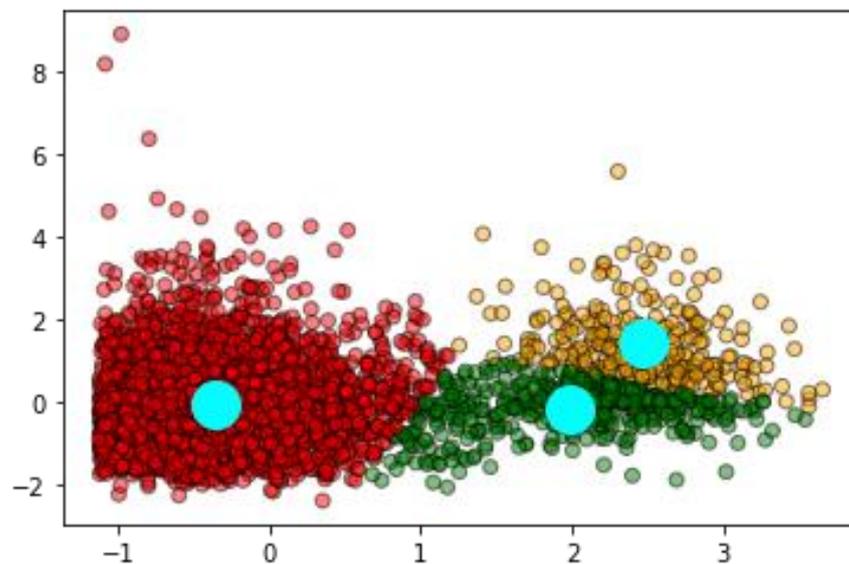
Pengujian akan dilakukan pada Data Model yaitu dataset Penyakit Stroke. Pengujian dilakukan adalah uji *DBI* dan *CHI* dengan menggunakan metode *K-means Dynamic Binary Search Centroid* pada total jumlah baris

data sebesar 5.110. Contoh dataset Penyakit Stroke dapat dilihat pada Gambar 4.27.

| id | gender | age | hypertension | heart_disease | ever_married | work_type | Residence_type | avg_glucose_level | bmi | smoking_status | stroke |
|-------|--------|------|--------------|---------------|--------------|---------------|----------------|-------------------|-----------|-----------------|--------|
| 9046 | Male | 67.0 | 0 | 1 | Yes | Private | Urban | 228.69 | 36.600000 | formerly smoked | 1 |
| 51676 | Female | 61.0 | 0 | 0 | Yes | Self-employed | Rural | 202.21 | 28.893237 | never smoked | 1 |
| 31112 | Male | 80.0 | 0 | 1 | Yes | Private | Rural | 105.92 | 32.500000 | never smoked | 1 |
| 60182 | Female | 49.0 | 0 | 0 | Yes | Private | Urban | 171.23 | 34.400000 | smokes | 1 |
| 1665 | Female | 79.0 | 1 | 0 | Yes | Self-employed | Rural | 174.12 | 24.000000 | never smoked | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 18234 | Female | 80.0 | 1 | 0 | Yes | Private | Urban | 83.75 | 28.893237 | never smoked | 0 |
| 44873 | Female | 81.0 | 0 | 0 | Yes | Self-employed | Urban | 125.20 | 40.000000 | never smoked | 0 |
| 19723 | Female | 35.0 | 0 | 0 | Yes | Self-employed | Rural | 82.99 | 30.600000 | never smoked | 0 |
| 37544 | Male | 51.0 | 0 | 0 | Yes | Private | Rural | 166.29 | 25.600000 | formerly smoked | 0 |
| 44679 | Female | 44.0 | 0 | 0 | Yes | Govt_job | Urban | 85.28 | 26.200000 | Unknown | 0 |

Gambar 4.27 Contoh Dataset Penyakit Stroke

Dataset diatas terlebih dahulu dilakukan *cleaning* data, *scaling* data, *clustering* menggunakan metode *K-means Dynamic Binary Search Centroid*, setelah itu dievaluasi sehingga didapatkan hasil visualisasi persebaran *cluster* yang dapat dilihat pada Gambar 4.28.



Gambar 4.28 Hasil Visualisasi Persebaran *Cluster* Dataset Penyakit Stroke

Dari hasil visualisasi persebaran *cluster* pada Gambar 4.28 terdapat penjelasan bahwa titik warna biru menunjukkan pusat *centroid* atau pusat *cluster* yang dikelilingi anggota *cluster* dimana setiap *cluster* memiliki warna berbeda.

Tabel 4.4 Hasil Uji *DBI* Dan *CHI Cluster* Pada Dataset Penyakit Stroke

| <i>DBI</i> | <i>CHI</i> |
|------------|-------------|
| 0.8265968 | 2054.421625 |

Dari Tabel 4.4 terdapat penejelasan yaitu, semakin kecil nilai *DBI* yang diperoleh (non-negatif ≥ 0), maka semakin baik cluster yang diperoleh dari metode clustering yang digunakan. Jumlah cluster terbaik ditunjukkan dengan semakin besar nilai *CHI*, serta jumlah dan variasi data mempengaruhi kinerja algoritma dan nilai evaluasi algoritma yaitu *DBI* dan *CHI*.

4.5 Integrasi Islam

Sistem yang dibangun dapat membantu proses pengelompokan tingkat kesehatan setiap wilayah, sehingga dapat menentukan prioritas pemberian bantuan dalam wilayah tersebut berdasarkan pada prioritas komponennya terhadap komponen- komponen yang lain. Dengan adanya sistem ini, diharapkan dapat digunakan sebagai salah satu metode terbaik dalam evaluasi tingkat kesehatan yang ada di wilayah khususnya Pulau Jawa serta acuan pengambilan keputusan dalam menentukan kebijakan bagi instansi terkait. Sistem pengelompokan juga terdapat dalam Al-Qur'an yaitu pengelompokan ayat Makkiyah dan Madaniyah. Makkiyah adalah ayat

yang diturunkan kepada Nabi Muhammad ﷺ sebelum melakukan hijrah dan Madaniyah adalah ayat yang diturunkan setelah Nabi Muhammad ﷺ hijrah ke Madinah. Dalam hal ini dapat dipahami bahwasanya surat atau ayat Makkiah dan Madaniyah memiliki karakteristik tersendiri yaitu terdapat perbedaan mulai dari proses dan waktu turunnya ayat tersebut. Perbedaan lain secara umumnya adalah bahwa surah dan ayat Makkiah membahas beberapa hal yang terkait kepada tauhid, janji dan ancaman, serta akidah dan akhlak. Namun berbeda dengan surah dan ayat madaniyah yang lebih berfokus kepada hukum, syaria, dan hubungan sosial karena pada masa Nabi Muhammad ﷺ di Madinah, penduduknya bukan hanya orang-orang Muhajirin dan Anshor namun sudah mencakup kepada lintas agama dan aliran diantaranya adalah Yahudi dan Nasrani. Hal ini sesuai dengan konsep *clustering* yang merupakan sebuah proses untuk melakukan pengelompokan data, kedalam beberapa *cluster* sehingga data didalam *cluster* tersebut memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda pada antar *cluster*.

Pengelompokan lain yang terdapat dalam Al-Qur'an juga dilakukan berdasarkan jumlah ayatnya, yang disebut pembagian atau pengelompokan surah-surah Al-Qur'an menjadi empat kelompok: *al-sab' al-tiwal*, *al-mi'un*, *al-masani*, dan *al-mufassal*. Istilah-istilah ini berasal dari suatu hadis yang diriwayatkan oleh Imam Ahmad dalam Musnad-nya dengan sanad hasan dari Watsilah bin al-Asqa.

عَنْ وَائِلَةَ رَضِيَ اللَّهُ عَنْهُ رَفَعَهُ أُعْطِيَتْ مَكَانَ الثَّوْرَةِ السَّبْعِ الطُّوْلَ وَأُعْطِيَتْ مَكَانَ
الرَّبُورِ الْمَائِينَ وَأُعْطِيَتْ مَكَانَ الْإِنْجِيلِ الْمَثَانِي وَفُضِّلَتْ بِالْمُقَصَّلِ.
(رواه احمد في الكبير كذا في جمع الفوائد)

Dari Watsilah RA, Rasulullah ﷺ bersabda, “Aku telah diberi Sab'a Thuwal sebagai pengganti Taurat, Mi'in sebagai pengganti Zabur, Matsani sebagai pengganti Injil, dan Mufashshal sebagai anugerah istimewa kepadaku.” (HR. Ahmad 16982 dan dihasankan Syuaib al-Arnauth).

Al-Sab' al-tiwal adalah tujuh surah panjang setelah Surah Al-Fatihah. Meliputi: al-Baqarah, Ali Imran, an-Nisa, al-Maidah, al-An'am, al-A'raf, dan al-Anfal. *Al-Mi'un* adalah sekelompok surat Al-Qur'an yang jumlah ayatnya mencapai seratus ayat atau lebih. Jumlah surah *al-Mi'un* juga ada tujuh, yaitu; surah Yunus, Hud, Yusuf, ar-Ra'd, Ibrahim, al-Hijr, dan an-Nahl. *Al-Masani* adalah surah-surah yang jumlah ayatnya kurang dari seratus ayat, selain surah-surah *al-mi'un*. Dari kata (تَنَّى) yang artinya mengulang. Menurut keterangan Ibnu Abbas radhiyallahu 'anhu dan Said bin Jubair, disebut demikian karena Allah banyak mengulang tentang kewajiban-kewajiban (*al-Faraid*), hukum-hukum syariat, kisah-kisah (*al-Qashas*), dan perumpamaan (*al-Amsal*). *Al-Mufassal* yang artinya sesuatu yang terpisah-pisah. Dinamakan demikian, karena jarak antar surahnya pendek-pendek dan banyak basmalah, sehingga terkesan banyak pemisahannya. Kemudian, menurut pendapat yang kuat, dimulai dari surat Qaf hingga surat an-Naas.

Islam yang pernah berjaya dengan segala bidang ilmu pengetahuannya, terkhusus di bidang kesehatan, sekarang hanya tinggal kenangan. Salah satu penyebab kemunduran ini adalah kurangnya pengetahuan dan perhatian umat islam terhadap ilmu kesehatan. Oleh karena itu, ilmu kesehatan dijadikan hal yang penting untuk mengembalikan dan memajukan peradaban muslimin.

Salah satu ciri yang membedakan Islam dengan yang lainnya adalah penekanannya terhadap masalah ilmu, Al-Qur'an dan hadist mengajak kaum muslim untuk mencari dan mendapatkan Ilmu dan kearifan, serta menempatkan orang-orang yang berpengetahuan pada derajat tinggi. Pentingnya menjaga nikmat sehat juga terdapat pada hadis Nabi Muhammad ﷺ.

عَنْ ابْنِ عُمَرَ رَضِيَ اللَّهُ عَنْهُمَا قَالَ : أَخَذَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ بِمَنْكِبِي فَقَالَ : كُنْ فِي الدُّنْيَا كَأَنَّكَ غَرِيبٌ أَوْ غَائِبٌ سَبِيلِي. وَكَانَ ابْنُ عُمَرَ رَضِيَ اللَّهُ عَنْهُمَا يَقُولُ : إِذَا أُمْسَيْتَ فَلَا تَنْتَظِرِ الصَّبَاحَ، وَإِذَا أَصْبَحْتَ فَلَا تَنْتَظِرِ الْمَسَاءَ، وَخُذْ مِنْ صِحَّتِكَ لِمَرَضِكَ، وَمِنْ حَيَاتِكَ لِمَوْتِكَ. [رواه البخاري]

Dari Ibn Umar ra berkata: Rasulullah ﷺ memegang kedua pundak saya seraya bersabda: "Hiduplah engkau di dunia seakan-akan orang asing atau pengembara," Ibnu Umar berkata: Jika kamu berada di sore hari jangan tunggu pagi hari, dan jika kamu berada di pagi hari jangan tunggu sore hari, gunakanlah kesehatanmu untuk (persiapan saat) sakitmu dan kehidupanmu untuk kematianmu (HR.al-Bukhari).

Islam merupakan agama yang sangat kompleks, karena mengatur segala aspek kehidupan baik masalah dunia ataupun akhirat. Al-Qur'an yang merupakan kalam ilahi dan menjadi pedoman bagi kehidupan manusia mengandung segala ilmu pengetahuan, termasuk ilmu kesehatan. Kesehatan merupakan hal yang sangat diperhatikan dalam ajaran islam. Karena kesehatan menjadi modal awal untuk beribadah kepada Allah secara optimal. Perhatian Islam terhadap kesehatan dibuktikan dengan banyaknya ayat-ayat Al-Qur'an dan hadis yang menjelaskan segala hal tentang kesehatan. Oleh karena itu, kita sebagai seorang muslim sudah seharusnya kita mendalami dan mengkaji Al-Qur'an dan Hadis yang merupakan

pedoman hidup dan sumber segala ilmu pengetahuan terkhusus ilmu kedokteran.

Penerapan *mu'amalah ma'a* Allah pada penelitian ini yaitu, dengan adanya pengelompokan data berbasis sistem manusia diberi kemampuan untuk mengembangkan teknologi sehingga memudahkan manusia dalam melakukan pekerjaan. Hasil pengujian dapat digunakan sebagai salah satu metode terbaik dalam evaluasi tingkat kesehatan yang ada di wilayah khususnya Pulau Jawa serta acuan pengambilan keputusan dalam menentukan kebijakan bagi instansi terkait. Melalui penelitian ini sekiranya dapat membantu untuk memudahkan para masyarakat dalam bekerja serta memudahkan masyarakat dalam mengakses informasi, sehingga masyarakat dapat selalu mengingat Allah *Subhanahu wa ta'ala* serta bersyukur dengan apa yang sudah diberikan Allah *Subhanahu wa ta'ala*. Pada Qur'an Surah Al-A'raf ayat 56 Allah *Subhanahu wa ta'ala* berfirman:

وَلَا تُفْسِدُوا فِي الْأَرْضِ بَعْدَ إِصْلَاحِهَا وَادْعُوهُ خَوْفًا وَطَمَعًا إِنَّ رَحْمَتَ اللَّهِ قَرِيبٌ مِّنَ الْمُحْسِنِينَ

“Dan janganlah kamu berbuat kerusakan di bumi setelah (diciptakan) dengan baik. Berdoalah kepada-Nya dengan rasa takut dan penuh harap. Sesungguhnya rahmat Allah sangat dekat kepada orang yang berbuat kebaikan.” (QS. Al-A'raf 7:56).

Pada *Zubdatut Tafsir Min Fathil Qadir* (TafsirWeb, n.d.) menjelaskan “Dalam ayat إِنَّ رَحْمَتَ اللَّهِ قَرِيبٌ مِّنَ الْمُحْسِنِينَ terdapat dorongan dan pemberian semangat bagi hamba-hamba-Nya untuk berbuat kebaikan. Dan (المحسنون) (yaitu orang-orang yang mengumpulkan dalam dirinya keimanan kepada Allah dan kepada hal yang ghaib, dan menjalankan kewajiban-kewajiban kepada Allah, dan menjauhi larangan-larangan-Nya,

serta merasa diawasi oleh Allah sehingga mereka memperbaiki amalan mereka”. Pada ayat tersebut Allah *Subhanahu wa ta’ala* menjelaskan bahwa rahmat Allah sangat dekat kepada orang yang berbuat kebaikan, diharapkan penelitian ini dapat bermanfaat sebagaimana semestinya serta dapat memberi kebaikan bagi para penggunanya dalam mengelompokan data data secara efektif dan efisien.

Selanjutnya yaitu penerapan *mu’amalah ma’a An-nas* pada penelitian ini yaitu, memudahkan pegawai instansi terkait dalam mengelompokan data wilayah berdasarkan faktor kesehatan sehingga pegawai instansi tidak perlu mengelompokan data secara manual serta memudahkan para pembaca dalam mengakses informasi dengan cepat. Oleh karena itu diharapkan penelitian ini dapat memberikan kemudahan bagi pegawai instansi terkait dalam mengelompokan data serta memberikan kemudahan bagi masyarakat dalam mencari informasi. Sesuai dengan hadis dari nabi yang terdapat dalam kitab *Arba’in An Nawawi* hadis ke 36

عَنْ أَبِي هُرَيْرَةَ رَضِيَ اللَّهُ عَنْهُ عَنِ النَّبِيِّ ﷺ قَالَ: مَنْ نَفَسَ عَنْ مُؤْمِنٍ كُرْبَةً مِنْ كُرْبِ الدُّنْيَا نَفَسَ اللَّهُ عَنْهُ كُرْبَةً مِنْ كُرْبِ يَوْمِ الْقِيَامَةِ، وَمَنْ يَسَّرَ عَلَى مُعْسِرٍ يَسَّرَ اللَّهُ عَلَيْهِ فِي الدُّنْيَا وَالْآخِرَةِ، وَمَنْ سَتَرَ مُسْلِمًا سَتَرَهُ اللَّهُ فِي الدُّنْيَا وَالْآخِرَةِ، وَاللَّهُ فِي عَوْنِ الْعَبْدِ مَا كَانَ الْعَبْدُ فِي عَوْنِ أَخِيهِ، وَمَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا سَهَّلَ اللَّهُ لَهُ بِهِ طَرِيقًا إِلَى الْجَنَّةِ، وَمَا اجْتَمَعَ قَوْمٌ فِي بَيْتٍ مِنْ بُيُوتِ اللَّهِ يَتْلُونَ كِتَابَ اللَّهِ وَيَتَادَرَسُونَ بَيْنَهُمْ إِلَّا نَزَلَتْ عَلَيْهِمُ السَّكِينَةُ وَغَشِيَتْهُمُ الرَّحْمَةُ وَالْحَفَنَةُ الْمَلَائِكَةُ وَذَكَرَهُمُ اللَّهُ فِيمَنْ عِنْدَهُ، وَمَنْ بَطَأَ بِهِ عَمَلُهُ لَمْ يُسْرَعْ بِهِ نَسْبُهُ (رواه مسلم بهذا اللفظ)

“Dari Abu Hurairah radhiallahuanhu, dari Rasulullah wasallam bersabda: Siapa yang menyelesaikan kesulitan seorang mu’min dari berbagai kesulitan-kesulitan dunia, niscaya Allah akan memudahkan kesulitan-kesulitannya di hari kiamat. Dan siapa yang memudahkan orang yang sedang kesulitan niscaya Allah mudahkan baginya di dunia dan akhirat dan siapa yang menutupi (aib) seorang muslim Allah akan tutupkan

aibnya di dunia dan akhirat. Allah selalumenolong hambanya selama hambanya menolong saudaranya. Siapa yangmenempuh jalan untuk mendapatkan ilmu, akan Allah mudahkan baginya jalankesyurga. Sebuah kaum yang berkumpul di salah satu rumah Allah membaca kitab- kitab Allah dan mempelajarinya di antara mereka, niscaya akan diturunkankepada mereka ketenangan dan dilimpahkan kepada mereka rahmat, dan mereka dikelilingi malaikat serta Allah sebut-sebut mereka kepada makhluk disisi-Nya. Dan siapa yang lambat amalnya, hal itu tidak akan dipercepat oleh nasabnya”. (HR. Muslim, no. 2699)

Salah satu perintah di dalam hadis tersebut menjelaskan bahwa apabila memudahkan kesulitan orang lain maka Allah akan memudahkan kesulitan kesulitannya di hari kiamat Diharapkannya sistem yang dibangun dapat membantu proses pengelompokan tingkat kesehatan setiap wilayah, sehingga dapat menentukan prioritas pemberian bantuan dalam wilayah tersebut berdasarkan pada prioritas komponennya terhadap komponen-komponen yang lain. Dengan adanya sistem ini, diharapkan dapat digunakan sebagai salah satu metode terbaik dalam evaluasi tingkat kesehatan yang ada di wilayah khususnya Pulau Jawa serta acuan pengambilan keputusan dalam menentukan kebijakan bagi instansi terkait.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Algoritma *K-means Binary Search Centroid (KBSC)* dan *Dynamic K-means (DK)* menghasilkan rata-rata kualitas *cluster* yang lebih baik dibanding algoritma *kmeans* tradisional dalam evaluasi metode *Elbow*, *Davies Bound Index (DBI)* dan *Calinski Harabasz Index (CHI)*. Berdasarkan pengujian diperoleh 5 jumlah *cluster* di wilayah sebaran, di antaranya *cluster* 1 dengan kualitas kesehatan sangat tinggi adalah 11 kabupaten/kota, *cluster* 2 dengan kualitas kesehatan sangat rendah adalah 24 kabupaten/kota, dan *cluster* 3 dengan kualitas kesehatan rendah adalah 28 kabupaten/kota. dan kualitas kesehatan 45 kabupaten/kota untuk *cluster* 4 dengan kualitas kesehatan cukup serta *cluster* 5 dengan kualitas kesehatan tinggi adalah 11 kabupaten/kota. Secara keseluruhan optimasi algoritma *K-means Dynamic Binary Search Centroid* menghasilkan rata-rata kualitas *cluster* yang lebih baik dibanding algoritma *K-means* tradisional. Hal ini ditunjukkan dengan evaluasi metode *Elbow* menghasilkan jumlah *cluster* terbaik sebanyak 5, perbedaan nilai *DBI* sebesar 0.6825546 dimana metode yang diusulkan memiliki nilai lebih baik dibanding metode *K-means* tradisional. Serta jumlah dan variasi data mempengaruhi kinerja algoritma dan besar nilai evaluasi algoritma.

5.2 Saran

Peneliti mengakui bahwa penelitian ini masih perlu untuk dikembangkan, adapun saran untuk penelitian selanjutnya sebagai berikut

1. Data yang diuji lebih beragam lagi dan bervariasi, sehingga dapat mengetahui kinerja gabungan algoritma *K-means Dynamic Binary Search Centroid*.
2. Dalam penelitian ini tidak mempertimbangan teknik pemrograman, diperlukan teknik pemrograman yang lebih baik, sehingga proses bisa lebih cepat dan efisien.
3. Untuk penelitian selanjutnya diharapkan untuk menggabungkan algoritma yang berbeda lainnya dalam *clustering* suatu data untuk mendapatkan hasil yang lebih baik.

DAFTAR PUSTAKA

- Akbari, G., & Kerlooza, Y. (2018). Peningkatan Hasil *Cluster* Menggunakan Algoritma *Dynamic K-means* dan *K-means Binary Search Centroid*. *Jurnal Tata Kelola dan Kerangka Kerja Teknologi Informasi*, 4(2), 25-33.
- Ardan Febrianto, H. (2017). *Peningkatan Hasil Clustering Menggunakan Algoritma Enhanced K-means Dan Dynamic Cluster (Studi Kasus: Pengelompokan Metode Pelatihan Sepak Bola)* (Doctoral dissertation, Universitas Komputer Indonesia).
- Ariasa, K., Gunadi, I. G. A., & Candiasa, I. M. (2020). Optimasi Algoritma *Cluster* Dinamis Pada *K-means* Dalam Pengelompokan Kinerja Akademik Mahasiswa (Studi Kasus: Universitas Pendidikan Ganesha). *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, 9(2), 181-192.
- Baarsch, J., & Celebi, M. E. (2012). Investigation of *Internal Validity Measures* for *K-means Clustering*. *International Multiconference of Engineers and Computer scientists 1* (hal. 14-16). LA: Louisiana Board of Regents.
- Badan Pusat Statistik Pulau Jawa, “Statistik Kesehatan Pulau Jawa 2020”, Available:
<https://jatim.bps.go.id/publication/2021/08/05/a70cbc1ca224552d5e0f5000/statistik-kesehatan-provinsi-jawa-timur-2020.html>. [Online]. [Accessed: 05-August-2022].
- Bansal, A., Sharma, M., & Goel, S. (2017). Improved k-mean *clustering* algorithm for prediction analysis using classification technique in data mining. *International Journal of Computer Applications*, 157(6), 0975-8887.
- Dinas Kesehatan Pulau Jawa, “Profil Kesehatan Pulau Jawa 2020”, Available:
<https://dinkes.jatimprov.go.id/userfile/dokumen/Profil%20kesehatan%202020.pdf>. [Online]. [Accessed: 05-August-2022].
- Harani, N. H., Prianto, C., & Nugraha, F. A. (2020). Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma *K-means* Berbasis Python. *Jurnal Manajemen Informatika (JAMIKA)*, 10(2), 133-146.
- Hardyanto, C. (2017). *Peningkatan Kualitas Hasil Clustering Menggunakan Algoritma HierarCHical Agglomerative Clustering K-means-Particle Swarm Optimization (Studi Kasus: Segmentasi Pasar Film)* (Doctoral dissertation, Universitas Komputer Indonesia).
- Hatamlou, A. (2012). In search of optimal *centroids* on data *clustering* using a binary search algorithm. *Pattern Recognition Letters*, 33(13), 1756-1760.
- Hossain, M. Z., Akhtar, M. N., Ahmad, R. B., & Rahman, M. (2019). A *Dynamic K-means clustering* for data mining. *Indonesian Journal of Electrical engineering and computer science*, 13(2), 521-526.
- Irani, J., Pise, N., & Phatak, M. (2016). *Clustering Techniques and The Similarity Measures Used in Clustering : A Survey*. *International Journal of Computer*

Applications, 9-14.

- Jing, J., Ke, S., Li, T., & Wang, T. (2021). Energy method of geophysical logging lithology based on *K-means Dynamic clustering* analysis. *Environmental Technology & Innovation*, 23, 101534.
- K. A. Seputra, I. Made Sudarma, and L. Jasa, "The Optimization of the *Dynamic K-means Clustering* Algorithm with the *Cluster Initialization* in Grouping Travelers Perception to the Beach Tourist Destinations in Bali, Indonesia," *International Journal of Research in IT*, vol. 07, no. 04, pp. 1–7, 2017, [Online]. Available: <http://indusedu.org>.
- Kamila, C. (2021). Systematic Literature Review: Penggunaan Algoritma *K-means* untuk *Clustering* di Indonesia dalam Bidang Pendidikan. *INTECH*, 2(1), 19-24.
- Kumar, Y., & Sahoo, G. (2014). A new initialization method to originate initial *cluster* centers for *K-means* algorithm. *International Journal of Advanced Science and Technology*, 62, 43-54.
- Nurdiana, N., Nilogiri, A., & Rahman, M. (2022). Penerapan Algoritma Fuzzy C-Means dan Metode *Elbow* untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Indeks Demokrasi Indonesia. *Jurnal Smart Teknologi*.
- Oti, E. U., Olusola, M. O., Esemokumo, P. A., & Enogwe, S. U. (2021). *K-Means Clustering* Method That Generates Initial *Cluster* Centers And Also Updates *Cluster Centroids*. *Journal Of Basic Physical Research Vol*, 10(2).
- Prasetyo, E. (2014). Data mining mengolah data menjadi informasi menggunakan matlab.
- Santoso, H., Magdalena, H., & Wardhana, H. (2022). Aplikasi *Dynamic Cluster* pada *K-means* Berbasis Web untuk Klasifikasi Data Industri Rumahan. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 21(3), 541-554.
- Satria, C., & Anggrawan, A. (2021). Aplikasi *K-means* berbasis Web untuk Klasifikasi Kelas Unggulan. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 21(1), 111-124.
- Satriawan, M. A. (2021). Segmentasi Pelanggan Telkomsel Menggunakan Metode *Clustering* Dengan Rfm Model Dan Algoritma *K-means*. *eProceedings of Engineering* 8.2.
- Shafeeq, A., & Hareesha, K. S. (2012). *Dynamic clustering* of data with modified *K-means* algorithm. In *Proceedings of the 2012 conference on information and computer networks* (pp. 221-225).
- Wahyudin, I., Djatna, T., & Kusuma, W. A. (2016). *Cluster* analysis for SME risk analysis documents based on Pillar *K-means*. *Telkomnika*, 14(2), 674.
- Widiarina, & Romi. (2015). Algoritma *Cluster* Dinamik untuk Optimasi *Cluster* pada Algoritma *K-means* dalam Pemetaan Nasabah Potensial. *Journal of Intelligent Systems*, 1.

Wijaya, Y. A., Kurniady, D. A., Setyanto, E., Tarihoran, W. S., Rusmana, D., & Rahim, R. (2021). Davies bouldin index algorithm for optimizing *clustering* case studies mapping school facilities. TEM J, 10(3), 1099-1103.

LAMPIRAN

| No. | Kabupaten/Kota Se Pulau Jawa |
|-----|---------------------------------|
| 1 | Kabupaten Pacitan |
| 2 | Kabupaten Ponorogo |
| 3 | Kabupaten Trenggalek |
| 4 | Kabupaten Tulungagung |
| 5 | Kabupaten Blitar |
| 6 | Kabupaten Kediri |
| 7 | Kabupaten Malang |
| 8 | Kabupaten Lumajang |
| 9 | Kabupaten Jember |
| 10 | Kabupaten Banyuwangi |
| 11 | Kabupaten Bondowoso |
| 12 | Kabupaten Situbondo |
| 13 | Kabupaten Probolinggo |
| 14 | Kabupaten Pasuruan |
| 15 | Kabupaten Sidoarjo |
| 16 | Kabupaten Mojokerto |
| 17 | Kabupaten Jombang |
| 18 | Kabupaten Nganjuk |
| 19 | Kabupaten Madiun |
| 20 | Kabupaten Magetan |
| 21 | Kabupaten Ngawi |
| 22 | Kabupaten Bojonegoro |
| 23 | Kabupaten Tuban |
| 24 | Kabupaten Lamongan |
| 25 | Kabupaten Gresik |
| 26 | Kabupaten Bangkalan |
| 27 | Kabupaten Sampang |
| 28 | Kabupaten Pamekasan |
| 29 | Kabupaten Sumenep |
| 30 | Kota Kediri |
| 31 | Kota Blitar |
| 32 | Kota Malang |
| 33 | Kota Probolinggo |
| 34 | Kota Pasuruan |
| 35 | Kota Mojokerto |
| 36 | Kota Madiun |

| No. | Kabupaten/Kota Se Pulau Jawa |
|-----|---------------------------------|
| 37 | Kota Surabaya |
| 38 | Kota Batu |
| 39 | Kabupaten Cilacap |
| 40 | Kabupaten Banyumas |
| 41 | Kabupaten Purbalingga |
| 42 | Kabupaten Banjarnegara |
| 43 | Kabupaten Kebumen |
| 44 | Kabupaten Purworejo |
| 45 | Kabupaten Wonosobo |
| 46 | Kabupaten Magelang |
| 47 | Kabupaten Boyolali |
| 48 | Kabupaten Klaten |
| 49 | Kabupaten Sukoharjo |
| 50 | Kabupaten Wonogiri |
| 51 | Kabupaten Karanganyar |
| 52 | Kabupaten Sragen |
| 53 | Kabupaten Grobogan |
| 54 | Kabupaten Blora |
| 55 | Kabupaten Rembang |
| 56 | Kabupaten Pati |
| 57 | Kabupaten Kudus |
| 58 | Kabupaten Jepara |
| 59 | Kabupaten Demak |
| 60 | Kabupaten Semarang |
| 61 | Kabupaten Temanggung |
| 62 | Kabupaten Kendal |
| 63 | Kabupaten Batang |
| 64 | Kabupaten Pekalongan |
| 65 | Kabupaten Pemalang |
| 66 | Kabupaten Tegal |
| 67 | Kabupaten Brebes |
| 68 | Kota Magelang |
| 69 | Kota Surakarta |
| 70 | Kota Salatiga |
| 71 | Kota Semarang |
| 72 | Kota Pekalongan |

| No. | Kabupaten/Kota Se Pulau Jawa |
|------------|---|
| 73 | Kota Tegal |
| 74 | Kabupaten Bogor |
| 75 | Kabupaten Sukabumi |
| 76 | Kabupaten Cianjur |
| 77 | Kabupaten Bandung |
| 78 | Kabupaten Garut |
| 79 | Kabupaten Tasikmalaya |
| 80 | Kabupaten Ciamis |
| 81 | Kabupaten Kuningan |
| 82 | Kabupaten Cirebon |
| 83 | Kabupaten Majalengka |
| 84 | Kabupaten Sumedang |
| 85 | Kabupaten Indramayu |
| 86 | Kabupaten Subang |
| 87 | Kabupaten Purwakarta |
| 88 | Kabupaten Karawang |
| 89 | Kabupaten Bekasi |
| 90 | Kabupaten Bandung Barat |
| 91 | Kabupaten Pangandaran |
| 92 | Kota Bogor |
| 93 | Kota Sukabumi |
| 94 | Kota Bandung |
| 95 | Kota Cirebon |
| 96 | Kota Bekasi |

| No. | Kabupaten/Kota Se Pulau Jawa |
|------------|---|
| 97 | Kota Depok |
| 98 | Kota Cimahi |
| 99 | Kota Tasikmalaya |
| 100 | Kota Banjar |
| 101 | Kabupaten Lebak |
| 102 | Kabupaten Pandeglang |
| 103 | Kabupaten Serang |
| 104 | Kabupaten Tangerang |
| 105 | Kota Tangerang |
| 106 | Kota Cilegon |
| 107 | Kota Serang |
| 108 | Kota Tangerang Selatan |
| 109 | Kabupaten Kulon Progo |
| 110 | Kabupaten Bantul |
| 111 | Kabupaten Gunung Kidul |
| 112 | Kabupaten Sleman |
| 113 | Kota Yogyakarta |
| 114 | Kota Jakarta Pusat |
| 115 | Kota Jakarta Utara |
| 116 | Kota Jakarta Barat |
| 117 | Kota Jakarta Selatan |
| 118 | Kota Jakarta Timur |
| 119 | Kepulauan Seribu |