Jurnal Smart Teknologi
Vol. 1, No. 1, Maret 2021, Halaman 100 – 102
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Algoritma *Fuzzy C-Means* dan Metode *Davies Bouldin Index (DBI)* untuk Mengolompokkan Kabupaten/Kota di Jawa Timur Berdasarkan Sanitasi Layak, Air Minum Layak, dan Rumah Layak Huni

Fuzzy C-Means Algorithm and Davies Bouldin Index (DBI) Method for Grouping Regencies/Cities in East Java Based on Adequate Sanitation, Adequate Drinking Water, and Liveable Houses

Ferdiyan Teguh Permadi[1)], Agung Nilogiri[2)], Ulya Anisatur Rosyidah[3)]
[1]Mahasiswa Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Email: ferdiyanteguh@gmail.com
[2,3]Dosen Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Email: agungnilogiri@unmuhjember.ac.id, ulyaanisatur@unmuhjember.ac.id

Abstrak

Sanitasi layak, air minum layak, dan rumah layak huni merupakan sektor penting pada kesehatan lingkungan untuk kebutuhan dasar manusia. Upaya pemenuhan kebutuhan tersebut sejalan dengan Agenda Global Tujuan Pembangunan Berkelanjutan 2030, yaitu SDGs (*Sustainable Development Goals*). Demi tercapainya tujuan tersebut, pemerintah pusat maupun daerah membuat berbagai macam program dan target. Dari permasalahan tersebut, diperlukan suatu kajian yang dapat memberikan referensi atau alternatif dalam menentukan target daerah yang tepat, salah satunya dengan algoritma *Fuzzy C-Means Clustering* dan metode *Davies Bouldin Index* (DBI). Data yang digunakan adalah persentase rumah tangga memiliki akses terhadap sanitasi layak, sumber air minum layak, dan rumah layak huni di 38 kabupaten/kota di Jawa Timur tahun 2019-2020. Dari serangkaian pengujian mulai dari 2 *cluster* sampai 10 *cluster*, dihasilkan *cluster* optimum berdasarkan nilai DBI terkecil berada pada 2 *cluster* yaitu dengan nilai DBI 0,740173.

Kata Kunci: *SDGs, clustering, fuzzy c-means, DBI.*

Abstract

Adequate sanitation, proper drinking water, and decent housing are important sectors in environmental health for basic human needs. Efforts to fulfill these needs are in line with the Global Agenda for the 2030 Sustainable Development Goals, namely the SDGs (Sustainable Development Goals). In order to achieve this goal, the central and local governments have made various programs and targets. From these problems, we need a study that can provide references or alternatives in determining the right target area, one of which is the Fuzzy C-Means Clustering algorithm and the Davies Bouldin Index (DBI) method. The data used is the percentage of households that have access to proper sanitation, adequate drinking water sources, and livable houses in 38 districts/cities in East Java in 2019-2020. From a series of tests ranging from 2 clusters to 10 clusters, the optimum cluster based on the smallest DBI value is in 2 clusters with a DBI value of 0.740173.

Keywords: *SDGs, clustering, fuzzy c-means, DBI.*

Jurnal Smart Teknologi
Vol. 1, No. 1, Maret 2021, Halaman 100 – 102
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

1. PENDAHULUAN

Sanitasi layak, air minum layak, dan rumah layak huni merupakan bagian dari kebutuhan dasar manusia. Perumahan dan permukiman yang layak sama halnya dengan pendidikan dan kesehatan yang pemenuhannya dijamin dalam Undang-Undang Dasar 1945 pasal 28(h). Upaya pemenuhan tersebut semakin ditingkatkan setiap tahunnya. Komitmen terhadap pemenuhan kebutuhan dasar tersebut juga sejalan dengan Agenda Global Tujuan Pembangunan Berkelanjutan 2030 (*Sustainable Development Goals*) atau yang dikenal dengan SDGs. Tujuan utamanya pada sektor lingkungan hidup, yaitu akses air minum, akses sanitasi, dan akses rumah layak huni dengan memastikan masyarakat mencapai akses universal air bersih dan sanitasi (Lubis, dkk. 2019).

Wujud komitmen Pemerintah Indonesia kemudian ditetapkan dalam Peraturan Presiden Republik Indonesia Nomor 59 Tahun 2017 tentang Pelaksanaan Pencapaian Tujuan Pembangunan Berkelanjutan yang mengamanatkan perlunya sinergi SDGs dengan rencana pembangunan jangka menengah nasional dan daerah. Tahun 2024, Pemerintah Indonesia menargetkan seluruh masyarakat telah memiliki akses air minum layak dan 90 persen imasyarakat mendapatkan akses sanitasi layak, termasuk di dalamnya 20 persen akses sanitasi layak. Demi tercapainya tujuan tersebut pemerintah pusat maupun daerah membuat berbagai macam program dengan target adalah wilayah dimana masyarakatnya belum memiliki akses sanitasi dan air minum layak (Pangeswari, dkk. 2021).

Dari permasalahan tersebut, diperlukan suatu kajian yang dapat memberikan referensi atau alternatif dalam menentukan target daerah yang tepat dengan berbagai program tersebut, salah satunya dapat gunakan algoritma *clustering* pada *data mining* untuk mengelompokkan wilayah atau daerah. Dari hasil *clustering*, akan diketahui perbedaan karakteristik antar *cluster* dengan cara mencari nilai terendah dan tertinggi dari masing-masing *cluster*. Salah satu algoritma *clustering* yang dapat digunakan adalah algoritma *Fuzzy C-Means*.

Menurut Nugraha & Riyandari (2020), *Fuzzy C-Means* memiliki kelebihan dimana perhitungan dan alur algoritmanya sederhana dan tidak membutuhkan komputasi yang banyak sehingga akan mudah untuk diterapkan dengan mempertimbangkan hasil yang bagus. Selain itu, Siti Aminah, dkk (2019) menjelaskan bahwa metode *Fuzzy C-Means* dipilih kerena berdasarkan penelitian sebelumnya yang membahas mengenai perbandingan antara metode *K-Means* dan *Fuzzy C-Means* dalam melakukan pengklusteran, metode *Fuzzy C-Means* lebih unggul karena *cluster* yang dihasilkan lebih mendekatai ketepatan (*valid*) dan kualitas *cluster* yang lebih baik.

Selain algoritma *clustering*, dapat juga digunakan metode pencarian *cluster* optimum untuk mengetahui pengelompokkan cluster yang terbaik dari beberapa *cluster* pada skenario uji. Salah satu metode yang dapat digunakan adalah metode *Davies Bouldin Index* (DBI). Pengujian metode DBI ini dengan melakukan beberapa perhitungan pada beberapa *cluster* yang sudah ditentukan, kemudian hasil nilai DBI beberapa *cluster* yang didapat, nilai DBI yang terkecil adalah yang paling optimum.

Berdasarkan pemaparan uraian sebelumnya, maka penulis akan melakukan penelitian dengan judul "Implementasi Algoritma *Fuzzy C-Means* dan Metode *Davies Bouldin Index* (DBI) untuk mengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Sanitasi Layak, Air Minum Layak, dan Rumah Layak Huni" menggunakan data persentase rumah tangga memiliki akses terhadap sanitasi layak, sumber air minum layak, dan rumah layak huni di 38 kabupaten/kota tahun 2019-2020. Data didapat dari Badan Pusat Statistika (BPS) Jawa Timur dengan judul "Statistik Perumahan dan Pemukiman Provinsi Jawa Timur, 2020". Algoritma yang akan digunakan adalah algoritma clustering *Fuzzy C-Means* dan metode *Davies Bouldin Index*

(DBI) untuk menentukan cluster optimum yang akan diuji dari 2-10 *cluster*.

2. TINJAUAN PUSTAKA

A. Sanitasi Layak

Sanitasi dikatakan layak apabila jenis kloset yang digunakan adalah leher angsa, tempat pembuangan akhir tinja menggunakan tangki/SPAL, dan fasilitas tempat buang air besar yang digunakan pada rumah tangga sendiri atau bersama rumah tangga lain bukan di MCK umum. Di dalam penghitungan rumah layak huni salah satu syarat yang harus dipenuhi adalah adanya akses terhadap sanitasi layak. Semakin baik akses terhadap sanitasi layak berarti juga akan mempengaruhi dari angka pada rumah layak huni (Pangeswari, dkk. 2021).

B. Air Minum Layak

Sumber air minum layak dan tidak layak adalah rincian dari sumber air minum rumah tangga. Rumah tangga dikatakan memiliki sumber air minum layak apabila pada rumah tangga tersebut memiliki akses air minum (*access to improved water*) yaitu sumber air minum utama yang digunakan adalah leding, air terlindungi, dan air hujan. Air terlindungi mencakup sumur bor/pompa, sumur terlindung dan mata air terlindung. Bagi rumah tangga yang menggunakan sumber air minum berupa air kemasan, maka rumah tangga dikategorikan memiliki akses air minum layak jika sumber air untuk mandi/cuci berasal dari leding, iumur bor/pompa, sumur terlindung, mata air terlindung, dan air hujan (Pangeswari, dkk. 2021).

C. Rumah Layak Huni

Kecukupan minimum luas bangunan dan keselamatan bangunan serta kesehatan penghuninya merupakan persyaratan untuk memenuhi kriteria rumah layak huni. Berdasarkan Permenpupr RI No. 29/PRT/M/2018 tentang Rumah Layak Huni harus memenuhi beberapa kriteria meliputi kecukupan luas minimum sebesar 7,2 m2 – 12 m2 per penduduk, memiliki keselamatan bangunan,dan dari aspek kesehatan harus memiliki pencahayaan, sirkulasi udara yang baik, adanya sanitasi layak juga tersedianya air bersih. Di dalam UU No. 01 tahun 2011 tentang Perumahan dan Kawasan Permukiman, pemerintah turut menjamin terwujudnya rumah layak huni dan terjangkau dalam lingkungan yang sehat, aman, serasi, teratur, terencana, terpadu, dan berkelanjutan (Pangeswari, dkk. 2021).

D. *Data Mining*

Data mining adalah proses untuk menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah data dalam jumlah besar yang disimpan di dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika (Larose & Larose, 2014). *Data mining* menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui pada data yang telah dikumpulkan (Han, dkk, 2012).

E. *Clustering*

Clustering adalah proses mengelompokkan atau penggolongan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/*cluster*. *Clustering* dalam *data mining* berguna untuk menemukan pola distribusi di dalam sebuah *dataset* yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek objek data, sedangkan objek-objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi (Hermawati, 2013).

Clustering mengacu pada pengelompokan seperti *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan dari *record* yang memiliki kemiripan satu sama lain, dan berbeda dengan record di *cluster* lain. *Clustering*

Jurnal Smart Teknologi
Vol. 1, No. 1, Maret 2021, Halaman 100 – 102
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

mencoba untuk membagi seluruh kumpulan data menjadi kelompok-kelompok yang relatif memiliki kemiripan, di mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal (Larose & Larose, 2014).

F. *Fuzzy C-Means*

Fuzzy C-Means (FCM) adalah suatu tenik pengklusteran data yang keberadaan setiap titik data dalam suatu *cluster* di tentukan oleh derajat keanggotaan. Teknik ini pertama di perkenalkan oleh Jim Bezdek pada tahu 1981. Output dari *FCM* adalah deretan pusat *cluster* dan beberapa derajat keanggotaan untuk tiap-tiap titik data (Aminah, dkk. 2019).

Menurut Nugraha dan Riyandari (2020), langkah-langkah untuk algoritma *Fuzzy C-Means* adalah sebagai berikut:

1. Masukkan data yang akan di *cluster X*, berupa matriks berukuran n x m (n=jumlah sampel data, m = atribut setiap data). Xij= data sampel ke-i (i=1,2,.....,n),atribut ke-j (j = 1,2.....,m).
2. Menentukan :
   - Jumlah *cluster (c)*
   - Pangkat (w)
   - Maksimum iterasi (MaxIter)
   - Error terkecil yang diharapkan (ξ)
   - Fungsi objektif awal (P_0 = 0)
   - Iterasi awal (t=1)
3. Bangkitkan bilangan *random* μ_{ik}, i = 1, 2, 3, …,n; k = 1, 2, 3, …,n; sebagai elemen dari matriks partisi awal U (derajat keanggotaan pada *cluster*). μ_{ik} merupakan derajat keanggotaan yang merujuk pada seberapa besar kemungkinan beberapa data akan menjadi anggota kedalam suatu *cluster*. Posisi & nilai matriks dihasilkan secara acak. Dimana nilai keanggotaan terletak dalam interval 0 hingga 1. Posisi awal matriks partisi *U* masih belum akurat begitu juga pusat *cluster*nya. Oleh karena itu, kecenderungan data yang masuk ke dalam *cluster* juga tidak akurat.
   Hitung setiap kolom atribut:
   $$Qj = \sum_{k=1}^{c} \mu_{ik} \qquad (1)$$
   Qj merupakan jumlah nilai menggunakan derajat keanggotaan per kolom = 1
   Dengan j = 1,2,.....,m.
4. Menghitung pusat *cluster* ke-k: V_{kj} dengan k=1, 2, 3, …,c; dan j=1, 2, 3, …,m
   $$V_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^{w} \times X_{ij}}{\sum_{i-1}^{n}(\mu_{ik})^{w}} \qquad (2)$$
   V_{kj} = Pusat *cluster*
   i = iterasi
   μ_{ik} = perubahan matrix partisi
   X_{ij} = atribut
5. Menghitung fungsi objektif pada iterasi ke-t. Fungsi objektif dipakai sebagai kondisi iterasi untuk mendapatkan pusat *cluster* yang tepat. Sehingga diperoleh kesamaan data untuk masuk ke *cluster* mana pada langkah akhir.
   $$P_t = \sum_{i=1}^{n} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{m} (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^{w} \right) \qquad (3)$$
   P_t = Fungsi objektif
   $\sum_{i=1}^{n}$ = Jumlah data yang di*cluster*
   $\sum_{i=1}^{n}$ = jumlah perhitungan *cluster* awal
6. Menghitung perubahan matrix partisi:
   $$\mu_{ik} = \frac{\left[\sum_{j-1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]^{\frac{-1}{w-1}}}{\sum_{k=1}^{C}\left[\sum_{j-1}^{m}\left(X_{ij} - V_{kj}\right)^2\right]^{\frac{-1}{w-1}}} \qquad (4)$$
   Dengan i=1,2,…,n; dan k=1,2,…c.
7. Memeriksa kondisi berhenti:
   - Jika (|Pt – Pt-1|< ξ) atau (t > MaxIter) maka berhenti.
   - Jika tidak : t = t + 1, mengulang ke langkah empat.

G. *Davies Bouldin Index*

David L. Davies dan Donald W. Bouldin adalah yang pertama kali mengusulkan sebuah metode evaluasi *cluster* yang dinamai oleh mereka "*Davies-Bouldin Index* (DBI)". Evaluasi

Jurnal Smart Teknologi
Vol. 1, No. 1, Maret 2021, Halaman 100 – 102
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

menggunakan DBI ini memiliki skema evaluasi *cluster* internal, dimana baik atau tidaknya hasil *cluster* dilihat dari kuantitas dan kedekatan antar data hasil *cluster*. Semakin kecil nilai DBI menunjukan skema *cluster* yang paling optimal (Jollyta, dkk. 2019).

Langkah-langkah perhitungan *Davies-Bouldin Index* (DBI) adalah sebagai berikut (Rohmah, 2019):

1. Menghitung nilai *Sum of Square Within-Cluster* (SSW).
   Sum of Square Within-Cluster (SSW) adalah salah satu cara untuk mengetahui kohesi dalam sebuah *cluster* ke-i. Kohesi diartikan sebagai jumlah dari kedekatan atau kemiripan data terhadap titik pusat *cluster* dari sebuah *cluster* yang diikuti. Persamaan untuk menentukan nilai *Sum of Square Within Cluster* (SSW) adalah sebagai berikut:
   $$SSW_i = \frac{1}{m_i}\sum_{j=i}^{m_i} d(x_j, c_i)$$
   Dimana,
   mi = jumlah data dalam *cluster* ke-i
   ci = *centroid cluster* ke-i
   d(*xj*,*ci*) = jarak setiap data ke *centroid i* yang dihitung menggunakan jarak *euclidean*.
2. Menghitung nilai *Sum of Square Between-Cluster* (SSB).
   Sum of Square Between-Cluster (SSB) dihitung untuk mengetahui separasi atau jarak antar *cluster*. Berikut ini adalah persamaan untuk menentukan nilai *Sum of Square Between-Cluster* (SSB):
   $SSB_i = (c\ , cj)$
   Dimana, $d(xi\ ,\ xj\ )$ merupakan jarak antara data ke *i* dengan data ke *j* di *cluster* lain.
3. Melakukan perhitungan Rasio dari nilai SSW dan nilai SSB.
   Cluster yang baik adalah *cluster* yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Perhitungan rasio (Ri,j) ini bertujuanuntuk mengetahui nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j untuk menghitung nilai rasio yang dimiliki oleh masing-masing *cluster*. Indeks i dan j merupakan mereprsentasikan jumlah *cluster*, dimana jika terdapat 3 *cluster* maka terdapat indeks sebanyak 3 yaitu i, j, dan k. Persamaan untuk menentukan nilai Rasio yaitu:
   $$R_{ij} = \frac{SSW_i + SSW_i}{SSB_{ij}}$$
   Dimana,
   $SSWi$ = *Sum of Square Within Cluster* pada *centroid* i
   $SSBi,j$ = *Sum of Square Between Cluster* data ke i dengan j pada *cluster* yang berbeda
4. Menghitung nilai *Davies-Bouldin Index* (DBI) dari nilai Rasio yang diperoleh sebelumnya dengan menggunakan persamaan berikut:
   $$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i\neq j}(R_{ij})$$
   Dimana, k adalah jumlah *cluster* yang digunakan. Semakin kecil nilai *Davies Bouldin Index* (DBI) yang diperoleh (non negatif >= 0) maka *cluster* tersebut semakin baik, namun jika *cluster* yang diperoleh dari pengelompokan menggunakan algoritma *clustering.*

H. *RStudio*

RStudio adalah *tools software* yang digunakan untuk menjalankan bahasa pemrograman R. Bahasa Pemrograman R ini merupakan salah satu pemrograman yang digunakan untuk analisis data dan grafik yang dikembangkan oleh Rick Becker, John Chambers, dan Allan Wilks dari AT&T Bell Laboratories (1976). *Tools* RStudio ini tersedia secara gratis atau *open source* yang dapat diunduh, dan juga memiliki beberapa keunggulan yaitu cepat, pengembangan metode dengan menggunakan *package*, memiliki grafik yang baik, dan unggul dalam menganalisis dan simulasi data (Asdi, 2017).

Jurnal Smart Teknologi
Vol. 1, No. 1, Maret 2021, Halaman 100 – 102
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

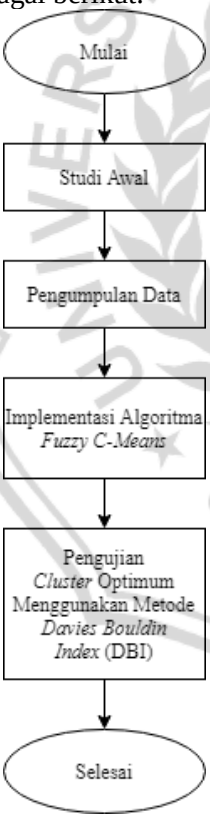
3. METODOLOGI PENELITIAN

A. Tahapan Penelitian

Metodologi yang dilakukan dalam penlitian ini melalui beberapa tahapan yaitu studi awal, pengumpulan data, proses *clustering* menggunakan algoritma *Fuzzy C-Means*, kemudian pengujian metode *Davies Bouldin Index* (DBI) untuk menentukan *cluster* optimum. Setelah melakukan pengujian metode *Davies Bouldin Index* (DBI), *cluster* yang optimum kemudian dilakukan *profiling cluster* optimum untuk mengetahui karakteristik data pada masing-masing *cluster* optimum.

B. Diagram Alur Penelitian

Dari tahapan penelitian yang sudah dijelaskan sebelumnya, maka diagram alur penelitian sebagai berikut:



Gambar 1. Tahapan Penelitian
Sumber: Hasil Penelitian (2022)

C. Studi Awal

Pada tahap awal penelitian ini, dimulai dengan mencari dan mempelajari masalah yang akan diteliti, lalu menentukan ruang lingkup dan latar belakang masalah, kemudian mempelajari beberapa literatur seperti jurnal, paper, artikel yang berkaitan dengan permasalahan & bagaimana mencari solusi berdasarkan masalah tersebut. Untuk mencapai tujuan yang ditentukan, maka penulis perlu mempelajari beberapa literatur yang dipakai, lalu literatur tersebut diseleksi untuk ditentukan menjadi literatur yang akan digunakan dalam penelitian.

D. Pengumpulan Data

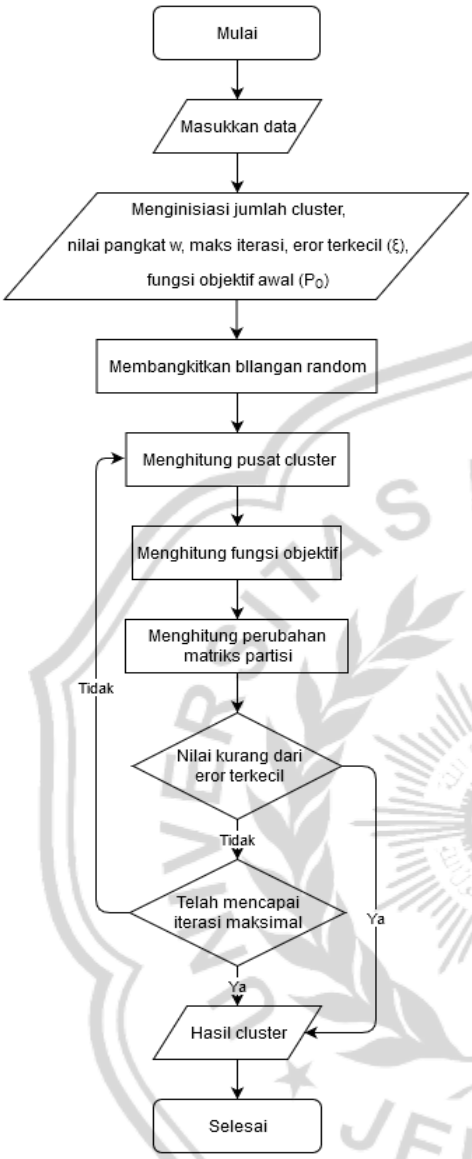
Tahap pengumpulan data ini menggunakan data sekunder yang diperoleh dari sumber yang sudah ada yaitu dari *website* Badan Pusat Statistik Provinsi Jawa Timur yang dapat diakses pada *link*: https://jatim.bps.go.id/Publikasi. Dari *link* tersebut didapatkan publikasi yang berjudul "Statistik Perumahan dan Permukiman Provinsi Jawa Timur 2020". Pada publikasi tersebut, terdapat beberapa data per kabupaten/kota di Jawa Timur tentang statistik perumahan dan pemukiman pada tahun 2019-2020.

Dataset pada penelitian ini digunakan sebanyak 38 kabupaten/kota di Jawa Timur yang terdiri dari 29 kabupaten dan 9 kota berdasarkan persentase rumah tangga memiliki akses terhadap sanitasi layak, sumber air minum layak, dan rumah layak huni tahun 2019-2020. Dimana data tersebut akan di *cluster* menggunakan algoritma *Fuzzy C-Means* dan menggunakan metode *Davies Bouldin Index* (DBI) sebagai metode optimasi *cluster* untuk mengetahui jumlah *cluster* optimum yang akan digunakan.

E. Alur Algoritma *Fuzzy C-Means*

Algoritma *Fuzzy C-Means* memiliki tahapan alur yang pada umumnya tahapan pada *flowchart* adalah sebagai berikut:

Jurnal Smart Teknologi
Vol. 1, No. 1, Maret 2021, Halaman 100 – 102
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST



Gambar 2. *Flowchart Fuzzy C-Means*
Sumber: Hasil Penelitian (2022)

F. Metode *Davies Bouldin Index* (DBI)

Metode DBI digunakan untuk menentukan jumlah *cluster* yang paling optimum. Sama pada bab sebelumnya, bab ini menggunakan 15 data sampel. Langkah-langkah perhitungan *Davies-Bouldin Index* (DBI):

1. Mengelompokkan data dari *cluster* yang di ikuti.
2. Menghitung nilai *Sum of Square Within-Cluster* (SSW).
3. Menghitung nilai *Sum of Square Between-Cluster* (SSB).
4. Melakukan perhitungan Rasio dari nilai SSW dan nilai SSB.
5. Menghitung nilai *Davies-Bouldin Index* (DBI).
6. Dihitung sampai DBI pada *cluster* terakhir, kemudian nilai DBI yang terkecil atau terendah adalah *cluster* yang paling optimum.

4. HASIL DAN PEMBAHASAN

A. Data Pengujian

Keseluruhan data yang didapat akan di *cluster* menggunakan algoritma *Fuzzy C-Means* kemudian diolah untuk mengetahui hasil *cluster* optimum dengan menggunakan metode *Davies Bouldin Index* (DBI). Data yang digunakan adalah persentase 3 parameter Kesehatan lingkungan yang terdiri dari rumah tangga memiliki akses terhadap sanitasi layak, sumber air minum layak, dan rumah layak huni tahun 2019-2020 pada 38 kabupaten/kota di Jawa Timur yang terdiri dari 29 kabupaten dan 9 kota.

B. *Fuzzy C-Means* pada RStudio

Data uji yang digunakan di *input* ke RStudio kemudian diolah menggunakan perintah *Fuzzy C-Means* dari 2 *cluster* sampai dengan 10 *cluster*. Hasil dari menjalankan perintah di RStudio akan muncul di bagian *console* dengan menampilkan beberapa *output* diantaranya adalah jumlah iterasi, pusat *cluster*, fungsi objektif, dan derajat keanggotaan setiap objek terhadap tiap *cluster*. Hasil *cluster* yang telah dijalankan dapat juga ditampilkan berupa *plot* dari setiap *cluster* yang terbentuk dengan simbol dan warna tertentu.

C. Hasil *Davies Bouldin Index* (DBI) pada RStudio

Setelah melakukan proses *cluster* menggunakan *Fuzzy C-Means*, kemudian dilakukan proses penentuan *cluster* optimum menggunakan metode *Davies Bouldin Index* (DBI) dari 2 *cluster* sampai dengan 10 *cluster*. Hasil perhitungan DBI berupa bilangan desimal yang dimana semakin kecil nilai DBI yang dihasilkan, maka *cluster* tersebut

semakin optimum. Berikut ini adalah hasil metode DBI pada RStudio:

Tabel 1. Hasil DBI

Cluster	DBI
2	**0,740173**
3	0,895274
4	1,096592
5	1,013756
6	0,961210
7	1,035917
8	0,958686
9	0,863929
10	0,796467
Nilai Terkecil (MIN)	**0,740173**

Sumber: Hasil Perhitungan (2022)

Pada hasil tabel 4.2 menunjukkan nilai DBI terkecil terdapat pada 2 *cluster* dengan nilai DBI yaitu 0,740173. Jadi, *cluster* optimum terdapat pada 2 *cluster*.

D. *Profiling Cluster* Optimum

Dari proses pencarian *cluster* optimum menggunakan *Davies Bouldin Index* (DBI) telah didapat *cluster* optimum terdapat pada 2 *cluster* yang terdiri dari *cluster* 1 dan *cluster* 2. Untuk mengetahui karakteristik dari masing-masing cluster optimum tersebut, perlu dilakukan *profiling cluster* dengan mencari nilai terendah dan tertinggi pada masing-masing *cluster* optimum.

Pada *cluster* 1 terdapat 12 anggota data yang terdiri dari 12 kabupaten dan 0 kota. Karakteristik *cluster* 1, pada sanitasi layak tahun 2019 dengan persentase 43,1 – 76,34 dan tahun 2020 dengan persentase 44,07 – 83,67, pada air minum layak tahun 2019 dengan persentase 77,72 – 96,28 dan tahun 2020 dengan persentasi 80,64 – 99,49, pada rumah layak huni tahun 2019 dengan persentase 38,71 - 61,72 dan pada tahun 2020 dengan persentase 36,49 – 69,59.

Pada *cluster* 2 terdapat 26 anggota data yang terdiri dari 17 kabupaten dan 9 kota. Karakteristik *cluster* 2, pada sanitasi layak tahun 2019 dengan persentase 69,59 – 98,06 dan tahun 2020 dengan persentase 77,44 – 98,71, pada air minum layak tahun 2019 dengan persentase 84,31 – 99,87 dan tahun 2020 dengan persentasi 91,15 – 100, pada rumah layak huni tahun 2019 dengan persentase 57,28 - 90,55 dan pada tahun 2020 dengan persentase 52,02 – 93,71.

Dari hasil karakteristik data rumah tangga memiliki akses terhadap sanitasi layak, sumber air minum layak, dan rumah layak huni tahun 2019-2020, *cluster* 1 memiliki persentase lebih rendah dibandingkan dengan persentase pada *cluster* 2.

5. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan uraian permasalahan dan pembahasan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

1. Penerapan metode *Davies Bouldin Index (DBI)* pada algoritma *Fuzzy C-Means* dari pengelompokan kabupaten/kota di Jawa Timur berdasarkan sanitasi layak, air minum layak, dan rumah layak huni menghasilkan cluster optimum pada 2 *cluster* dengan nilai DBI 0,740173. Pada 2 *cluster* ditetapkan sebagai *cluster* optimum karena nilai DBI yang diuji dari 2 *cluster* – 10 *cluster* yang paling terkecil terdapat pada 2 *cluster*.
2. Hasil pengelompokkan pada *cluster* 1 terdiri dari 12 kabupaten dan 0 kota yaitu Kab. Pacitan, Kab. Trenggalek, Kab. Jember, Kab. Bondowoso, Kab. Situbondo, Kab. Probolinggo, Kab. Pasuruan, Kab. Ngawi, Kab. Bangkalan, Kab. Sampang, Kab. Pamekasan, dan Kab. Sumenep. Sedangkan hasil pengelompokkan pada *cluster* 2 terdiri dari 17 kabupaten dan 9 kota yaitu Kab. Ponorogo, Kab. Tulungagung, Kab. Blitar, Kab. Kediri, Kab. Malang, Kab. Lumajang, Kab. Banyuwangi, Kab. Sidoarjo, Kab. Mojokerto, Kab. Jombang, Kab. Nganjuk, Kab. Madiun, Kab. Magetan, Kab. Bojonegoro, Kab. Tuban, Kab. Lamongan, Kab. Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota

Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu.

B. Saran

Saran dari penelitian ini yaitu:

1. Validitas *cluster* untuk mencari *cluster* terbaik, dapat menggunakan alternatif lain selain metode *Davies Bouldin Index* seperti metode *Silhouette, Gap Statistic, Elbow*, dll.
2. Sebagai pembanding hasil *cluster* pada RStudio digunakan perhitungan manual *cluster Fuzzy C-Means* di *Microsoft Excel*. Perhitungan manual di Microsoft Excel dapat menggunakan bilangan acak/*random* yang diambil dari RStudio untuk menentukan cluster awal agar menghasilkan pengelompokkan yang serupa. Jika bilangan *random* tidak diambil dari RStudio, maka hasil *cluster* dapat berbeda.

6. REFERENSI

Aminah, S., Irfan, A. P., & Zulkarnaim, N. 2019. *Clustering Wilayah berdasarkan Data Kesehatan Lingkungan menggunakan Fuzzy C-Means*. Majene: JCIS (Journal of Computer and Information System).

Asdi, Y. 2017. *Pengenalan Software R*. Padang: FMIPA Universitas Andalas.

Fahmi, M., Suprapto, Y., & Wirawan. 2016. *Segmentation and distribution of watershed using K-Modes Clustering Algorithm and Davies-Bouldin Index based on Geographic information System (GIS)*. Surabaya: International Seminar on Application for Technology of Information and Communication.

Han, J., Kamber, M., & Pei, J. 2012. *Data Mining Concepts And Techniques, Thrid Edition.* Waltham: Morgan Kaufmann Publishers.

Hermawati, F. A. 2013. *Data Mining*. Yogyakarta: Andi.

Jollyta, D., Efendi, S., Zarlis, M., & Mawengkang, H. 2019. *Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index*. Medan: Prosiding Seminar Nasional Riset Information Science (SENARIS).

Larose, D. T., & Larose, C. D. 2014. *Discovering Knowledge In Data: An Introduction To Data Mining: Second Edition.* Canada: John Wiley & Sons, Inc.

Lubis, I., Aisyah, N., & Mardikanto, A. K. 2019. *Pedoman Pengukuran Capaian Pembangunan Perumahan dan Permukiman Berbasis Hasil (Outcome)*. Jakarta: Kementerian Perencanaan Pembangunan Nasional.

Mayasari, T. R. 2019. *Clustering Akses Air Bersih dan Sanitasi Layak Kabupaten/Kota di Provinsi Lampung*. Lampung: Seminar Nasional Official Statistics 2019.

Muningsih, E., Maryani, I., & Handayani, V. R. 2021. *Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa*. Jakarta: Evolusi: Jurnal Sains dan Manajemen.

Ningsi, L. N., Poningsih, & Tambunan, H. S. 2021. *Implementasi Data Mining Kluster pada Rumah Tangga yang Memiliki Akses Hunian Layak Berdasarkan Provinsi*. Pematangsiantar: KLIK: Kajian Ilmiah Informatika dan Komputer.

Nugraha, G. S., & Riyandari, B. A. 2020. *Implementasi Fuzzy C-Means untuk Mengelompokan Daerah berdasarkan Indikator Kesehatan*. Mataram: Jurnal Teknologi Informasi, Vol.4, No.1.

Pangeswari, D. 2021. *Statistik Perumahan dan Pemukiman Provinsi Jawa Timur, 2020*. Surabaya: Badan Pusat Statistika Provinsi Jawa Timur.

Peng, L., Su, H., Peng, H., Qiao, X., Wu, P., & Wang, K. 2020. Wind Farm Clustering Optimization Method Using Gap Statistic. Beijing, China: IEEE, doi: 10.1109/ICPRE51194.2020.9233199.

Rohmah, R. L. 2019. *Zonasi Daerah Terdampak Bencana Angin Puting Beliung Menggunakan K-Means Clustering dengan Silhouette Coefficient, Davies Bouldin Index, dan Purity*. Surabaya: Universitas Islam Negeri Sunan Ampel Surabaya.