

## PEMETAAN PENDUDUK CALON PENERIMA BANTUAN RENOVASIRUMAH DESA PESANGKALAN MENGGUNAKAN ALGORITMA *CLUSTERING K-MEANS*

Asep Muhidin<sup>1)</sup>, Indarwista Baragiratri<sup>2)</sup>

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa  
asep.muhidin@pelitabangsa.ac.id

Disetujui, 25 Maret 2019

### Abstraksi

Rumah atau tempat tinggal adalah satu dari tiga kebutuhan dasar manusia. Di Indonesia banyak sekali rumah yang masuk dalam kategori Rumah Tidak Layak Huni. Pemerintah Indonesia memberikan bantuan Rehabilitasi Sosial Rumah Tidak Layak Huni (RS-RLTH) berupa uang untuk pembelian bahan bangunan guna pemugaran rumah. Di desa Pesangkalan, program bantuan renovasi rumah tergolong baru dan pemerintah desa belum dapat menentukan calon penerima bantuan yang tepat. Dibutuhkan teknik pemanfaatan data menjadi sebuah informasi baru atau disebut juga *data mining*. Metode yang cocok untuk memetakan calon penerima bantuan adalah *clustering k-means* karena metode ini dapat mengolah data tanpa diketahui label kelasnya. Penelitian ini akan menghasilkan tiga kelompok penduduk sesuai dengan pendapatan rata-rata dan status kepemilikan bangunan. Hasil tiga kelompok tersebut yaitu penduduk yang layak, kurang layak dan tidak layak menerima bantuan. Pengolahan data penduduk menggunakan algoritma *k-means* mendapatkan hasil *davies bouldinindex* sebesar 0.428 yang dinilai cukup baik karena semakin dekat hasil dengan angka nol, maka semakin baik pula hasil *clusternya*.

**Kata kunci:** Rumah tidak layak huni, *clustering, k-means, davies bouldin index*

### Abstract

*Home or residence is one of the three basic human needs. In Indonesia, a lot of houses in the category House No Livable. The Indonesian government to provide assistance Social Rehabilitation House Not Livable (RS-RLTH) in cash for the purchase of building materials for home renovations. In the village Pesangkalan, home renovation assistance program is relatively new and the village government can not determine the proper beneficiary selection. It takes data utilization techniques into a new information or called data mining. The method is suitable for mapping the prospective beneficiaries are k-means information or called data minin. The method is suitable for mapping the prospective beneficiaries are k-means clustering because this method can process data without a known class label. The study will include three groups of the population according to the average income and the status of ownership of the building. The results of the three groups, namely eligible population, less worthy and not worthy of receiving help. Using population data processing algorithms k-means receiving help. Using population data processing algorithms k-means get results Davies Bouldin index by 0428 that is considered good because the closer result with a zero, it means better the cluster results.*

**Keywords:** *The house is not habitable, clustering, k-means, Davies Bouldin index*

### 1. Pendahuluan

Tiga kebutuhan pokok manusia adalah sandang, pangan dan papan. Sandang berarti pakaian, pangan berarti makanan dan papan berarti tempat tinggal. Pentingnya tempat tinggal atau rumah layak huni untuk setiap orang dapat terlindungi, istirahat, dan berkumpul dengan keluarga dengan aman dan nyaman perlu diperhatikan. Di Indonesia masih banyak sekali rumah yang masuk dalam kategori tak layak huni. Pemerintah memberikan bantuan Rumah Tidak Layak Huni (RLTH) berupa uang untuk pembelian bahan bangunan guna pemugaran Rumah Tidak Layak Huni dimana rumah tersebut tidak memenuhi persyaratan keselamatan bangunan dan kecukupan minimal bangunan.

Penyaluran bantuan Rumah Tidak Layak Huni ditangani oleh pemerintah daerah yang disalurkan kepada kelurahan/desa yang masuk dalam kategori kelurahan/desa miskin dengan prioritas tinggi. Kelurahan/desa Pesangkalan termasuk salah satu desa yang masuk dalam kategori desa miskin di Banjarnegara, Jawa Tengah. Namun, dalam menentukan sasaran penanganan bantuan Rumah Tidak Layak

Huni, pemerintah kelurahan/desa Pesangkalan masih memiliki beberapa kendala. Kendala tersebut dikarenakan banyaknya data yang berubah sehingga memungkinkan proses penentuan penerimaan bantuan Rumah Tidak Layak Huni subjektif atau tidak tepat sasaran. Untuk mengatasi masalah tersebut, dibutuhkan metode penerapan data mining agar penyaluran bantuan ini tepat sasaran.

Program bantuan Rumah Tidak Layak Huni ini masih tergolong baru untuk kelurahan/desa Pesangkalan sehingga tidak terdapat label pada dataset. Oleh karena itu dibutuhkan metode klasterisasi (clustering) yang merupakan pengelompokan item data ke dalam sejumlah kecil grup sehingga masing-masing grup mempunyai persamaan yang esensial. Metode klasterisasi juga tidak menggunakan label dalam prosesnya dan label dapat diberikan ketika cluster sudah terbentuk.

Banyak metode klasterisasi yang diusulkan oleh para ahli, salah satunya adalah *k-Means* method atau metode *k-Means* yang termasuk dalam metode berbasis partisi, dimana metode ini bekerja dengan cara membagi data ke dalam sejumlah kelompok. Metode *k-Means* merupakan algoritma klasterisasi yang paling sering digunakan karena prosesnya sederhana untuk diimplementasikan dan dijalankan. Selain mampu melakukan klasterisasi dan mudah diadaptasi, operasi matematis dalam metode *k-Means* juga relatif lebih sederhana. Berdasarkan uraian diatas, peneliti tertarik untuk melakukan penelitian terkait membantu pemerintah kelurahan/desa Pesangkalan dalam menentukan sasaran penerima bantuan Rumah Tak Layak Huni secara tepat menggunakan algoritma *k-Means method*.

## 2. Tinjauan Studi

### 2.1 Metode *K-Means*

Suyanto (2017:262) mendefinisikan bahwa ide dasar algoritma *k-Means* sangat sederhana yaitu meminimalkan *Sum of Squared Error* (SSE) antar objek- objek data dengan sejumlah *k centroid*. Algoritma *k-Means* bekerja dengan empat langkah. Pertama, dari himpunan data yang akan diklasterisasi, dipilih sejumlah *k* objek secara acak sebagai *centroid* awal. Kedua, setiap objek yang bukan *centroid* dimasukkan ke klaster terdekat berdasarkan ukuran jarak tertentu. Ketiga, setiap *centroid* diperbaharui berdasarkan rata-rata dari objek yang ada di dalam setiap klaster. Keempat, langkah kedua dan ketiga tersebut diulang-ulang (diiterasi) hingga seluruh *centroid* stabil atau konvergen dalam arti semua *centroid* yang dihasilkan dalam iterasi saat ini sama dengan semua *centroid* yang dihasilkan pada iterasi sebelumnya.

Kompleksitas waktu *k-Means* adalah  $O(nkt)$ , di mana  $n$  adalah jumlah keseluruhan objek data,  $k$  adalah jumlah klaster, dan  $t$  adalah jumlah iterasi ( JHan dalam Suyanto, 2012:263).

Menurut Vlandari, (2017:54) *k-Means* merupakan algoritma *clustering* yang berulang-ulang. Algoritma *k-Means* menetapkan nilai-nilai *cluster* ( $k$ ) secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau biasa disebut dengan *centroid*, *means* atau “*means*”. Kemudian menghitung jarak setiap data yang ada terhadap masing-masing *centroid* menggunakan rumus *Euclidian* hingga ditemukan jarak paling dekat dari setiap data dengan *centroid*. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan *centroid*. Dan melakukan langkah tersebut hingga nilai *centroid* tidak berubah (stabil).

*K-means clustering* merupakan salah satu metode data *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil.

### 2.2 Langkah-langkah Proses *K-Means*

Menurut Santosa (2007), langkah-langkah melakukan clustering dengan metode *k-means* adalah sebagai berikut:

1. Pilih jumlah *cluster*  $k$ .
2. Inisialisasi  $k$  pusat kluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian jugakedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots (1)$$

dimana:

$D(i,j)$  = Jarak data ke  $i$  ke pusat cluster  $j$   
 $X_{kj}$  = Titik pusat ke  $j$  pada atribut ke  $k$

4. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan *median* dari *cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.

5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

Hasil dari operasi *clustering* yang terbentuk selanjutnya akan di evaluasi menggunakan *Davies bouldin index* yang dihitung dengan persamaan :

Dimana:

$$DBI = \frac{1}{K} \sum_{i=1}^k R_i$$

$$R_i = \max_{j=1 \dots k, i \neq j} R_{ij}$$

$$R_{ij} = \frac{\text{var}(C_i) + \text{var}(C_j)}{\|c_i - c_j\|}$$

$C_i$  = Cluster  $i$  dan  $c_i$  adalah *centroid* dari *cluster*  $i$

### 2.3 Kelebihan dan Kekurangan K-Means

Adapun keuntungan atau kelebihan dari metode *klustering k-Means* menurut Han, (2012:445) adalah:

- K-Means* dapat digunakan untuk mendekteksi outlier (nilai yang “jauh” dari manapun klaster).
- K-Means* juga disebut dengan segmentasi data di beberapa aplikasi karena pengelompokan mempartisi set data besar ke dalam grup sesuai dengan kemiripannya.
- K-Means* mempartisi sekumpulan objek data (atau pengamatan) ke dalam himpunan bagian, sehingga banyak digunakan dalam banyak aplikasi seperti interlijen bisnis, pengenalan pola gambar, pencarian web, biologi dan keamanan.

Selain itu, metode *klustering k-Means* juga mempunyai beberapa kekurangan. Berikut adalah kekurangan-kekurangan yang diuraikan oleh Suyanto, (2017:262) antara lain:

- K-Means* tidak dapat menjamin konvergen dalam optimum global.
- K-Means* sering terjebak pada *optimum local*, dimana *centroid* akhir yang dihasilkan tidak benar-benar menjadi pusat kluster yang sesungguhnya.
- Keluaran dari *k-Means* bergantung pada *centroid* awal yang ditentukan secara acak.

### 2.4 Rumah Layak Huni

Menurut Peraturan Menteri Negara Perumahan Rakyat Republik Indonesia Nomor 22/Permen/M/2008 tentang standar pelayanan bidang perumahan rakyat daerah provinsi dan daerah kabupaten mendefinisikan bahwa rumah layak huni adalah rumah yang memenuhi persyaratan keselamatan bangunan dan kecukupan minimum luas bangunan serta kesehatan penghuninya.

#### Rehabilitasi Sosial Rumah Tidak Layak Huni

Kementrian Sosial Republik Indonesia mengembangkan kebijakan sosial Penanggulangan Kemiskinan (P2K) melalui Rehabilitasi Sosial Ruma Tidak Layak Huni (RS-RUTILAHU) dimaksudkan untuk memenuhi kebutuhan rumah layak huni sebagai unsur kesejahteraan sosial. Konsep program bantuan

Rehabilitasi Sosial Rumah Tidak Layak Huni adalah memperbaiki kondisi rumah baik sebagian ataupun seluruhnya yang dilakukan secara gotong royong di atas tanah dengan legalitas kepemilikan yang sah.

Dalam Peraturan Menteri Sosial Republik Indonesia No.20 tahun 2017 tentang Rehabilitasi Rumah Tidak Layak Huni dan Sarana Prasarana Lingkungan disebutkan bahwa Fakir Miskin adalah orang yang sama sekali tidak memiliki sumber mata pencaharian atau mempunyai sumber mata pencaharian tetapi tidak mempunyai kemampuan untuk memenuhi kebutuhan dasar yang layak bagi kehidupan diri dan keluarganya.

### 2.5 Data Mining

*Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Istilah *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. *Data mining*, sering juga disebut sebagai *Knowledge Discovery Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemuka keteraturan, pola atau hubungan dalam kumpulan data berukuran besar (Ridwan, Suyono, & Sarosa, 2013).

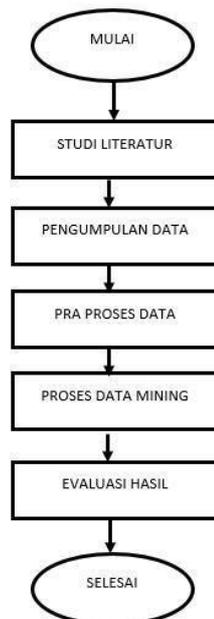
Suyanto, (2017:2) mendefinisikan bahwa *data mining* ditujukan untuk mengekstrak pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basis data dan manajemen data, prapemrosesan data, pertimbangan model dan inferensi, ukuran keterterikan, pertimbangan kompleksitas, prepemrosesan terhadap struktur yang ditemukan, visualisasi dan *online updating*.

### 2.6 Rapid Miner

Baskoro,dkk (2013:8) mengemukakan bahwa *rapidminer* merupakan perangkat lunak yang bersifat terbuka (*open source*). *Rapidminer* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining*, dan analisis prediksi. *Rapidminer* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.

## 3. Kerangka Konsep

Penelitian merupakan rangkaian kegiatan ilmiah. Langkah awal yang harus dilakukan sebelum memulai penelitian adalah studi literatur untuk menentukan *data mining* guna mengolah data dan penentuan alternatif solusi. Kemudian dilakukan pencarian dan pengumpulan data berkaitan dengan masalah yang akan diteliti. Berikut ini merupakan tahapan penelitian yang akan dilakukan pada penelitian ini adalah sebagai berikut:



Gambar 1. Tahapan Penelitian

Secara umum proses-proses tahapan yang dilakukan dalam pelaksanaan penelitian ini adalah:

1. Studi literatur dilakukan dengan membaca dan memahami teori dari buku atau jurnal yang relevan bertujuan untuk mendapatkan teori-teori tentang permasalahan yang telah dirumuskan. Teori ini dapat dijadikan pedoman untuk mendapatkan solusi.

2. Pengumpulan data mentah didapat dari pemerintah daerah kelurahan/desa Pesangkalan berupa data penduduk kelurahan/desa Pesangkalan dengan beberapa atribut yang menentukan pengelompokan penerima bantuan renovasi rumah tidak layak huni.
3. Pra proses *data mining*, meliputi *data cleaning* pembersihan data yang di dalamnya terdapat atribut kosong maupun tidak lengkap, *data integration* atau pengubahan data menjadi format yang sesuai untuk digunakan dalam proses *data mining*, dan *task relevan data* sejalan dengan *data integration* yang memilih atribut yang relevan dengan rumusan masalah.
4. Proses *data mining*. Data yang telah didapatkan kemudian diolah dengan proses *data mining* dimana pengolahan ini menggunakan algoritma *K- Means Clustering*. Hasil proses ini adalah data yang dikelompokkan berdasar kemiripan karakteristik setiap data sehingga ditemukan pola dan informasi yang tersembunyi dari data-data tersebut.

#### 4. Desain Penelitian/ Metodologi

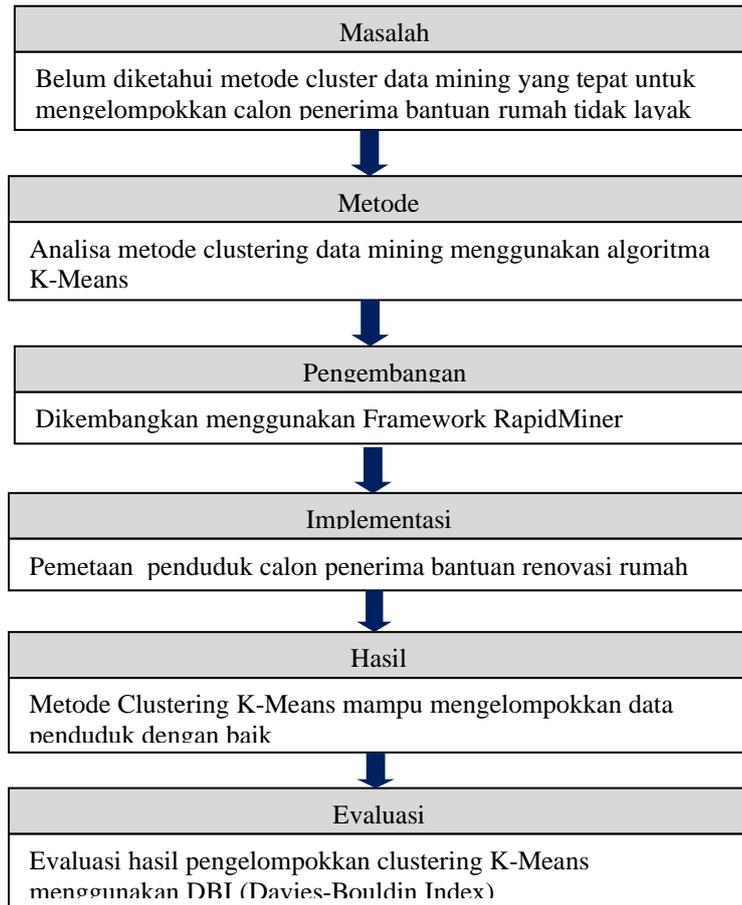
Metode yang akan digunakan pada penelitian ini adalah algoritma *K-Means*. Data yang diperlukan dalam penelitian ini diperoleh dari pemerintah desa/kelurahan Pesangkalan Banjarnegara yang berupa data kependudukan sensus tahun 2016. Selain berisi data kependudukan, didalamnya juga terdapat beberapa atribut data yang menunjang atau mempengaruhi proses data mining *clustering* menggunakan algoritma *K-Means* seperti pendapatan perbulan dan status kepemilikan bangunan tempat tinggal. Berikut adalah data kependudukan kelurahan/desa Pesangkalan tahun 2016.

Data mentah yang telah diperoleh tidak serta merta dapat langsung diolah. Agar dapat diolah menggunakan metode *clustering K-Means* diperlukan pra proses data guna memperoleh atribut dan tipe data yang cocok dan berguna untuk diolah dalam proses *data mining*. Metode yang dilakukan melalui tahapan pra proses data yang terdiri dari :

- a. *Task Relevan Data* adalah melakukan seleksi data yang memiliki atribut yang relevan. Dengan melakukan seleksi data akan membantu tahapan proses *data mining* dalam menemukan pola yang berguna. Oleh karena itu tidak semua atribut dapat digunakan, hanya atribut yang dianggap peneliti berguna dan sebarannya tidak terlalu acak.
- b. Pembersihan Data (*Data Cleaning*). Data-data yang tidak lengkap isinya tidak dilibatkan dalam penelitian, proses ini merupakan bagian dari *data cleaning* atau pembersihan data. Dan data-data yang lengkap isinya, dipilih beberapa atribut agar data yang akan diolah benar-benar relevan dengan kebutuhan.
- c. Integrasi Data (*Data Integration*). *Data Integration* merupakan proses pengubahan atau penggabungan data ke format yang sesuai untuk diproses dalam *data mining*. Seringkali data yang digunakan dalam proses *data mining* mempunyai format yang belum langsung dapat digunakan. Oleh karena itu beberapa format data perlu dirubah. Beberapa data formatnya akan diinisialisasikan ke dalam bentuk angka agar dapat diolah dengan menggunakan metode *clustering K-Means*.

d.

Tabel 1. Desain Penelitian



**5. Hasil Penelitian Dan Pembahasan**

**a. Deskripsi Linguistik Hasil Clustering**

Berikut adalah depenelitian linguistik dan tabel hasil pemetaan data kependudukan berdasarkan pendapatan rata-rata perbulan dan status kepemilikan bangunan tempat tinggal:

**Tabel 2. Depenelitian Linguistik Hasil Clustering**

Anggota Cluster	Pendapatan	Status Kepemilikan	Jumlah KK
Cluster 0	Rp 0 - sd Rp 500.000,- Rp 500.000,- sd Rp 1.500.000,-	Bebas Sewa Milik Orang Tua/ Sanak/ Saudara	11 keluarga
Cluster 1	Rp 1.500.000,- sd Rp 2.500.000,- Rp 2.500.000,- sd Rp 3.500.000,- Rp 3.500.000,- sd Rp 4.500.000,- > Rp 4. 500.000,-	Milik SendiriBebas Sewa Kontrak	65 keluarga
Cluster 2	Rp 0 - sd Rp 500.000,- Rp 500.000,- sd Rp 1.500.000,-	Milik SendiriKontrak	246 keluarga

**b. Pemetaan Kelayakan Calon Penerima Bantuan**

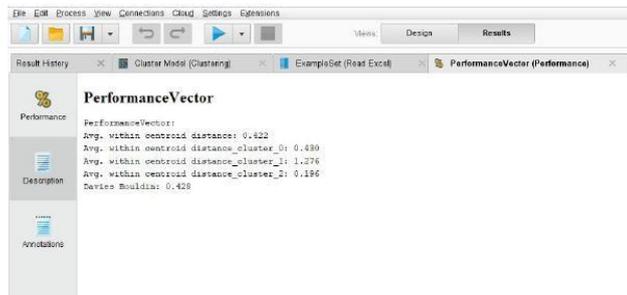
Berdasarkan Upah Minimum Regional (UMR) Jawa Tengah khususnya Kabupaten Banjarnegara yaitu Rp 1.490.000,- perbulan. Serta peraturan pemerintah tentang kriteria calon penerima bantuan Renovasi Rumah Tidak Layak Huni adalah rumah dengan kepemilikan yang sah atas nama milik sendiri maka didapatkan pemetaan/pengelompokkan sebagai berikut:

**Tabel 3 Pemetaan Kelayakan Penerima Bantuan**

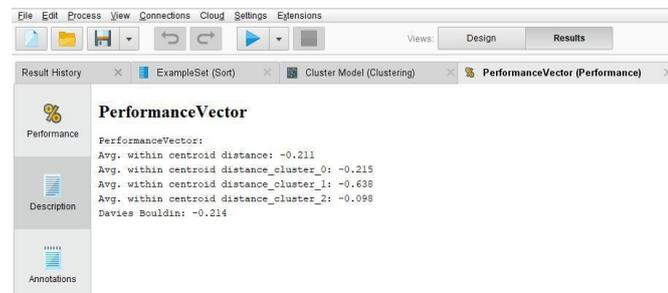
Cluster	Keterangan
0	Kurang cocok menerima bantuan
1	Tidak cocok menerima bantuan
2	Layak menerima bantuan

**c. Nilai Davies Bouldin Index (DBI)**

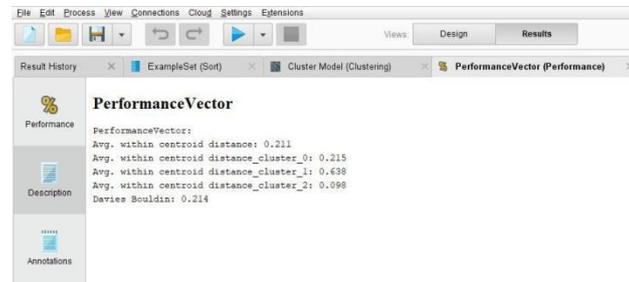
Setelah proses clustering selesai, maka dilakukan validasi clustering menggunakan proses Performance agar dapat diketahui nilai Davies Bouldin Index (DBI) dimana metode ini bertujuan untuk memaksimalkan pengukuran jarak cluster antara satu cluster dengan cluster yang lain.



**Gambar 1. Gambar Performance Vector Description (Average)**



**Gambar 2. Gambar Performance Vector Description (Minimum)**



**Gambar 3. Gambar Performance Vector Description (Maximum)**

Berikut adalah hasil yang diperoleh dari Performance Vector bilangan yang mendekati 0:

**Tabel 4. Hasil Operasi Performance Vector Mendekati 0**

Performance Vector	Value
Average within Centroid Distance	0.211
Average within Centroid Distance Cluster 0	0.215
Avreage within Centroid Distance Cluster 1	0.638
Davies Bouldin	0.214

Semakin jauh jarak antar cluster dan semakin dekat jarak antar anggota cluster maka semakin baik hasil clusternya. DBI bertujuan memaksimalkan jarak antar cluster dan meminimalkan jarak antar anggota

cluster. Dalam hasil performance vector diatas, nilai rata-rata jarak antar centroid adalah 0.211 dan nilai DBI 0.214 yang mana nilai tersebut mendekati 0 yang dapat diartikan bahwa tiap anggota dalam satu cluster berdekatan dan dapat disimpulkan pula hasil nilai DBI cukup baik.

## 6. Kesimpulan

Dari penelitian yang sudah peneliti lakukan maka dapat ditarik kesimpulan sebagai berikut:

1. Pemetaan penduduk calon penerima bantuan renovasi rumah menggunakan algoritma clustering k-means berdasarkan pendapatan rata-rata perbulan dan status kepemilikan bangunan tempat tinggal menghasilkan tiga cluster yaitu penduduk yang layak menerima bantuan, penduduk yang kurang layak menerima bantuan dan penduduk yang tidak layak menerima bantuan.
2. Hasil pengujian mendapatkan nilai DBI (Davies Bouldin Index) sebesar 0.214 yang berarti memiliki jarak antar anggota *cluster* yang cukup baik karena mendekati angka nol.

## Daftar Pustaka

- Aprilla, D., Ambarwati, L., Baskoro, D. A., Wicaksana, I. W. S. 2013. *Belajar Data Mining dengan RapidMiner*. Jakarta: Open Content Model
- Han, J., Kamber, M., Pei, J. 2012. *Data Mining Concepts and Techniques*. Waltham: Elsevier.
- Jumadi, B.D.S. 2018. *Peningkatan Hasil Evaluasi Clustering Davies Bouldin Index Dengan Penentuan Titik Pusat Cluster Awal Algoritma K-means* [penelitian]. Medan. Universitas Sumatera Utara
- Kementrian Sosial Republik Indonesia, <https://kemsos.go.id/berita/pemerintah-berikan-bantuan-rs-rutilahu-untuk-fakir-miskin>
- Munir, R. 2012. *Matematika Diskrit*. Bandung: Informatika
- Santosa, B., 2007. *Data mining (Teknik Pemanfaatan Data untuk Keperluan Bisnis)*. Surabaya: Graha Ilmu.
- Santosa, B., 2007. *Data mining Terapan*. Surabaya: Graha Ilmu
- Suyanto. 2017. *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika
- Vulandari, R. T. 2017. *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Gavamedia