

# Implementasi Metode XGBoost Dalam Seleksi Atribut Pada Algoritma K-Means Untuk *Clustering* Masyarakat Penerima Bantuan Langsung Tunai

(Studi Kasus: Desa Sinorang, Kecamatan Batui Selatan, Kabupaten Banggai)

Amiruddin, Maryam Hasan, Muhammad Erdiansyah

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Ichsan Gorontalo, Gorontalo, Indonesia

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Ichsan Gorontalo, Gorontalo, Indonesia

Email: \*amier.76@gmail.com, \*maryam@gmail.com, mohersiansyah02@gmail.com

**Abstrak** Penelitian ini mengkaji implementasi metode XGBoost dalam seleksi atribut pada algoritma K-Means untuk clustering masyarakat penerima Bantuan Langsung Tunai (BLT). Dalam konteks ini, 14 atribut awal digunakan untuk menggambarkan karakteristik rumah tangga penerima BLT: Luas Lantai, Lantai Rumah, Dinding Rumah, MCK, Sumber Listrik, Sumber Air, Bahan Bakar, Konsumsi, Pakaian, Tidak Sanggup Berobat, Sumber Penghasilan KK, Penghasilan KK, Pendidikan KK, dan Tabungan. Metode XGBoost diaplikasikan untuk menyeleksi atribut yang paling relevan dalam menentukan kelompok penerima BLT. Dari hasil seleksi, ditemukan tiga atribut penting yaitu Luas Lantai, Lantai Rumah, dan Penghasilan KK. Implementasi K-Means clustering dilakukan dua kali, pertama menggunakan seluruh atribut dan kedua menggunakan tiga atribut penting yang telah diseleksi oleh XGBoost. Analisis hasil clustering menunjukkan bahwa sebelum seleksi atribut, nilai Davies-Bouldin Index (DBI) sebesar 1.325. Setelah seleksi atribut penting, nilai DBI menurun menjadi 0.800. Penurunan nilai DBI sebesar 0.525 ini mengindikasikan bahwa hasil clustering menjadi lebih optimal setelah penerapan XGBoost. Dengan demikian, penelitian ini menyimpulkan bahwa penggunaan XGBoost untuk seleksi atribut dapat meningkatkan kinerja K-Means dalam clustering masyarakat penerima BLT, menghasilkan grup yang lebih jelas dan homogen. Temuan ini memiliki implikasi penting untuk meningkatkan efisiensi dan efektivitas program penyaluran BLT dengan mendasarkan keputusan pada atribut yang paling berpengaruh.

**Kata Kunci:** XGBoost, K-Means, Clustering, Seleksi Atribut, Bantuan Langsung Tunai

**Abstract** This study examines the XGBoost method implementation in attribute selection on the K-Means algorithm for clustering community recipients of Direct Cash Assistance. In this context, 14 initial attributes are employed to describe the characteristics of the household recipients of Direct Cash Assistance, namely Floor Area, House Floor, House Wall, Toilets, Electricity Source, Water Source, Fuel, Consumption, Clothes, Inability to Get Medical Treatment, Source of Household Income, Household Income, Household Education, and Savings. The XGBoost method was applied to select the most relevant attributes in determining the Direct Cash Assistance recipient group. Based on the selection results, three important attributes are found, namely Floor Area, House Floor, and Household Income. The implementation of K-Means clustering is twice. The first employs all attributes and the second takes the three important attributes selected by XGBoost. The analysis of the clustering results shows that before attribute selection, the Davies-Bouldin Index (DBI) value indicates 1.325. After the selection of important attributes, the DBI value decreases to 0.800. The decrease in DBI value by 0.525 indicates that the clustering results become more optimal after the XGBoost implementation. This study infers that the XGBoost use for attribute selection can improve the K-Means performance in clustering Direct Cash Assistance beneficiaries, resulting in clearer and more homogenous groups. This finding has important implications for improving the efficiency and effectiveness of the Direct Cash Assistance distribution program by basing decisions on the most influential attributes.

**Keywords:** XGBoost, K-Means, Clustering, attribute selection, Direct Cash Assistance

## 1. PENDAHULUAN

Program bantuan langsung tunai memiliki peran penting dalam menyediakan dukungan finansial kepada masyarakat yang membutuhkan, namun, pemilihan penerima yang tepat menjadi kunci dalam memastikan alokasi dana yang efisien dan manfaat yang adil. Algoritma K-Means, yang digunakan untuk mengelompokkan penerima berdasarkan atribut tertentu, dapat ditingkatkan kualitasnya dengan mengintegrasikan metode XGBoost yang efektif dalam seleksi atribut. Hal ini dapat membantu dalam mengidentifikasi kelompok masyarakat yang paling membutuhkan bantuan, mengurangi risiko kesalahan pemilihan, serta memaksimalkan dampak positif dari program bantuan langsung tunai. Oleh karena itu, implementasi kombinasi ini memiliki potensi besar untuk

meningkatkan efektivitas program bantuan sosial dan mendukung tujuan kesejahteraan masyarakat yang lebih baik.[1]

Desa Sinorang sebagai salah satu desa yang berada di Kecamatan Batui selatan, Kabupaten Banggai yang terletak di Provinsi Sulawesi Tengah sejak tahun 2019 sampai dengan sekarang telah melaksanakan Program Bantuan Langsung Tunai Dana Desa (BLT-DD) yang di salurkan kepada warga desa. Akan tetapi, berdasarkan hasil evaluasi pelaksanaan Program Bantuan Langsung Tunai di lapangan pada tahun 2020 sampai 2022 di Desa Sinorang, masih terdapat masalah penyaluran bantuan yang tidak tepat sasaran dan merata diantaranya karena kurangnya ketelitian dan kerumitan dalam pengolahan data dalam menentukan masyarakat miskin yang menjadi prioritas utama penerima bantuan diantara banyaknya data penduduk yang mengusulkan menerima bantuan. Oleh karena itu, pemerintah desa harus lebih teliti lagi dalam pemilihan warga yang akan di beri bantuan sehingga warga yang diberi benar-benar layak menerimanya dan dari golongan yang kurang mampu.[1]

Sementara itu, kelemahan atau permasalahan pada algoritma K-Means ialah tidak dapat menentukan atribut yang paling penting dalam menghasilkan clustering yang ideal, sehingga diperlukan algoritma lain untuk dapat mengatasi masalah tersebut. Ada beberapa algoritma yang bisa digunakan yaitu: Silhouette Score, Mean Difference dan Gradient Boosting (XGBoost). Algoritma Silhouette Score dan Mean Difference tidak dapat secara otomatis menentukan urutan atribut yang penting karena pada umumnya kedua algoritma tersebut digunakan untuk mengevaluasi performa clustering..[2]

Dengan demikian, untuk mengetahui seberapa baik hasil clustering yang nanti diperoleh dari kinerja Metode XGBoost Dalam Seleksi Atribut Pada Algoritma K-Means Untuk Clustering Masyarakat Penerima Bantuan Langsung Tunai, menggunakan Davies Bouldin Index (DBI) merupakan metrik evaluasi yang umum digunakan untuk mengukur kualitas hasil clustering dalam algoritma k-means. Penggunaan DBI memberikan pemahaman yang lebih mendalam tentang sejauh mana objek-objek dalam satu kluster saling berdekatan dan seberapa jauh kluster tersebut berbeda satu sama lain. DBI juga dapat digunakan sebagai panduan untuk menentukan jumlah kluster yang optimal dalam k-means, karena nilai DBI yang rendah menunjukkan bahwa kluster-kluster tersebut memiliki karakteristik yang baik. Oleh karena itu, integrasi DBI dalam analisis clustering k-means dapat memberikan wawasan yang lebih baik dalam mengukur dan meningkatkan kualitas hasil clustering.[3]

## 2. TINJAUAN PUSTAKA

### 2.1 Bantuan Langsung Tunai

Program Bantuan Langsung Tunai atau sering dikenal dengan (BLT) merupakan salah satu program bantuan sosial yang dikeluarkan pemerintah melalui anggaran dana desa yang diberikan kepada keluarga kurang mampu pada suatu wilayah atau Desa berupa pemberian dana tunai sebesar Rp.600.000 untuk setiap keluarga miskin yang memenuhi kriteria dan diberikan selama 3 bulan dan Rp.300.000 setiap bulan untuk tiga bulan berikutnya. Dalam mengatasi pemuliharaan ekonomi nasional yang diakibatkan oleh covid-19. Dalam upaya mencapai target penurunan kemiskinan yang tercantum dalam Rancangan Pembangunan Jangka Menengah Desa (RPJM-Desa) yang mengedepankan “pembangunan dari masyarakat, oleh masyarakat dan untuk masyarakat” dalam mewujudkan kesejahteraan masyarakat desa, mempercepat pemerataan dan keadilan.[4]

### 2.2. Data Mining

Menurut Fayaad et al. (1996) didalam buku yang ditulis Suyanto (2017), tugas-tugas data mining dapat dikelompokkan ke dalam enam kelompok sebagai berikut:[5]

1. Klasifikasi (classification): men-generalisasi struktur yang diketahui untuk diaplikasikan pada data-data baru.
2. Klasterisasi (clustering): mengelompokkan data, yang tidak di ketahui label kelasnya, kedalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya.
3. Regresi (regresion): menentukan suatu fungsi yang memodelkan data dengan galat (kesalahan prediksi) seminimal mungkin.
4. Deteksi anomali (anomali detection): mengidentifikasi data yang tidak umum, bisa berupa outlier (pencilan), perubahan atau deviasi yang mungkin sangat penting dan perlu investigasi lebih lanjut.
5. Pembelajaran aturan asosiasi (association rule learning) atau pemodelan kebergantungan (dependency modeling): mencari relasi antar variabel
6. Perangkuman (summarization): menyediakan representasi data yang lebih sederhana, meliputi visualisa si dan pembuatan laporan.

### 2.3 Clustering

*Clustering* adalah proses pembagian satu set objek data ke dalam kelompok-kelompok yang disebut cluster. Objek yang terdapat pada suatu cluster memiliki karakteristik yang serupa antara satu dengan yang lain, dan berbeda dengan objek yang terdapat dalam cluster lainnya.[6]

Du (2010) menjelaskan bahwa clustering adalah sebuah proses dalam memisahkan data yang tidak berlabel menjadi beberapa kelompok data yang memiliki kesamaan satu sama lain. Clustering memiliki bebrapa variabel

antara lain: variabel K yang merupakan jumlah kluster, variabel C yang merupakan label dari suatu cluster, dan variabel P merupakan dataset.[7]

#### 2.4 Algoritms K-Means

Menurut Suwarno (2006) dalam penjelasannya tahapan algoritma K-Means yaitu sebagai berikut:[8]

- Pada tahap pertama, tentukan  $K$  terlebih dahulu sebagai jumlah *cluster* yang ingin dibentuk
- Tahap kedua, mengambil nilai random yang akan dijadikan untuk pusat *cluster* awal (*centroid*) sebanyak  $K$
- Tahap ketiga, menghitung jarak setiap data yang diinput terhadap masing-masing *centroid* atau kluster awal dengan menggunakan rumus jarak *Euclidean* hingga menentukan jarak yang paling mendekati dari setiap data yang ada dengan *centroid*. Rumus persamaannya *Euclidean Distance* berikut:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \dots\dots\dots (2.1)$$

Dimana :

- $d$  = jarak
- $i$  = indeks dari cluster
- $j$  = indeks dari variable
- $x_i$  = data kriteria ke- $i$
- $\mu_j$  = *centroid* pada cluster ke- $j$

- Menentukan setiap data yang memiliki kedekatan dengan *centroid* dalam jarak terkecil
- Memperbarui nilai *centroid* yang telah diperoleh dari rata-rata *cluster* yang berkaitan, rumus yang digunakan yaitu :

$$\mu_j(t + 1) = \frac{1}{N_{sj}} \sum x_j \dots\dots\dots (2.2)$$

Dimana :

- $j$  = indeks dari variable
- $\mu_j(t + 1)$  = *centroid* baru pada iterasi ke- $j$
- $N_{sj}$  = banyak data pada cluster  $S_j$
- $S_j$  = Silhoutte data ke- $j$
- $X_j$  = data kriteria ke- $j$

- Melakukan perulangan dari Langkah 2 sampai pada Langkah 5 sampai anggota dari setiap *cluster* tidak ada yang berubah. Dan apabila Langkah ke-6 telah terpenuhi, maka nilai pada *cluster* ( $\mu_j$ ) di iterasi terakhir yang akan diambil sebagai kesimpulan akhir.

#### 2.5 XGBoost

XGBoost merupakan metode yang banyak digunakan dalam pemodelan prediktif, seperti regresi dan klasifikasi. Meskipun XGBoost dapat menghasilkan skor kepentingan atribut, biasaya akan digunakan untuk menentukan atribut yang paling penting dalam pemodelan prediktif, bukan untuk seleksi atribut *clustering* K-means. Namus demikian, XGBoost dapat digunakan sebagai alat tambahan untuk menganalisis atribut pada *clustering* K-means.

Skor kepentingan dalam model XGBoost dihitung berdasarkan seberapa sering sebuah fitur digunakan untuk membuat keputusan pemisahan di semua pohon keputusan dengan menggunakan rumus *Weight* (bobot).

$$Weight(f) = \sum_{i=1}^N I(f = f_i) \dots\dots\dots (2.3)$$

Ket:

- Weight* ( $f$ ) adalah bobot fitur
- $f$  adalah fitur yang sedang dinilai
- $N$  adalah jumlah pohon keputusan dalam model
- $f_i$  adalah fitur yang digunakan untuk membuat keputusan pemisahan model
- $I(-)$  adalah fungsi indikator yang bernilai 1 jika pernyataan didalamnya benar, dan 0 jika tidak.

Skor kepentingan (*importance score*) dari model XGBoost dapat memberikan indikasi bahwa atribut tersebut berpengaruh, skor kepentingan yang tinggi menunjukkan bahwa fitur tersebut sering digunakan dalam pemisahan pohon keputusan dan memiliki dampak yang signifikan pada hasil prediksi model.[9]

#### 2.6 Metode Elbow

Metode siku, atau yang lebih dikenal sebagai metode *elbow* dalam analisis kluster, merupakan suatu pendekatan yang digunakan untuk menentukan jumlah optimal dari kluster dalam suatu dataset. Dalam metode ini, kita menggambarkan kurva antara "*within the sum of squares*" (WSS) dan jumlah kluster yang digunakan. Ketika kurva tersebut diplot, bentuknya menyerupai lengan manusia, dan inilah sebabnya metode ini sering disebut sebagai metode siku.[10]

#### 2.7 Davies Bouldin Index

*Davies Bouldin Index* (DBI) adalah metrik evaluasi yang digunakan untuk mengukur seberapa baik kualitas *clustering* yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin bertepatan pada tahun 1997. Dengan menerapkan DBI suatu cluster dapat dianggap mempunyai skema *clustering* yang optimal yakni yang memiliki DBI minimal.[10]

Untuk menghitung setiap kluster yang merupakan rata-rata jarak antara *centroid* kluster tersebut dan semua titik dalam kluster itu sendiri dengan rumus:

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \dots\dots\dots (2.4)$$

Dimana:

- $R_{ij}$  merupakan nilai R untuk kluster i dan j
- $d_{ij}$  yaitu jarak antara *centroid* kluster i dan j
- $S_i$  dan  $S_j$  yaitu jarak rata-rata antara setiap titik kluster i dan *centroid* dari kluster j

Selanjutnya menghitung nilai DBI, yang merupakan rata-rata dari semua nilai DB untuk setiap kluster, rumusnya adalah:

$$DB = \frac{1}{k} \sum_{i=1}^k \max R_{ij, i \neq j} \dots\dots\dots (2.5)$$

Dimana:

- DBI yaitu *Davies Bouldin Index*
- $K$  ialah jumlah kluster yang digunakan

### 2.8 Perangkat Lunak Pendukung

Perangkat lunak pendukung yang digunakan penulis dalam penelitian ini yaitu *Python*.

*Python* merupakan bahasa pemrograman tingkat tinggi, penulisan kode/sintaks lebih sederhana, bersifat *open-source*, *cross-platfrom* dan cocok digunakan untuk Data Scientist serta didukung banyak library seperti *Numpy*, *SciPy*, *Pandas*, *Scikit-Learn*, dan *Matplotlib*.

## 3. METODE PENELITIAN

### 3.1 Jenis Metode, Subjek, Waktu dan Lokasi Penelitian

Berdasarkan latar belakang dan kerangka pikir seperti yang sudah diuraikan dalam pembahasan 1 dan pembahasan 2, maka yang menjadi objek penelitian yaitu “Implementasi Metode XGBoost Dalam Seleksi Atribut Pada Algoritma K-means Untuk *Clustering* Masyarakat Penerima Bantuan Langsung Tunai.(Studi Kasus Desa Sinorang Kec.Batui Selatan)”. Penelitian ini dimulai pada tanggal 12 september tahun 2023 yang berlokasi di Kantor Desa Sinorang, Kecamatan Batui Selatan.

### 3.2 Pengumpulan data

Metode pengumpulan data yang digunakan untuk mendapatkan data dan informasi digunakan 2 (dua) jenis data, yaitu data primer dan data sekunder sebagai berikut:

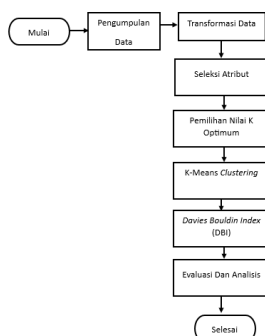
#### 3.2.1 Data Primer (Observasi, Wawancara)

Data primer dalam penelitian ini adalah data yang di kumpulkan langsung oleh peneliti di lokasi penelitian yang berkaitan dengan objek yang diteliti. Kemudian wawancara dilakukan dengan mengajukan pertanyaan ke pihak Aparat Desa Sinorang.

#### 3.2.2 Data Sekunder (Keperustakaan)

Data sekunder merupakan pengambilan informasi dengan melakukan pengkajian ke perpustakaan yang berisi dasar-dasar teori. Metode ini digunakan untuk mengambil contoh dokumen yang berhubungan dengan objek penelitian

### 3.3 Pemodelan



### 3.4 Pra Pengolahan Data

Sebelum data akan diolah, terlebih dahulu dilakukan proses preprocessing agar peneliti dapat menghindari gangguan terhadap data-data yang tidak konsisten, tujuannya agar hasil output memiliki tingkat keakuratan yang tinggi.

### 3.5 Transformasi Data

Pada tahap ini transformasi data dilakukan bertujuan untuk mengubah data dari nominal menjadi numerik dikarenakan dalam proses perhitungan K-means data yang digunakan adalah nilai numerik.

### 3.6 Hasil Clustering

Hasil dari klasterisasi merupakan keluaran, pada data yang diperoleh dari proses klasterisasi menggunakan metode K-means berdasarkan data yang diperoleh dari data penerima bantuan langsung tunai.

### 3.7 Evaluasi

Evaluasi bertujuan untuk mengetahui hasil performansi dan metode yang digunakan, evaluasi dilakukan terhadap output data yang dihasilkan akan dimasukkan kedalam cluster hasil untuk menghitung nilai akurasi.

## 4. HASIL DAN PEMBAHASAN

Implementasi metode XGBoost dalam seleksi atribut pada algoritma K-Means untuk *clustering* masyarakat penerima bantuan langsung tunai. Dilakukan berdasarkan pemodelan penelitian yang telah diuraikan sebelumnya, yaitu dimulai dari pengumpulan data, transformasi data, seleksi atribut, pemilihan nilai K optimum, K-Means *Clustering*, *Davies Bouldin Index*, evaluasi dan analisis hingga visualisasi hasil clustering. Proses ini dilakukan dengan menggunakan Google Colaboratory sebagai platform dan menggunakan bahasa pemrograman Python. Pada proses ini, digunakan beberapa library seperti *Numpy*, *Pandas*, *Seabron*, *Matplotlib*, dan *Scikit Learn* untuk proses analisis dan implementasi algoritma *clustering* yakni K-Means.

### 4.1 Pengumpulan Dataset

Dataset yang akan digunakan pada penelitian ini bersumber dari Kantor Desa Sinorang dengan jumlah penerima bantuan langsung tunai pada tahun 2022 sebanyak 107 data, dengan atribut kriteria penerima sebanyak 14 atribut.

Tabel 4.1 Sampel Dataset

No	Nama Penerima	Luas Lantai	Lantai Rumah	Penghasilan KK
1	Waser	8m2	Tanah	Rp.600.000
2	Ikram	12m2	Tanah	Rp.600.000
3	Hasdin	<8m2	Kayu murah	Rp.500.000
4	Mawar	<8m2	Kayu muarah	Rp.500.000
5	Tani	11m2	Tanah	Rp.400.000
6	Rambitan	8m2	Tanah	Rp.500.000
7	Siti Khodija	15m2	Tanah	Rp.600.000
8	Ruiya	12m2	Tanah	Rp.400.000
9	Sinayim	8m2	Tanah	Rp.500.000
10	Aisya	<8m2	Tanah	Rp.500.000

### 4.2 Transformasi Dataset

Sebelum dilakukan pemodelan dengan metode K-Means clustering terlebih dahulu dilakukan beberapa preprocessing data terhadap dataset pada Tabel 4.1 diatas khususnya untuk atribut penerima bantuan langsung tunai dilakukan konversi dari data jenis kategorikal menjadi numerik agar dapat menjadi suatu jenis type data yaitu numerik.

#### 1. Konversi Data Untuk Atribut Luas Lantai

Tabel 4.2 Konversi Atribut Luas Lantai

No	Nama Atribut	Nilai Atribut
1	<8m2	1
2	8m2	2
3	11m2	3
4	12m2	4
5	15m2	5

#### 2. Konversi Data Untuk Atribut Lantai Rumah

Tabel 4.3 Konversi Atribut Lantai Rumah

No	Nama Atribut	Nilai Atribut
1	Tanah	1
2	Kayu Murah	2

### 3. Konversi Data Untuk Atribut Penghasilan KK

Tabel 4.4 Konversi Atribut Penghasilan KK

No	Nama Atribut	Nilai Atribut
1	Rp.300000	1
2	Rp.800000	2
3	Rp.700000	3
4	Rp.400000	4
5	Rp.600000	5
6	Rp.500000	6

Tabel 4.5 Sampel Dataset Akhir

No	Nama Penerima	Luas Lantai	Lantai Rumah	Penghasilan KK
1	Waser	2	1	5
2	Ikram	4	1	5
3	Hasdin	1	2	6
4	Mawar	1	2	6
5	Tani	3	1	4
6	Rambitan	2	1	6
7	Siti Khodija	5	1	5
8	Ruiya	4	1	4
9	Sinayim	2	1	6
10	Aisyah	1	1	6

### 4.3 Seleksi Atribut

Berdasarkan dataset 4.1 dilakukan seleksi atribut yang akan digunakan dalam proses clustering yaitu Luas Lantai, Lantai Rumah dan Penghasilan KK. Ketiga atribut tersebut di pilih karena berdasarkan nilai skor kepentingan terbesar yang dihitung dengan pohon keputusan dari metode XGBoost dengan bantuan *tools Python*.

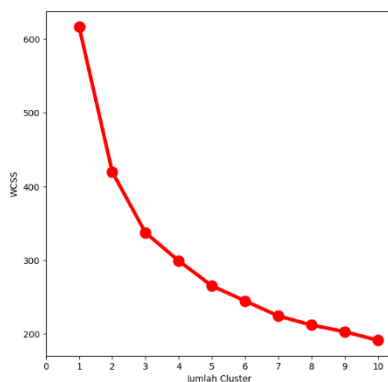
### 4.4 Pemilihan Nilai K Optimum

Penentuan nilai K optimum pada pemodelan K-Means dapat ditentukan dengan menggunakan salah satu metode yaitu metode Elbow. Berikut potongan kode program untuk penggunaan metode *Elbow*.

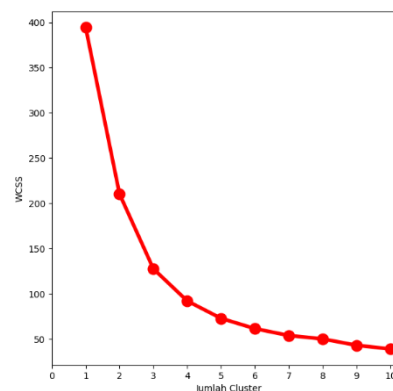
```
# Menentukan jumlah cluster yang optimum dengan teknik Elbow

WCSS = []
for i in range(1,11):
    model = KMeans(n_clusters = i, init = "k-means++", max_iter = 300, n_init = 10, random_state = 0)
    model.fit(x)
    WCSS.append(model.inertia_)
fig = plt.figure(figsize = (7,7))
plt.plot(range(1,11),WCSS, linewidth=4, markersize=12,marker='o',color = 'red')
plt.xticks(np.arange(11))
plt.xlabel("Jumlah Cluster")
plt.ylabel("WCSS")
plt.show()
```

Gambar 4.1 Kode Program Metode *Elbow*



Gambar 4.2 Sebelum Seleksi Atribut



Gambar 4.3 Setelah Seleksi atribut

Berdasarkan Gambar 4.1, sebelum dilakukan seleksi atribut posisi jumlah *cluster* yang optimal terletak pada nilai 3 (tiga) kluster. Selanjutnya setelah dilakukan seleksi atribut posisi jumlah *cluster* yang optimal terletak pada 4 (empat) kluster. Hal ini dapat dilihat dari titik kedua grafik pada gambar diatas di mana penurunan inersia tidak

lagi signifikan setelah titik jumlah *cluster* 4 (empat). Oleh karena itu, nilai  $K = 4$  dijadikan sebagai nilai optimal dalam pemodelan K-Means pada *Clustering* penerima bantuan langsung tunai dipenelitian ini.

#### 4.5 Davies Bouldin Index

Hasil *clustering* menggunakan K-Means, dapat diukur atau dinilai berdasarkan jumlah kluster yang dibuat. Sebuah indeks yang mengukur tingkat perbedaan antar kluster dikenal sebagai *Davies Bouldin Index*. Dimana nilai indeks yang lebih rendah, maka hasil *clustering*nya dianggap lebih baik. Hasil penilaian *clustering* bantuan langsung tunai ditunjukkan pada tabel 4.

Tabel 4.6 Hasil Pengujian

No	Uraian	Jumlah Cluster	Hasil DBI
1	Sebelum seleksi atribut	3	1.325
2	Setelah seleksi atribut	4	0.800

Berdasarkan Tabel 4. Dapat dilihat bahwa nilai *index* terkecil yaitu 0.800 pada jumlah cluster 4 (empat) setelah seleksi atribut, sehingga pada *clustering* masyarakat penerima bantuan langsung tunai dipilih 4 (empat) *cluster* sebagai jumlah *cluster* yang paling ideal.

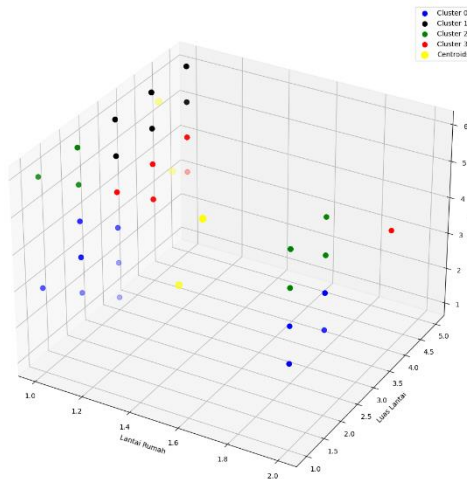
#### 4.6 Evaluasi dan Analisis

Evaluasi hasil *clustering* masyarakat penerima bantuan tunai dapat di analisis dengan hasil visualisasi *clustering* menggunakan *tools Python* di tunjukan pada kode program berikut.

```
# Visualisasi hasil cluster dengan 3d scatterplot menggunakan matplotlib
fig = plt.figure(figsize = (13,13))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(x[y_clusters == 0],x[y_clusters == 0,1],x[y_clusters == 0,2], s = 50 , color = 'blue', label = "Cluster 0")
ax.scatter(x[y_clusters == 1],x[y_clusters == 1,1],x[y_clusters == 1,2], s = 50 , color = 'black', label = "Cluster 1")
ax.scatter(x[y_clusters == 2],x[y_clusters == 2,1],x[y_clusters == 2,2], s = 50 , color = 'green', label = "Cluster 2")
ax.scatter(x[y_clusters == 3],x[y_clusters == 3,1],x[y_clusters == 3,2], s = 50 , color = 'red', label = "Cluster 3")
#ax.scatter(x[y_clusters == 4],x[y_clusters == 4,1],x[y_clusters == 4,2], s = 50 , color = 'purple', label = "Cluster 4")
ax.scatter(model.cluster_centers_[:,0],model.cluster_centers_[:,1],model.cluster_centers_[:,2], s = 100, c = "yellow", label = "Centroids")
ax.set_xlabel('Dinding Rumah')
ax.set_ylabel('sumber listrik')
ax.set_zlabel('pakaian')
ax.legend()
plt.show()
```

Gambar 4.4 Kode Program Visualisasi

Hasil visualisasi *clustering* masyarakat penerima bantuan langsung tunai sesuai hasil kode program pada gambar 4.4 Hasilnya ditunjukkan pada gambar 4.5.



Gambar 4.5 Visualisasi Hasil Clustering

Berdasarkan Gambar 5.6 diatas, ditunjukkan bahwa setiap *cluster* sudah dikelompokkan cukup baik namun penyebaran anggota setiap *cluster* masih kurang, hal ini disebabkan nilai ketiga atribut yang digunakan nilainya hanya berkisar 1 sampai dengan 6. Dan hasil *clustering* secara detail dapat ditunjukkan pada Tabel 4.7 berikut:

Tabel 4.7 Sampel Hasil Clustering

No	Nama Penerima	Luas Lantai	Lantai Rumah	Penghasilan KK	Cluster
1	Waser	2	1	5	C3
2	Ikram	4	1	5	C2
3	Hasdin	1	2	6	C3
4	Mawar	1	2	6	C3
5	Tani	3	1	4	C4
6	Rambitan	2	1	6	C3
7	Siti Khodija	5	1	5	C2

No	Nama Penerima	Luas Lantai	Lantai Rumah	Penghasilan KK	Cluster
8	Ruiya	4	1	4	C4
9	Sinayim	2	1	6	C3
10	Aisya	1	1	6	C3

## 5. KESIMPULAN

1. Hasil analisis seleksi atribut penting pada K-Means untuk *clustering* masyarakat penerima bantuan langsung tunai dengan menggunakan XGBoost dari 14 atribut yakni Luas Lantai, Lantai Rumah, Dinding Rumah, MCK, Sumber Listrik, Sumber Air, Bahan Bakar, Konsumsi, Pakaian, Tidak Sanggup Berobat, Sumber Penghasilan KK, Penghasilan KK, Pendidikan KK, Tabungan. didapatkan 3 atribut penting yaitu Luas Lantai, Lantai Rumah dan Penghasilan KK. Dan analisis masing-masing *cluster* pada data penerima bantuan langsung tunai tahun 2022 memperoleh hasil yaitu sebanyak 24 data C1, 40 data C2, 21 data C3 dan 22 data C4.
2. Hasil *clustering* sebelum seleksi atribut dengan XGBoost didapatkan nilai DBI sebesar 1.325, setelah seleksi atribut penting didapatkan nilai DBI sebesar 0.800, didapatkan hasil bahwa nilai DBI setelah penerapan XGBoost lebih kecil sehingga ini menunjukkan bahwa hasil *clustering* yang dihasilkan lebih optimum atau terjadi penurunan nilai DBI sebesar 0.525.

## DAFTAR PUSTAKA

- [1] Sari, Sarmila, and Joy Nashar Utamajaya. "Sistem Pendukung Keputusan Penerima Bantuan Langsung Tunai Dana Desa Menggunakan Metode Algoritma K-Means Clustering." *JUPITER (Jurnal Penelitian Ilmu dan Teknik Komputer)* 14.1 (2022): 150-160.
- [2] Bengnga, Amiruddin, and Reziwati Ishak. "Penerapan XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah." *Jambura Journal of Electrical and Electronics Engineering* 5.2 (2023): 192-196.
- [3] M. Jafar, "Perencanaan Pembangunan Desa," Kementrian Desa, Pembangun Daerah Tertinggal, Dan Transmigrasi Republik Indonesia, 2015. (accessed Mar. 28, 2022).
- [4] Suyanto. (Mei 2017). *DATA MINING Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung.
- [5] Nurahman, Nurahman, and Jetri Susanto. "Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma K-Means." *JURIKOM (Jurnal Riset Komputer)* 10.2 (2023): 461-470.
- [6] Asroni, Asroni, and Ronald Adrian. "Penerapan metode K-means untuk clustering mahasiswa berdasarkan nilai akademik dengan Weka Interface studi kasus pada jurusan Teknik Informatika UMM Magelang." *Semesta Teknika* 18.1 (2015): 76-82.
- [7] Kusumawardani, Y., & Hamzah, A. (2018). Perbandingan Metode Clustering Menggunakan Hierarchical Clustering dan Partional Clustering Untuk Mengelompokan Dokumen Berita. *Jurnal Script*, 23-36.
- [8] Nusrhendratno, S. S. (2022, November). Sintesis Fitur Density Based Feature Selection (DBFS) dan AdaBoots dengan XGBoost Untuk Meningkatkan Performa Model Prediksi. In *Prosiding Seminar Sains Nasional dan Teknologi* (Vol. 12, No. 1, pp. 305-313).
- [9] Priya Pedamkar K- Means Clustering Algorithm. (2019, May 15). EDUCBA. <https://www.educba.com/k-means-clustering-algorithm/>
- [10] Pratama, A. R., Maulana, B., Rianda, R. D., & El Hasyim, S. (2023). Comparison of K-Means and K-Medoids Algorithms for Grouping Video Game Sales Data in North America: Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Penjualan Video Game di Amerika Utara. *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, 3(2), 111-118.