

**IMPLEMENTASI METODE XGBOOST DALAM
SELEKSI ATRIBUT PADA ALGORITMA
K-MEANS UNTUK CLUSTERING
MASYARAKAT PENERIMA
BANTUAN LANGSUNG
TUNAI**

(Studi Kasus: Desa Sinorang, Kecamatan Batui Selatan)

Oleh
MUHAMMAD ERDIANSYAH
T3120116

SKRIPSI

**Untuk memenuhi salah satu syarat ujian
guna memperoleh gelar Sarjana**



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2024**

PERSETUJUAN SKRIPSI

IMPLEMENTASI METODE XGBOOST DALAM SELEKSI ATRIBUT PADA ALGORITMA K-MEANS UNTUK CLUSTERING MASYARAKAT PENERIMA BANTUAN LANGSUNG TUNAI

(Studi Kasus: Desa Sinorang, Kecamatan Batui Selatan)

Oleh

MUHAMMAD ERDIANSYAH

T3120116

SKRIPSI

Untuk memenuhi salah satu syarat ujian
Guna memperoleh gelar Sarjana
Program Studi Teknik Informatika,
Ini telah disetujui oleh Tim Pembimbing

Gorontalo, 6 Juni 2024

Pembimbing I



Amiruddin, M.Kom., MCF
NIDN : 0910097601

Pembimbing II



Maryam Hasan S.Kom., M.Kom
NIDN : 0907099002

PENGESAHAN SKRIPSI

IMPLEMENTASI METODE XGBOOST DALAM SELEKSI ATRIBUT PADA ALGORITMA K-MEANS UNTUK CLUSTERING MASYARAKAT PENERIMA BANTUAN LANGSUNG TUNAI

(Studi Kasus: Desa Sinorang, Kecamatan Batui Selatan)

Oleh

MUHAMMAD ERDIANSYAH

T3120116

Diperiksa oleh Panitia Ujian Sastra Satu (S1)

Universitas Ichsan Gorontalo

1. Ketua Pengudi
Haditsah Annur, M.Kom
 2. Anggota
Muis Nanja, M.kom
 3. Anggota
Serwin, M.Kom
 4. Anggota
Amiruddin, M.Kom, MCF
 5. Anggota
Maryam Hasan S.Kom, M.Kom
- 
- 

Mengetahui

Dekan Fakultas Ilmu Komputer



Irvan Abraham Salihi, M.Kom
NIDN: 0928028101

Ketua Program Studi



Sudirman S.Pana, M.Kom
NIDN: 094038205

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa:

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan dari Tim Pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Penyataan ini saya buat dengan sesungguhnya dan apabila kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, 6 / Juni / 2024
Yang Membuat Pernyataan



Muhammad Erdiansyah

ABSTRACT

MUHAMMAD ERDIANSYAH. T3120116. THE IMPLEMENTATION OF THE XGBOOST METHOD IN ATTRIBUTE SELECTION ON THE K-MEANS ALGORITHM FOR CLUSTERING COMMUNITY RECIPIENTS OF DIRECT CASH ASSISTANCE

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-Mean's 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

ABSTRAK

MUHAMMAD ERDIANSYAH. T3120116. IMPLEMENTASI METODE XGBOSST DALAM SELEKSI ATRIBUT PADA ALGORITMA K-MEANS UNTUK CLUSTERING MASYARAKAT PENERIMA BANTUAN LANGSUNG TUNAI

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

KATA PENGANTAR

Alhamdulillah puji syukur kepada Allah SWT atas berkat dan rahmat-Nyalah peneulis dapat menyelesaikan proposal/skripsi ini dengan judul **“IMPLEMENTASI METODE XGBOOST DALAM SELEKSI ATRIBUT PADA ALGORITMA K-MEANS UNTUK CLUSTERING MASYARAKAT PENERIMA BANTUAN LANGSUNG TUNAI”**. Skripsi ini disusun sebagai salah satu syarat ujian guna memperoleh gelar Sarjana pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis menyadari sepenuhnya bahwa dalam penyusunan skripsi ini tidak akan selesai tanpa bantuan dan dorongan dari berbagai pihak, baik bantuan moril maupun materil. Untuk itu, dengan segala keikhlasan dan kerendahan hati, penulis mengucapkan banyak terima kasih dan penghargaan setinggi-tingginya kepada:

1. Dr. Yuriko Abdussamad, M.Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Dr. Abdul Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan Abraham Salihi, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
4. Bapak Sudirman Melangi, M.Kom, selaku Pembantu Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Ibu Irma Surya Kumala Idris, M.Kom, selaku Pembantu Dekan II Bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
6. Bapak Sudirman S Panna, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
7. Bapak Amiruddin, M.Kom, MCF sebagai Pembimbing Utama dalam penelitian ini yang telah membimbing penulis selama penelitian;
8. Ibu Maryam Hasan, M.Kom, sebagai Pembimbing Pendamping dalam penelitian ini yang telah membimbing penulis selama Penelitian;

9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;
10. Ayahanda Halidun A Sadi. Beliau yang menjadi inti tulang punggung keluarga, merkipun beliau tidak sempat merasakan pendidikan hingga bangku perkuliahan namun beliau mampu mendidik penulis menjadi laki-laki yang kuat dan tegar dalam segala rintangan hingga penulis mampu menyelesaikan studinya sampai sarjana
11. Ibunda Bahiya, pintu surgaku. Beliau sangat berperan penting dalam menyelesaikan program studi penulis. Beliau juga memang tidak sempat merasakan pendidikan hingga bangku perkuliahan, namun gigih dalam memenjatkan doa yang selalu;
12. Seluruh kakak perempuan saya Erlin Aliya S.Pd, Erni Aliya S.Pd, Erva Aliya Amd.Kom, dan Endang Aliya S.Pd yang tidak pernah berhenti memberi segala motivasi, doa yang terus dipanjatkan, dan dukungan sehingga dapat menyelesaikan pendidikan tinggi ini dengan penuh semangat dan tekad yang sangat besar;
13. Terakhir, kepada semua rekan-rekan penulis terkhusus kelas Reguler Pagi B angakatan 2020 yang banyak terlibat dalam membantu penulis untuk menyelesaian Penelitian ini yang tak sempat penulis sebutkan satu-persatu. Semoga Allah, SWT melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis menyadari sepenuhnya bahwa apa yang telah dicapai ini masih jauh dari kata sempurna dan masih banyak terdapat kekurangan, Terimah Kasih.

Gorontalo, 6 / Juni / 2024

Penulis

DAFTAR ISI

| | |
|---|-------------|
| PERSETUJUAN SKRIPSI..... | ii |
| PENGESAHAN SKRIPSI..... | iii |
| PERNYATAAN SKRIPSI..... | iv |
| ABSTRACT | v |
| ABSTRAK | vi |
| KATA PENGANTAR | vii |
| DAFTAR ISI..... | ix |
| DAFTAR GAMBAR..... | xii |
| DAFTAR TABEL..... | xiii |
| DAFTAR PSEUDOCODE | xv |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Identifikasi Masalah | 3 |
| 1.3 Rumusan Masalah | 4 |
| 1.4 Tujuan Penelitian..... | 4 |
| 1.5 Manfaat Penelitian..... | 4 |
| 1.5.1 Manfaat Teoritis..... | 4 |
| 1.5.2 Manfaat Praktis | 4 |
| BAB II LANDASAN TEORI | 5 |
| 2.1 Tinjauan Studi | 5 |
| 2.2 Tinjauan Pustaka..... | 9 |
| 2.2.1 Bantuan Langsung Tunai..... | 9 |
| 2.2.2 Data <i>Mining</i> | 11 |
| 2.2.3 <i>Clustering</i> | 15 |
| 2.2.4 Algoritma K-Means <i>Clustering</i> | 15 |
| 2.2.5 Metode XGBoost..... | 18 |
| 2.2.6 Metode <i>Elbow</i> | 19 |
| 2.2.7 <i>Davies Bouldin Index (DBI)</i> | 19 |

| | |
|--|-----------|
| 2.2.8 Contoh Penerapan Metode XGBoost Pada K-Means <i>Clustering</i> | 20 |
| 2.2.9 Perangkat Lunak Pendukung | 22 |
| 2.3 Keranka Pikir | 23 |
| BAB III METODE PENELITIAN | 24 |
| 3.1 Jenis Metode, Subjek, Waktu dan Lokasi Penelitian | 24 |
| 3.2 Pengumpulan Data | 24 |
| 3.2.1 Data Primer (Observasi, Wawancara) | 24 |
| 3.2.2 Data Sekunder (Keperpustakaan) | 25 |
| 3.3 Pemodelan | 27 |
| 3.4 Pra Pengolahan Data | 27 |
| 3.5 Transformasi Data..... | 28 |
| 3.6 Hasil <i>Clustering</i> | 28 |
| 3.7 Evaluasi | 28 |
| BAB IV HASIL PENELITIAN | 29 |
| 4.1 Hasil Pengumpulan Data..... | 29 |
| 4.2 Hasil pemodelan..... | 31 |
| 4.2.1 Data Preprocessing | 31 |
| 4.2.2 Pemodelan | 36 |
| 4.2.3 Pengujian Model..... | 42 |
| 4.3 Implementasi Hasil Pemodelan..... | 42 |
| 4.3.1 Implementasi Data Preprocessing | 42 |
| 4.3.2 Implementasi Pemodelan | 49 |
| BAB V HASIL DAN PEMBAHASAN | 53 |
| 5.1 Pembahasan Kinerja Model | 53 |
| 5.2 Pembahasan Hasil Pemodelan..... | 53 |
| 5.2.1 Hasil Pemodelan K-Means Sebelum Seleksi Atribut | 53 |
| 5.2.2 Hasil Seleksi Atribut Penting Dengan XGBoost | 55 |
| 5.2.3 Hasil Pemodelan K-Means Setelah Seleksi Atribut | 56 |
| 5.2.4 Analisis Hasil Clustering Berdasarkan Atribut Terpenting | 58 |
| BAB VI PENUTUP | 63 |
| 6.1 Kesimpulan | 63 |

6.2 Saran..... 63

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1 Proses Umum Data <i>Mining</i> | 12 |
| Gambar 2.2 Tahapan Data <i>mining</i> , Han (2006)..... | 14 |
| Gambar 2.3 Contoh Gambar Cluster Pada K-means | 17 |
| Gambar 2.4 Visualisasi Hasil Metode Elbow | 19 |
| Gambar 3.1 Model usulan..... | 27 |
| Gambar 4. 1 Sampel Pohon Keputusan 1 | 38 |
| Gambar 4. 2 Sampel Pohon Keputusan 2 | 39 |
| Gambar 4. 3 Sampel Pohon Keputusan 3 | 39 |
| Gambar 4. 4 Sampel Pohon Keputusan 4 | 39 |
| Gambar 4. 5 Sampel Pohon Keputusan 5 | 40 |
| Gambar 5.1 Jumlah cluster dengan teknik elbow | 54 |
| Gambar 5.2 Evaluasi Davies Bouldin Index (DBI) Sebelum Seleksi Atribut | 54 |
| Gambar 5.3 Ambil Hasil Perhitungan Skor Kepentingan..... | 55 |
| Gambar 5.4 Elbow Setelah Seleksi Atribut | 56 |
| Gambar 5.5 Evaluasi Davies Bouldin Index (DBI) Sesudah Seleksi Atribut | 57 |
| Gambar 5.6 Grafik Hasil Visualisasi | 58 |
| Gambar 5.7 Deskripsi Statistik Dari Python | 59 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Tabel Tinjauan Studi | 5 |
| Tabel 2.2 Tabel Kriteria Keluarga Miskin..... | 9 |
| Tabel 2.3 Atribut Dataset KIP Kuliah | 20 |
| Tabel 3.1 Atribut/Variabel Data | 25 |
| Tabel 4.1 Dataset Penerima Bantuan Langsung Tunai..... | 29 |
| Tabel 4.2 Konversi Atribut Luas Lantai..... | 31 |
| Tabel 4.3 Konversi Atribut Lantai Rumah..... | 31 |
| Tabel 4.4 Konversi Atribut Dinding Rumah..... | 32 |
| Tabel 4.5 Konversi Atribut MCK | 32 |
| Tabel 4.6 Konversi Atribut Sumber Listrik | 32 |
| Tabel 4.7 Konversi Atribut Sumber Air | 32 |
| Tabel 4.8 Konversi Atribut Bahan Bakar..... | 33 |
| Tabel 4.9 Konversi Atribut Konsumsi | 33 |
| Tabel 4.10 Konversi Atribut Pakaian..... | 33 |
| Tabel 4.11 Konversi Atribut Tidak Sanggup Berobat..... | 34 |
| Tabel 4.12 Konversi Atribut Sumber Penghasilan KK | 34 |
| Tabel 4.13 Konversi Atribut Penghasilan KK | 34 |
| Tabel 4.14 Konversi Atribut Pendidikan KK..... | 35 |
| Tabel 4.15 Konversi Atribut Tabungan..... | 35 |
| Tabel 4.16 Dataset Akhir Penrima Bantuan Langsung Tunai | 35 |
| Tabel 4.17 Hasil Clustering Awal..... | 36 |
| Tabel 4.18 Bobot Atribut Pohon Keputusan..... | 40 |
| Tabel 4.19 Skor Kepentingan Atribut..... | 41 |
| Tabel 4.20 Penjelasan Pseudocode 4.1 | 42 |
| Tabel 5.1 Perbandingan Hasil Cluster | 55 |
| Tabel 5.2 Skor Kepentingan | 54 |
| Tabel 5.3 Hasil Deskripsi Statistik | 58 |
| Tabel 5.4 Analisis Masing-masing Cluster | 59 |

Tabel 5.5 Hasil Cluster Penerima Bantuan Langsung Tunai..... 61

DAFTAR PSEUDOCODE

| | |
|---|----|
| Pseudocode 4.1 Import Liblary Python..... | 42 |
| Pseudocode 4.2 Import Google Colab dengan Google Drive | 44 |
| Pseudocode 4.3 Baca File Excel | 44 |
| Pseudocode 4.4 Info dataset..... | 44 |
| Pseudocode 4.5 Konversi Dataset ke CSV..... | 45 |
| Pseudocode 4.6 Info Dataset | 45 |
| Pseudocode 4.7 Menghitung Banyak Data Dalam Suatu Atribut | 45 |
| Pseudocode 4.8 Konversi Data (Luas lantai) Ke Angka | 45 |
| Pseudocode 4.9 Konversi Data (Lantai Rumah) Ke Angka | 46 |
| Pseudocode 4.10 Konversi Data (Dinding Rumah) Ke Angka | 46 |
| Pseudocode 4.11 Konversi Data (MCK) Ke Angka | 46 |
| Pseudocode 4.12 Konversi Data (Sumber Listrik) Ke Angka..... | 46 |
| Pseudocode 4.13 Konversi Data (Sumber Air) Ke Angka | 47 |
| Pseudocode 4.14 Konversi Data (Bahan Bakar) Ke Angka | 47 |
| Pseudocode 4.15 Konversi Data (Konsumsi) Ke Angka..... | 47 |
| Pseudocode 4.16 Konversi Data (Pakaian) Ke Angka | 47 |
| Pseudocode 4.17 Konversi Data (Tidak Sanggup Berobat) Ke Angka..... | 47 |
| Pseudocode 4.18 Konversi Data (Sumber Penghasilan KK) Ke Angka | 48 |
| Pseudocode 4.19 Konversi Data (Penghasilan KK) Ke Angka..... | 48 |
| Pseudocode 4.20 Konversi Data (Pendidikan KK) Ke Angka | 48 |
| Pseudocode 4.21 Konversi Data (Tabungan) Ke Angka..... | 48 |
| Pseudocode 4.22 Pemodelan K-Means | 48 |
| Pseudocode 4.23 Memilit Atribut Yang Terpenting..... | 49 |
| Pseudocode 4.24 Teknik Elbow | 50 |
| Pseudocode 4.25 Visualisasi 3D Hasil Cluster | 51 |

BAB I

PENDAHULUAN

1.1 Latar Belakang

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 Tuanai 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 karna 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]

Salah satu metode *Clustering* adalah Algoritma K-means *clustering*. Metode algoritma *clustering* merupakan metode klasifikasi yang cukup bagus karena tingkat akurasi yang sangat baik dibandingkan dengan metode lain. Algoritma k-means merupakan algoritma *clustering* yang masuk ke dalam kelompok *Unsupervised learning* yang dipakai untuk mengelompokan data ke dalam beberapa bagian dengan sistem partisi. Kesederhanaan k-means menyatakan bahwa algoritma sudah diadopsi dibanyak bidang ini terkemuka karena mampu dengan cepat dan efisien dalam mengelompokan sejumlah data yang besar.[2]

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*. Algoritma *Silhouette Score* mengukur sejauh mana setiap titik data dalam cluster serupa dengan titik-titik dalam cluster lainnya, sementara itu, *Mean Difference* mengukur perbedaan rata-rata antara titik-titik dalam suatu cluster dan titik-titik di luar cluster tersebut. Jadi kedua algoritma tersebut berfokus pada evaluasi hasil *clustering* tanpa memberikan informasi tentang atribut mana yang paling penting atau berkontribusi secara signifikan dalam pembentukan *clustering* tersebut. Sedangkan pada XGBoost hal tersebut dapat dilakukan secara otomatis. Dengan menggunakan XGBoost, penelitian ini berusaha mendapatkan hasil *clustering* yang ideal dengan memilih atribut-atribut yang memiliki pengaruh paling besar terhadap pembentukan *clustering* yang akurat. XGBoost menggunakan teknik pembelajaran mesin ensemble dan memanfaatkan algoritma pohon keputusan untuk mengevaluasi dan menilai kontribusi setiap atribut dalam membuat prediksi. Sehingga pada penelitian ini menggunakan metode XGBoost dalam pemilihan atribut yang penting untuk mendapatkan hasil *clustering* yang ideal.[3]

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 klaster saling berdekatan dan seberapa jauh klaster tersebut berbeda satu sama lain. DBI juga dapat digunakan sebagai panduan untuk menentukan jumlah klaster yang optimal dalam k-means, karena nilai DBI yang rendah menunjukkan bahwa klaster-klaster 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]

Sesuai dengan uraian latar belakang penelitian diatas yang telah peneliti deskripsikan maka peneliti tertarik untuk melakukan penelitian dengan judul, “**Implementasi Metode XGBoost Dalam Seleksi Atribut Pada Algoritma K-means Untuk Clustering Masyarakat Penerima Bantuan Langsung Tunai.(Studi Kasus Desa Sinorang Kec.Batui Selatan)**”. Dengan harapan hasil penelitian yang telah penulis lakukan nantinya dapat dijadikan sistem pendukung keputusan pemerintahan Desa Sinorang dalam menetukan masyarakat penerima bantuan langsung tunai (BLT) di Wilayah Desa Sinorang, secara merata dan tepat sasaran untuk memaksimalkan pelayanan bantuan pemerintahan desa agar dapat tersalurkan dengan baik dan tepat.

1.2 Identifikasi Masalah

Berdasarkan uraian latar belakang diatas maka yang menjadi permasalahan utama pada penelitian ini adalah bagaimana peranan Metode XGBoost dan K-Means yang perlu disesuaikan untuk mendapatkan *cluster* yang lebih akurat dan pemilihan atribut yang lebih relevan, untuk menghasilkan *clustering* yang ideal, serta mengevaluasi kinerja *Davies Bouldin Index* (DBI) dalam mengukur kualitas hasil *clustering* dalam Algoritma K-means.

1.3 Rumusan Masalah

1. Bagaimana hasil analisa seleksi atribut penting pada Algoritma K-Means untuk Clustering masyarakat penerima bantuan langsung tunai dengan menggunakan XGBoost?
2. Seberapa optimum hasil clustering K-Means apabila XGBoost diterapkan dalam seleksi atribut?

1.4 Tujuan Penelitian

1. Menganalisis hasil seleksi atribut penting pada Algoritma K-Means untuk clustering masyarakat penerima bantuan langsung tunai dengan menggunakan XGBoost.
2. Menerapkan XGBoost dalam seleksi atribut penting pada K-Means, sehingga didapatkan hasil clustering yang lebih optimum.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi pihak-pihak yang membutuhkan, baik secara teoritis maupun praktis, diantaranya:

1.5.1 Manfaat Teoritis

Manfaat Teoritis: memberikan masukan bagi perkembangan ilmu pengetahuan tentang teknologi, khususnya dalam bidang ilmu komputer, yaitu berupa penerapan metode XGBoost dalam seleksi atribut pada Algoritma K-Means dalam *Clustering* masyarakat penerima tetap, penerima bantuan langsung tunai.

1.5.2 Manfaat Praktis

Manfaat Praktis: Sumbangan pemikiran, karya, bahan pertimbangan, atau solusi bagi pihak pemerintahan desa guna untuk melakukan penentuan penerima bantuan langsung tunai sehingga berdampak pada peningkatan ketepatan dalam penyaluran bantuan langsung tunai.

BAB II

LANDASAN TEORI

2.1 Tinjauan Studi

Tinjauan studi sangat berguna bagi peneliti dalam memberikan pedoman serta pegangan peneliti yang selanjutnya nantinya dengan adanya penelitian sebelumnya akan mempermudah peneliti dalam melakukan penelitian yang sesuai dengan topik pembahasan. Berikut terdapat beberapa jurnal yang relevan dengan penelitian ini yaitu:

Tabel 2.1 Tabel Tinjauan Studi

| No | Peneliti | Jundul | Tahun | Metode | Hasil |
|----|---------------------------------------|---|-------|------------------|---|
| 1 | Amiruddin Bengga, Rezqiwati Ishak.[3] | Penerapan XGBoost Untuk Seleksi Atribut Pada K-Means Dalam Clustering Penerima KIP Kuliah | 2023 | XGBoost, K-Means | Disimpulkan bahwa dengan menerapkan Metode XGBoost untuk pemilihan atribut yang paling penting pada K-Means dalam clustering prioritas penerima bantuan KIP kuliah terbukti dapat menghasilkan clustering yang ideal, hal ini berdasarkan hasil dari pemilihan dari 12 atribut yang ada |

| No | Peneliti | Jundul | Tahun | Metode | Hasil |
|----|---------------------|--|-------|---------|--|
| | | | | | dipilih 3 atribut yang penting yaitu Pekerjaan Ayah, Penghasilan Ibu, dan Luas Bangunan. Ini berdasarkan pengukuran evaluasi hasil clustering didapatkan nilai index terkecil yaitu 0.819 dari penggunaan 3 atribut terbaik dan nilai K=4 sebagai nilai K optimum. |
| 2 | Dwi Astika sari.[4] | Pengelompokan Data Penduduk Penerima Bantuan sosial Tunai Pandemic Menggunakan Metode Algoritma K-Means Clustering (kantor desa) | 2022 | K-Means | Hasil dari penelitian ini dapat disimpulkan bahwa pada cluster 1 kelompok penerima bantuan sosial tunai pandemic (BSTP) pada grup pekerjaan (x) adalah pekerjaan pabrik, untuk |

| No | Peneliti | Jundul | Tahun | Metode | Hasil |
|----|----------|---------------------|-------|--------|---|
| | | padang Brahrang) | | | <p>kelompok pendapatan (y) ialah 0 -500.000 dan pada grub bantuan (z) sembako. Pada cluster 2 kelompok penerima bantuan sosial pandemic (BSTP) pada grub pekerjaan (x) adalah pekerjs pabrik untuk kelompok pendapatan (y) ialah 600.000 – 1.000.000, dan pada grub bentuk bantuan (z) uang tunai. Pada cluster 3 penerima bantuan sosial pamdemic (BSTP) pada grub pekerjaan(x) adalah tukang becak, untuk kelompok pendapatan(y) ialah 1.100.000 – 1.500.000 dan pada</p> |

| No | Peneliti | Jundul | Tahun | Metode | Hasil |
|----|-----------------|---|-------|---------|---|
| | | | | | bentuk bantuan (z) uang tunai. |
| 3 | Sri Ufriani.[5] | Penerapan Agoritma K-means Clustering Untuk Menentukan Prioritas penerima Bantuan Dana Sosial PKH Di Kelurahan Kampung Singkep | 2023 | K-means | Berdasarkan dari penelitian yang telah dilakukan peneliti bahwasanya data diri dari kelurahan dan data dari rekomendasi peneliti, bahwa dari data desa terdapat 95 warga yang menerima bantuan PKH, sedangkan berdasarkan data diri yang penulis rekomendasi terdapat 123 warga yang berhak menerima bantuan PKH pada Kelurahan Kampung Singkep tidak tercatat. |

2.2 Tinjauan Pustaka

2.2.1 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 pemulihuan ekonomi nasional yang diakibatkan oleh *covid-19*. Dalam upaya mencapai target penurunan kemiskinan yang tercantum dalam Rancangan Pembangunan Jangkah Menengah Desa (RPJM-Desa) yang mengedepankan “pembangunan dari masyarakat, oleh masyarakat dan untuk masyarakat” dalam mewujudkan kesejahteraan masyarakat desa, mempercepat pemerataan dan keadilan.[6]

kebijakan BLT merupakan suatu tindakan dalam menanggulangi kemiskinan akibat dampak pandemic covid-19 dengan pemanfaatan anggaran dana desa. Kebijakan ini diatur oleh peraturan Menteri Desa dan Pembangunan Daerah Tertinggal dan Transmigrasi (Permendes PDTT) Nomor 6 tahun 2020 Tentang Perubahan pada Peraturan Menteri Desa, Pembangunan Daerah Terpolosok, dan Transmigrasi Nomor 11 Tahun 2019 Tantang Prioritas Penggunaan Dana Desa Tahun 2020, merupakan dasar yuridis dan implementatif BLT-DD kepada penduduk miskin di desa.[7]

Tabel 2.2 Tabel Kriteria Keluarga Miskin

| No | Kriteria | Keterangan |
|----|--------------|---|
| 1 | Luas Lantai | <ul style="list-style-type: none"> - <8 m² - 8 – 16 m² - 16 – 25 m² - >25m² |
| 2 | Lantai Rumah | <ul style="list-style-type: none"> - Tanah - Bambu |

| No | Kriteria | Keterangan |
|----|-----------------------|---|
| | | <ul style="list-style-type: none"> - Kayu Murah |
| 3 | Dinding Rumah | <ul style="list-style-type: none"> - Bambu - Rumbia - Kayu Murah - Tembok Tampa Plester |
| 4 | MCK | <ul style="list-style-type: none"> - Tanpa Fasilitas - Bersama Orang Lain - Milik Sendiri |
| 5 | Sumber Listrik | <ul style="list-style-type: none"> - Tidak Ada - PLN subsidi - Menumpang Tetangga |
| 6 | Sumber Air | <ul style="list-style-type: none"> - Sumur - Mata Air Tidak Terlindungi - Sungai - Air Hujan |
| 7 | Bahan Bakar | <ul style="list-style-type: none"> - Kayu - Arang - Minyak Tanah |
| 8 | Konsumsi | <ul style="list-style-type: none"> - Daging 1 Kali/Minggu - Susu 1 Kali/Minggu - Ayam 1 Kali/Minggu |
| 9 | Pakaian | <ul style="list-style-type: none"> - 1 Stel Setahun - >1 Stel Setahun |
| 10 | Tidak Sanggup Berobat | <ul style="list-style-type: none"> - Puskesmas - Poloklinik |
| 11 | Sumber Penghasilan KK | <ul style="list-style-type: none"> - Petati Berlahan <500m² - Buruh Tani - Buruh Nelayan - Buruh Bangunan |

| No | Kriteria | Keterangan |
|----|----------------|---|
| | | - Buruh Perkebunan |
| 12 | Penghasilan KK | - Tidak Berpenghasilan - <Rp.600.000/Bulan - Rp. 600.000 – Rp.800.000 - Rp. >800.000 |
| 13 | Pendidikan KK | - Tidak Sekolah - Tidak Tamat SD - Tamat SD |
| 14 | Tabungan | - Tidak Ada -Tidak Memiliki barang Yang Mudah Dijual Minimal Rp.500.000 |

(Sumber: standar Badan Pusat Statistik,2020)

2.2.2 Data Mining

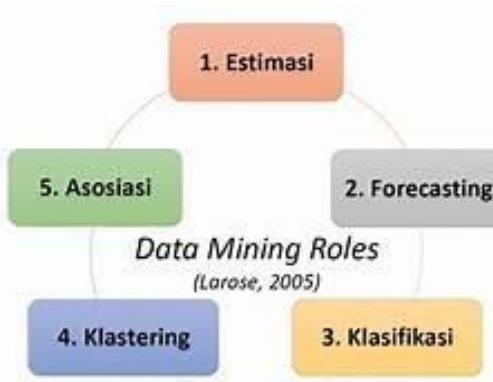
Data mining adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu database yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui.[8]

Menurut Fayaad *et al.* (1996) didalam buku yang ditulis Suyanto (2017), tugas-tugas *data mining* dapat dikelompokkan ke dalam enam kelompok sebagai berikut:[9]

1. Klasifikasi (*classification*): men-generalisasi struktur yang diketahui untuk diaplikasikan pada data-data baru.
2. Klasterisasi (*clustering*): mengelompokan data, yang tidak di ketahui label kelasnya, kedalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya.
3. Regresi (*regresion*): menetukan suatu fungsi yang memodelkan data dengan galat (kesalahan prediksi) seminimal mungkin.
4. Deteksi anomali (*anomaly detection*): mengidentifikasi data yang tidak umum, bisa berupa *outlier* (penculan), 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 visualisasi dan pembuatan laporan.

Menurut Larose (2005) menjelaskan proses yang umumnya dilakukan oleh data *mining* adalah *Estimasi*, *Forecasting*, *Klasifikasi*, *Clustering*, *Asosiasi*. Secara rinci proses pada data *mining* dijelaskan sebagai berikut:[10]



Gambar 2.1 Proses Umum Data *Mining*

1. *Estimasi*

Estimasi memiliki kesamaan dengan prediksi, tetapi variabel target dalam estimasi lebih bersifat numerik daripada kategorikal. Proses pembangunan model melibatkan penggunaan data lengkap yang memberikan nilai variabel target sebagai nilai prediksi. Pada tahap berikutnya, dalam evaluasi berikutnya, nilai *estimasi* untuk variabel target dihasilkan berdasarkan nilai variabel prediksi.

2. *Forecasting*

Bertujuan untuk mengidentifikasi pola yang secara berkala muncul dalam suatu dataset, dan kemudian mengonversi pola tersebut menjadi aturan dan kriteria yang dapat dipahami dengan mudah oleh para ahli di bidang aplikasinya. Aturan yang dihasilkan perlu disajikan dengan cara yang sederhana agar dapat meningkatkan pengetahuan sistem secara efektif. Tugas deskriptif, yang sering kali diperlukan dalam teknik postprocessing, bertujuan untuk melakukan validasi dan menjelaskan hasil dari proses data *mining*.

3. Klasifikasi

Klasifikasi adalah suatu proses di mana sebuah model atau fungsi ditemukan untuk menjelaskan dan membedakan data ke dalam kelas-kelas tertentu. Dalam konteks ini, *klasifikasi* melibatkan pemeriksaan karakteristik dari suatu objek dan penempatannya ke dalam salah satu kelas yang telah ditentukan sebelumnya. Dengan menggunakan model atau fungsi yang telah ditemukan, sistem dapat mengkategorikan objek dengan mempertimbangkan karakteristiknya dan menetapkan label kelas yang sesuai.

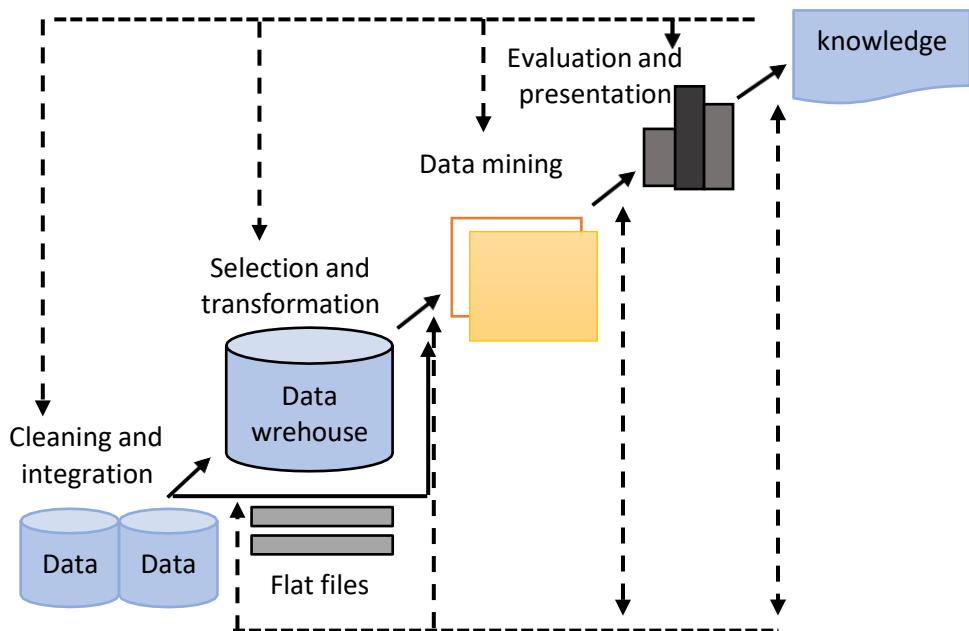
4. Clustering

Clustering adalah suatu metode pengelompokan data di mana tidak ada kelas tertentu yang menjadi dasar pengelompokannya; sebaliknya, objek-objek dikelompokkan bersama berdasarkan kemiripan mereka. Sebuah kluster dapat diartikan sebagai himpunan rekaman yang menunjukkan kemiripan satu sama lain dan berbeda dari rekaman dalam kluster lainnya. Fokus utamanya adalah menghasilkan kelompok-kelompok objek yang menunjukkan kemiripan, sehingga semakin besar kesamaan antar objek dalam satu kluster dan semakin besar perbedaan antara kluster yang berbeda.

5. Asosiasi

Proses *asosiasi* dalam data *mining* bertujuan untuk mengidentifikasi atribut yang sering muncul bersamaan dalam suatu konteks waktu. Dalam konteks bisnis, ini sering dikenal sebagai analisis keranjang belanja atau Market Basket Analysis. Tugas asosiasi berupaya untuk mengukur hubungan antara dua atau lebih atribut dengan tujuan memahami pola dan mencari tahu seberapa sering mereka muncul.[10]

Dan pada penelitian ini jenis metode yang digunakan dalam proses *data mining* ialah *Clustering* atau kalsterisasi data. Menurut Han (2006) menjelaskan bahwa data mining terdiri dari 6 tahapan yang terpola sebagai berikut:[11]



Gambar 2.2 Tahapan Data *mining*, Han (2006)

1. Pembersihan Data (*data cleaning*) merupakan proses untuk menghilangkan *noise* dan data yang tidak konsisten atau data yang tidak relevan. Karena pembersihan data juga dapat mempengaruhi performansi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
2. Integrasi Data (*data intergation*) yaitu penggabungan data dari berbagai database kedalam satu database baru.
3. Seleksi Data (*data section*) data pada *database* sering kali tidak semuanya dipakai. Oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.
4. Transformasi Data (*data transformation*) ialah pengubahan atau pengabungan data kedalam format yang sesuai untuk diproses dalam *data mining*. Sebagai

contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima input data kategorikal.

5. Proses *Mining*, merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi Pola (*pattren evaluation*) yaitu untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan.

2.2.3 *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.[12]

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 beberapa variabel antara lain: variabel K yang merupakan jumlah klaster, variabel C yang merupakan label dari suatu *cluster*, dan variabel P merupakan dataset.[13]

2.2.4 Algoritma K-Means *Clustering*

Algoritma K-Means merupakan algoritma yang tepat digunakan dalam penelitian yang menggunakan dataset yang tidak memiliki label. Algoritma K-Means ialah algoritma klasterisasi yang dapat digunakan untuk mengelompokan objek pengamatan kedalam kelompok (*cluster*) yang memiliki mean tersekat. K-Means tentunya mempunyai beberapa kelebihan seperti mudah untuk diterapkan dan bisa mengklaster data yang berjumlah besar.[14]

Menurut Suwarno (2006) dalam penjelasannya tahapan algoritma K-Means yaitu sebagai berikut:[15]

- a. Pada tahap pertama, tentukan K terlebih dahulu sebagai jumlah *cluster* yang ingin dibentuk
- b. Tahap kedua, mengambil nilai random yang akan dijadikan untuk pusat *cluster* awal (centroid) sebanyak K
- c. 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:

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

Dimana :

d = jarak

i = indeks dari cluster

j = indeks dari variable

x_i = data kriteria ke- i

μ_j = centroid pada cluster ke- j

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

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j \quad \dots \dots \dots \quad (2.2)$$

Dimana :

j = indeks dari variable

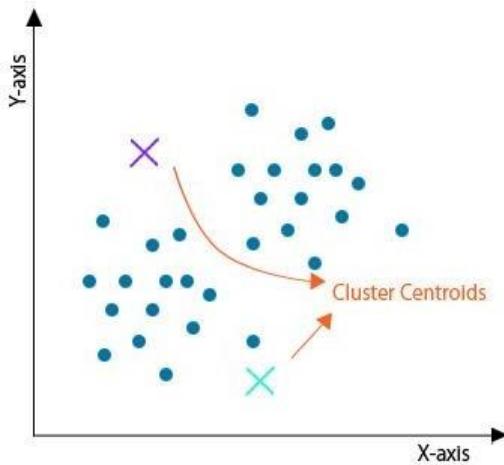
$\mu_j(t+1)$ = *centroid* baru pada iterasi ke- j

N_{sj} = banyak data pada cluster S_j

S_j = Silhouette data ke- j

x_j = data kriteria ke- j

- f. 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* (μ_j) di iterasi terakhir yang akan diambil sebagai kesimpulan akhir.



Gambar 2.3 Contoh Gambar *Cluster* Pada K-means

Algoritma K-Means juga memiliki beberapa kelebihan dan kelemahan diantaranya:

- 1) Kelebihan Algoritma K-means
 - a) Sederhana dan cepat
 - b) Skalabilitas K-means cocok untuk dataset besar dan berkinerja baik dalam hal skala data
 - c) Efisien dalam ruang atribut yang tinggi
 - d) Interpretasi sederhana, hasil K-means dapat dianggap sebagai klaster-klaster terpisah secara intuitif
- 2) Kekurangan Algoritma K-means
 - e) Sensitif terhadap inisialisasi, K-means sangat peka terhadap inisialisasi *centroid* awal, yang akan menyebabkan berbagai hasil *clustering*
 - f) Membutuhkan jumlah *cluster* yang diketahui, K-means memerlukan data sebelumnya tentang jumlah klaster yang diharapkan dalam dataset
 - g) Sensitif terhadap *outlier*, *outlier* dapat memengaruhi K-means dan dapat mengganggu pemisahan klaster yang ideal
 - h) Hanya berlaku untuk atribut numerik
 - i) K-means tidak dapat memberikan informasi secara langsung tentang fitur atau atribut yang paling penting untuk pembentukan klaster

2.2.5 Metode XGBoost

XGBoost merupakan metode yang banyak digunakan dalam pemodelan prediktif, seperti regresi dan klasifikasi. Meskipun XGBoost dapat menghasilkan skor kepentingan atribut, biasanya akan digunakan untuk menentukan atribut yang paling penting dalam pemodelan prediktif, bukan untuk seleksi atribut *clustering* K-means. Namun 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).

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.[16]

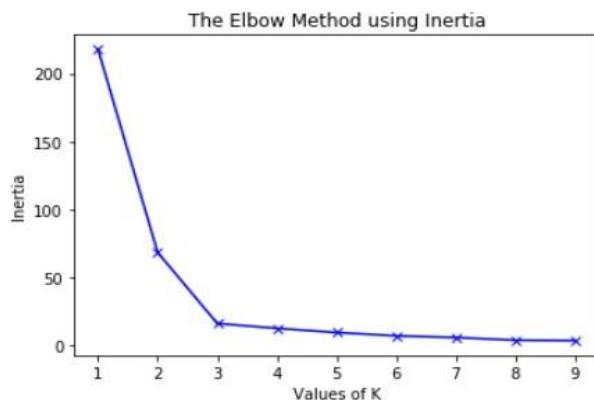
Algoritma XGBoost adalah algoritma yang ditingkatkan berdasarkan *gradient boosting decision tree* dan dapat membangun *boosted trees* secara efisien dan beroprasi secara paralel. XGBoost merupakan salah satu teknik pembelajaran mesin untuk mengatasi permasalahan regresi, rengking dan klasifikasi berdasarkan *Gradient Boosting*. XGBoost pada dasarnya adalah metode ensemble yang didasarkan pada *gradient boosting tree*. Didalam pohon regresi, *nodes* bagian dalam mewakili nilai-nilai untuk tes atribut dan *left nodes* dengan skor

mewakili keputusan. Hasil prediksi adalah jumlah skor yang di prediksi oleh pohon K .[16]

2.2.6 Metode Elbow

Metode siku, atau yang lebih dikenal sebagai metode *elbow* dalam analisis klaster, merupakan suatu pendekatan yang digunakan untuk menentukan jumlah optimal dari klaster dalam suatu dataset. Dalam metode ini, kita menggambarkan kurva antara "*within the sum of squares*" (WSS) dan jumlah klaster yang digunakan. Ketika kurva tersebut diplot, bentuknya menyerupai lengan manusia, dan inilah sebabnya metode ini sering disebut sebagai metode siku.[17]

Pada metode ini untuk menentukan nilai K pada K-means dengan melihat penerunan yang paling signifikan dan kestabilan hasil nilai K tersebut.



Gambar 2.4 Visualisasi Hasil Metode *Elbow*

2.2.7 Davies Bouldin Index (DBI)

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.[18]

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

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

Dimana:

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

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

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

Dimana:

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

2.2.8 Contoh Penerapan Metode XGBoost Pada K-Means *Clustering*

Berikut ini merupakan contoh penerapan perhitungan menggunakan metode yang sesuai dengan penelitian yang akan dilakukan dan diambil dari penelitian sebelumnya oleh Amiruddin Bengga, Rizqiwati Ishak dengan judul “Penerapan Algoritma XGBoost Untuk Seleksi Atribut Pada K-Means Dalam *Clustering* Penerima KIP Kuliah”.

Tabel 2.3 Atribut Dataset KIP Kuliah

| No | Atribut | Nilai Atribut |
|----|----------------|--|
| 1 | Pekerjaan Ayah | <ul style="list-style-type: none"> -Tidak Bekerja -Petani -Nelayan -Peg.Swasta -Wirausaha -Lainnya |
| 2 | Pekerjaan Ibu | <ul style="list-style-type: none"> -Tidak Bekerja -Petani -Wirausaha -Lainnya |

| No | Atribut | Nilai Atribut |
|----|-------------------|---|
| 3 | Penghasilan Ayah | <ul style="list-style-type: none"> -Tidak Berpenghasilan -<Rp.250.000 -Rp.250.001 – Rp.500.000 -Rp.500.001 – Rp.750.000 -Rp.750.001 –Rp.1.000.000 ->Rp. 1.000.000 |
| 4 | Penghasilan Ibu | <ul style="list-style-type: none"> -Tidak Berpenghasilan -<Rp.250.000 -Rp.250.001 – Rp.500.000 -Rp.500.001 – Rp.750.000 -Rp.750.001 –Rp.1.000.000 ->Rp. 1.000.000 |
| 5 | Jumlah Tanggungan | <ul style="list-style-type: none"> -1 – 2 Orang -3 – 4 Orang - >4 Orang |
| 6 | Kepemilikan Rumah | <ul style="list-style-type: none"> -Tidak Memiliki -Menumpang -Sendiri |
| 7 | Sumber Listrik | <ul style="list-style-type: none"> -Tidak Ada -Menumpang Tetangga -PLN -PLN dan Genset -Tenaga Surya |
| 8 | Luas Tanah | <ul style="list-style-type: none"> -<25m² -25 – 50 m² -50 – 99 m² -100 – 200 m² ->200 m² |
| 9 | Luas Bangunan | -<25m ² |

| No | Atribut | Nilai Atribut |
|----|------------|---|
| | | <ul style="list-style-type: none"> -25 – 50 m2 -50 – 99 m2 -100 – 200 m2 ->200 m2 |
| 10 | Sumber Air | <ul style="list-style-type: none"> -Sungai/Mata Air -Sumur -PDAM -Kemasan |
| 11 | MCK | <ul style="list-style-type: none"> -Berbagai Pakai -Kepemilikan Sendiri Diluar -Kepemilikan Sendiri Didalam |
| 12 | Prestasi | <ul style="list-style-type: none"> -Tidak Ada -Tingkat Kabupaten/Kota -Tingkat Nasional -Tingkat Provinsi |

(Sumber: Amiruddin, Resqiwaty Ishak 2023)

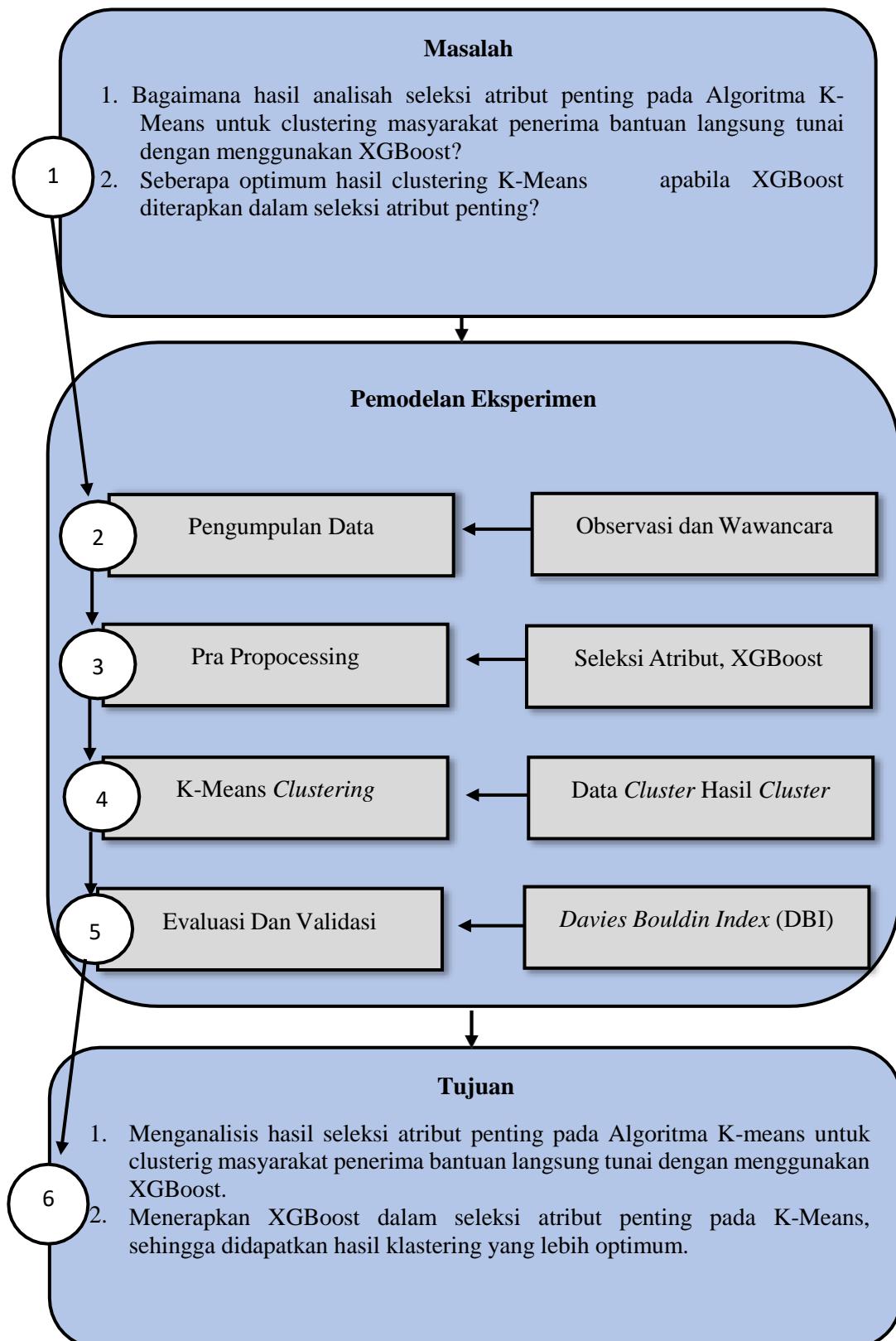
Berdasarkan **Tabel 2.3**, dengan menerapkan metode XGBoost untuk pemilihan atribut pada K-Means dalam *clustering* prioritas penerima bantuan KIP kuliah terbukti dapat menghasilkan *clustering* yang ideal, hal ini berdasarkan hasil pemilihan dari 12 atribut yang dipilih 3 atribut yang paling penting yaitu: (1). Pekerjaan Ayah, (2). Penghasilan Ibu, (3). Luas Bangunan berdasarkan pengukuran evaluasi hasil *clustering*.

2.2.9 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-platform* dan cocok digunakan untuk Data Scientist serta didukung banyak library seperti *Numpy*, *SciPy*, *Pandas*, *Scikit-Learn*, dan *Matplotlib*.

2.3 Kerangka Pikir



BAB III

METODE PENELITIAN

3.1 Jenis Metode, Subjek, Waktu dan Lokasi Penelitian

Dipandang dari tingkat penerapanya, maka penelitian ini merupakan penelitian kualitatif. Penelitian ini menggunakan metode studi kasus. Oleh karena itu data penelitian ini diperoleh dengan wawancara sehingga metode yang digunakan dalam masalah *clustering* penerima bantuan langsung tunai.

Berdasarkan latar belakang dan kerangka pikir seperti yang sudah diuraikan dalam Bab I dan Bab II, 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 (Keperpustakaan)

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.

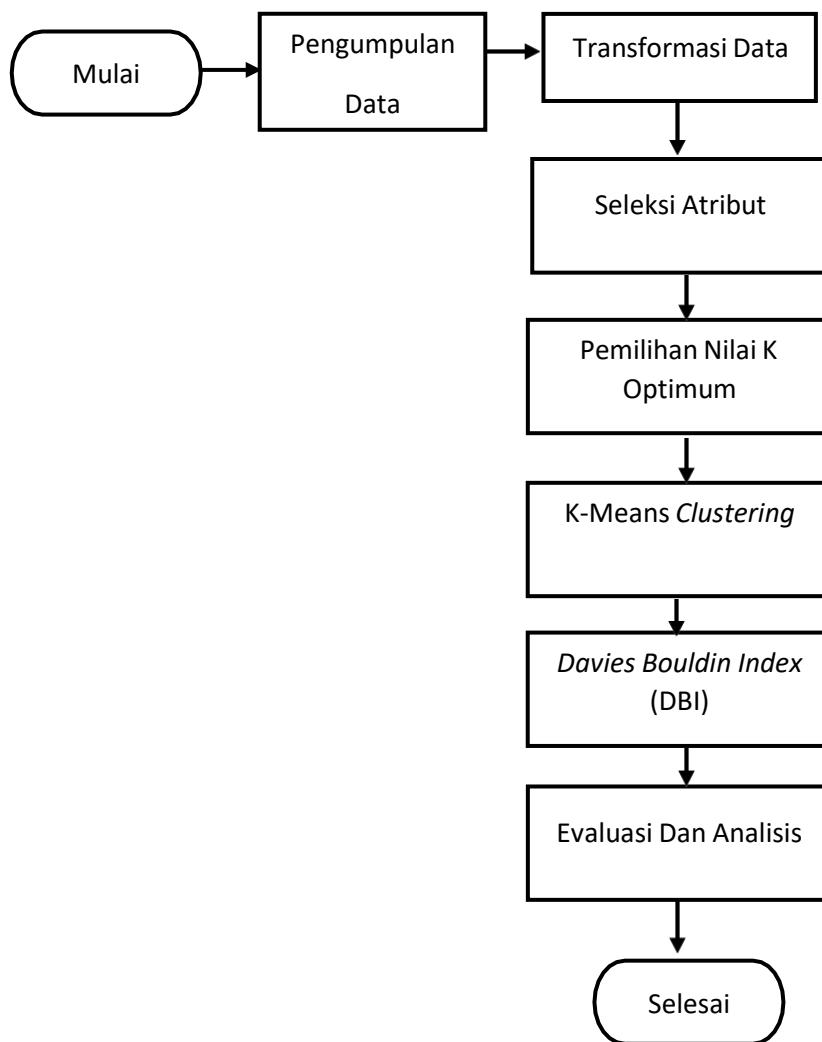
Adapun variabel/atribut dengan tipe datanya masing-masing di tunjukan pada tabel berikut.

Tabel 3.1 Atribut/Variabel Data

| No | Name | Type | Value | Keterangan |
|----|----------------|---------|--|-----------------------|
| 1 | Luas Lantai | Nominal | <8 m ² , 8 – 16 m ² , 16 – 25 m ² , >25m ² | <i>Variabel Input</i> |
| 2 | Lantai Rumah | Ordinal | Tanah, Bambu, Kayu Murah | <i>Variabel Input</i> |
| 3 | Dinding Rumah | Ordinal | Bambu, Rumbia, Kayu Mura, Tembok tanpa Plester | <i>Variabel Input</i> |
| 4 | MCK | Ordinal | Tanpa Fasilitas, Milik Sendiri, Bersama Orang lain | <i>Variabel Input</i> |
| 5 | Sumber Listrik | Ordinal | Tidak Ada, PLN Subsidi, Menumpang Tetangga | <i>Variabel Input</i> |
| 6 | Sumber Air | Ordinal | Sumur, Sungai, Air Hujan, Mata Air Tidak Terlindungi | <i>Variabel Input</i> |
| 7 | Bahan Bakar | Ordinal | Kayu, Arang, Minyak Tanah | <i>Variabel Input</i> |
| 8 | Konsumsi | Ordinal | Daging 1 Kali/Minggu, Susu 1 Kali/Minggu, Ayam 1 Kali/Minggu | <i>Variabel Input</i> |

| No | Name | Type | Value | Keterangan |
|----|-----------------------|---------|---|-----------------------|
| 9 | Pakaian | Nominal | 1 Stel Setahun, >1 Stel Setahun | <i>Variabel Input</i> |
| 10 | Tidak Sanggup Berobat | Ordinal | Puskesmas, Poliklinik | <i>Variabel Input</i> |
| 11 | Sumber Penghasilan KK | Ordinal | Petati Berlahan <500m Buruh Tani, Buru Nelayan, Buru Bangunan, Buru Perkebunan | <i>Variabel Input</i> |
| 12 | Penghasilan KK | Nominal | Tidak Berpenghasilan <Rp.600.000/Bulan Rp.600.000–800.000 Rp. >800.000 | <i>Variabel Input</i> |
| 13 | Pendidikan KK | Ordinal | Tidak Sekolah, Tidak Tamat SD, Tamat SD | <i>Variabel Input</i> |
| 14 | Tabungan | Ordinal | Tidak Ada, Tidak Memiliki Barang Yang Mudah Dijual Minimal Rp.500.000 | <i>Variabel Input</i> |

3.3 Pemodelan



Gambar 3, 1 Model usulan

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 dimasukan kedalam *cluster* hasil untuk menghitung nilai akurasi.

BAB IV

HASIL PENELITIAN

4.1 Hasil Pengumpulan Data

Data yang digunakan pada penelitian ini di ambil dari kantor desa sinorang, dengan jumlah penerima bantuan langsung tunai pada tahun 2022 sebanyak 107.

Tabel 4.1 Dataset Penerima Bantuan Langsung Tunai

| No | Nama Penerima | Luas Lantai | Lantai Rumah | Dinding Rumah | MCK | Sumber Listrik | Sumber Air | bahan bakar |
|-----|---------------|-------------|--------------|----------------------|--------------------|--------------------|------------|--------------|
| 1 | Waser | 8m2 | tanah | tembok tanpa plester | bersama orang lain | pln subsidi | sumur | minyak tanah |
| 2 | Ikram | 12m2 | tanah | kayu murah | milik sendiri | pln subsidi | sumur | arang |
| 3 | Hasdin | <8m2 | kayu murah | kayu murah | milik sendiri | menumpang tetangga | sumur | kayu |
| 4 | Mawar | <8m2 | kayu murah | bambu | milik sendiri | pln subsidi | sumur | minyak tanah |
| 5 | Tani | 11m2 | tanah | tembok tanpa plester | milik sendiri | pln subsidi | sumur | minyak tanah |
| 6 | Rambitan | 8m2 | tanah | tembok tanpa plester | milik sendiri | pln subsidi | sumur | minyak tanah |
| 7 | Siti Khodija | 15m2 | tanah | tembok tanpa plester | milik sendiri | pln subsidi | sumur | minyak tanah |
| 8 | Ruiya | 12m2 | tanah | tembok tanpa plester | milik sendiri | menumpang tetangga | sumur | minyak tanah |
| 9 | Sinayim | 8m2 | tanah | tembok tanpa plester | milik sendiri | menumpang tetangga | sumur | arang |
| 10 | Aisyah | <8m2 | tanah | tembok tanpa plester | milik sendiri | menumpang tetangga | sumur | arang |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 107 | Bejo Prayitno | 11m | tanah | tembok tanpa plester | milik sendiri | Menumpang tetangga | sumur | Minyak tanah |

| No | Komsumsi | Pakaian | Tidak Sanggup Berobat | Sumber Penghasilan KK | Penghasilan KK | Pendidikan KK | Tabungan |
|-----|---------------------|---------|-----------------------|-----------------------|----------------|---------------|-----------|
| 1 | daging 1kali/minggu | 1stel | puskesmas | buru nelayan | rp.600000 | tamat sd | tidak ada |
| 2 | ayam 1kali/minggu | 1stel | puskesmas | buru nelayan | rp.600000 | tamat sd | tidak ada |
| 3 | daging 1kali/minggu | >1stel | puskesmas | buru tani | rp.500000 | tamat sd | tidak ada |
| 4 | daging 1kali/minggu | >1stel | poliklinik | buru perkebunan | rp.500000 | tamat sd | tidak ada |
| 5 | daging 1kali/minggu | >1stel | poliklinik | buru tani | rp.400000 | tamat sd | tidak ada |
| 6 | ayam 1kali/minggu | 1stel | poliklinik | buru tani | rp.500000 | tamat sd | tidak ada |
| 7 | ayam 1kali/minggu | 1stel | puskesmas | buru perkebunan | rp.600000 | tamat sd | tidak ada |
| 8 | daging 1kali/minggu | 1stel | puskesmas | buru perkebunan | rp.400000 | tamat sd | tidak ada |
| 9 | ayam 1kali/minggu | >1stel | poliklinik | buru tani | rp.500000 | tamat sd | tidak ada |
| 10 | ayam 1kali/minggu | >1stel | poliklinik | buru perkebunan | rp.500000 | tamat sd | tidak ada |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 107 | ayam1kali/minggu | >1stel | poliklinik | buru tani | rp.600000 | tamat sd | tidak ada |

4.2 Hasil pemodelan

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

4.2.1 Data Preprocessing

1. Konversi Data Untuk Atribut Luas Lantai

Konversi nilai atribut untuk atribut jenis Luas Lantai di tunjukan pada Tabel 4.2 berikut:

Tabel 4.2 Konversi Atribut Luas Lantai

| No | Nama Atribut | Nilai Atribut |
|----|------------------|---------------|
| 1 | <8m ² | 1 |
| 2 | 8m ² | 2 |
| 3 | 11m ² | 3 |
| 4 | 12m ² | 4 |
| 5 | 15m ² | 5 |

2. Konversi Data Untuk Atribut Lantai Rumah

Konversi nilai atribut untuk atribut jenis Lantai Rumah di tunjukan pada Tabel

4.3 berikut:

Tabel 4.3 Konversi Atribut Lantai Rumah

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

3. Konversi Data Untuk Atribut Dinding Rumah

Konversi nilai atribut untuk atribut jenis Dinding Rumah di tunjukan pada Tabel

4.4 berikut:

Tabel 4.4 Konversi Atribut Dinding Rumah

| No | Nama Atribut | Nilai Atribut |
|----|----------------------|---------------|
| 1 | Tembok Tanpa Plester | 1 |
| 2 | Kayu Murah | 2 |
| 3 | Bambu | 3 |

4. Konversi Data Untuk Atribut MCK

Konversi nilai atribut untuk atribut jenis MCK di tunjukan pada Tabel 4.5 berikut:

Tabel 4.5 Konversi Atribut MCK

| No | Nama Atribut | Nilai Atribut |
|----|--------------------|---------------|
| 1 | Milik Sendiri | 1 |
| 2 | Bersama Orang Lain | 2 |

5. Konversi Data Untuk Atribut Sumber Listrik

Konversi nilai atribut untuk atribut jenis Sumber Listrik di tunjukan pada Tabel 4.6 berikut:

Tabel 4.6 Konversi Atribut Sumber Listrik

| No | Nama Atribut | Nilai Atribut |
|----|--------------------|---------------|
| 1 | Menumpang Tetangga | 1 |
| 2 | PLN Subsidi | 2 |

6. Konversi Data Untuk Atribut Sumber Air

Konversi nilai atribut untuk atribut jenis Bahan Bakar di tunjukan pada Tabel 4.7 berikut:

Tabel 4.7 Konversi Atribut Sumber Air

| No | Nama Atribut | Nilai Atribut |
|----|--------------|---------------|
| 1 | Sumur | 1 |

7. Konversi Data Untuk Atribut Bahan bakar

Konversi nilai atribut untuk atribut jenis Bahan Bakar di tunjukan pada Tabel 4.8 berikut:

Tabel 4.8 Konversi Atribut Bahan Bakar

| No | Nama Atribut | Nilai Atribut |
|----|--------------|---------------|
| 1 | Kayu | 1 |
| 2 | Arang | 2 |
| 3 | Minyak Tanah | 3 |

8. Konversi Data Untuk Atribut Konsumsi

Konversi nilai atribut untuk atribut jenis Konsumsi di tunjukan pada Tabel 4.9 berikut:

Tabel 4.9 Konversi Atribut Konsumsi

| No | Nama Atribut | Nilai Atribut |
|----|---------------------|---------------|
| 1 | Daging 1Kali/Minggu | 1 |
| 2 | Ayam 1kali/Minggu | 2 |

9. Konversi Data Untuk Atribut Pakaian

Konversi nilai atribut untuk atribut jenis Pakaian di tunjukan pada Tabel 4.10 berikut:

Tabel 4.10 Konversi Atribut Pakaian

| No | Nama Atribut | Nilai Atribut |
|----|--------------|---------------|
| 1 | 1Stel | 1 |
| 2 | >1Stel | 2 |

10. Konversi Data Untuk Atribut Tidak Sanggup Berobat

Konversi nilai atribut untuk atribut jenis Tidak Sanggup Berobat di tunjukan pada Tabel 4.11 berikut:

Tabel 4.11 Konversi Atribut Tidak Sanggup Berobat

| No | Nama Atribut | Nilai Atribut |
|----|--------------|---------------|
| 1 | Puskesmas | 1 |
| 2 | Poliklinik | 2 |

11. Konversi Data Untuk Atribut Sumber Penghasilan KK

Konversi nilai atribut untuk atribut jenis Sumber Penghasilan KK di tunjukan pada Tabel 4.12 berikut:

Tabel 4.12 Konversi Atribut Sumber Penghasilan KK

| No | Nama Atribut | Nilai Atribut |
|----|-----------------|---------------|
| 1 | Buru Perkebunan | 1 |
| 2 | Buru Nelayan | 2 |
| 3 | Buru Tani | 3 |

12. Konversi Data Untuk Atribut Penghasilan KK

Konversi nilai atribut untuk atribut jenis Penghasilan KK di tunjukan pada Tabel 4.13 berikut:

Tabel 4.13 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 |

13. Konversi Data Untuk Atribut Pendidikan KK

Konversi nilai atribut untuk atribut jenis Pendidikan KK di tunjukan pada Tabel 4.14 berikut:

Tabel 4.14 Konversi Atribut Pendidikan KK

| No | Nama Atribut | Nilai Atribut |
|----|--------------|---------------|
| 1 | Tamat SD | 1 |

14. Konversi Data Untuk Atribut Tabungan

Konversi nilai atribut untuk atribut jenis Tabungan di tunjukan pada Tabel 4.15 berikut:

Tabel 4.15 Konversi Atribut Tabungan

| No | Nama Atribut | Nilai Atribut |
|----|--------------|---------------|
| 1 | Tidak Ada | 1 |

Setelah dilakukan konversi untuk semua atribut yang bernilai kategorikal, maka hasil akhir dataset yang akan digunakan ditunjukan pada Tabel 4.16

Tabel 4.16 Dataset Akhir Penrima Bantuan Langsung Tunai

| No | Nama Penerima | Luas Lantai | Lantai Rumah | Dinding Rumah | MCK | Sumber Listrik | Sumber Air | bahan bakar |
|-----|---------------|-------------|--------------|---------------|-----|----------------|------------|-------------|
| 1 | Waser | 2 | 1 | 1 | 2 | 2 | 1 | 3 |
| 2 | Ikram | 4 | 1 | 2 | 1 | 2 | 1 | 2 |
| 3 | Hasdin | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 4 | Mawar | 1 | 2 | 3 | 1 | 2 | 1 | 3 |
| 5 | Tani | 3 | 1 | 1 | 1 | 2 | 1 | 3 |
| 6 | Rambitan | 2 | 1 | 1 | 1 | 2 | 1 | 3 |
| 7 | Siti Khodija | 5 | 1 | 1 | 1 | 2 | 1 | 3 |
| 8 | Ruiya | 4 | 1 | 1 | 1 | 1 | 1 | 3 |
| 9 | Sinayim | 2 | 1 | 1 | 1 | 1 | 1 | 2 |
| 10 | Aisyah | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 107 | Bejo Prayitno | 3 | 1 | 1 | 1 | 1 | 1 | 3 |

| No | Konsumsi | Pakaian | Tidak Sanggup Berobat | Sumber Penghasilan KK | Penghasilan KK | Pendidikan KK | Tabungan |
|-----|----------|---------|-----------------------|-----------------------|----------------|---------------|----------|
| 1 | 2 | 1 | 1 | 2 | 5 | 1 | 1 |
| 2 | 2 | 1 | 1 | 2 | 5 | 1 | 1 |
| 3 | 2 | 2 | 1 | 3 | 6 | 1 | 1 |
| 4 | 2 | 2 | 2 | 3 | 6 | 1 | 1 |
| 5 | 2 | 2 | 2 | 3 | 4 | 1 | 1 |
| 6 | 2 | 1 | 2 | 3 | 6 | 1 | 1 |
| 7 | 2 | 1 | 1 | 3 | 5 | 1 | 1 |
| 8 | 2 | 1 | 1 | 3 | 4 | 1 | 1 |
| 9 | 2 | 2 | 2 | 3 | 6 | 1 | 1 |
| 10 | 2 | 2 | 2 | 3 | 6 | 1 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 107 | 2 | 2 | 2 | 3 | 5 | 1 | 1 |

4.2.2 Pemodelan

Penerapan metode XGBoost dalam pemilihan atau seleksi atribut, terlebih dahulu dilakukan *clustering* terhadap dataset pada Tabel 4.16. adapun hasil *clustering* ditunjukan pada Tabel 4.17 berikut.

Tabel 4.17 Hasil *Clustering* Awal

| No | Nama Penerima | Luas Lantai | Lantai Rumah | Dinding Rumah | MCK | Sumber Listrik | Sumber Air | Bahan Bakar |
|-----|---------------|-------------|--------------|---------------|-----|----------------|------------|-------------|
| 1 | Waser | 2 | 1 | 1 | 2 | 2 | 1 | 3 |
| 2 | Ikram | 4 | 1 | 2 | 1 | 2 | 1 | 2 |
| 3 | Hasdin | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 4 | Mawar | 1 | 2 | 3 | 1 | 2 | 1 | 3 |
| 5 | Tani | 3 | 1 | 1 | 1 | 2 | 1 | 3 |
| 6 | Rambitan | 2 | 1 | 1 | 1 | 2 | 1 | 3 |
| 7 | Siti Khodija | 5 | 1 | 1 | 1 | 2 | 1 | 3 |
| 8 | Ruiya | 4 | 1 | 1 | 1 | 1 | 1 | 3 |
| 9 | Sinayim | 2 | 1 | 1 | 1 | 1 | 1 | 2 |
| 10 | Aisyah | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| | | | | | | | | |
| 107 | Bejo Prayitno | 1 | 1 | 1 | 1 | 1 | 1 | 3 |

| No | Konsumsi | Pakaian | Tidak Sanggup Berobat | Sumber Penghasilan KK | Penghasilan KK | Pendidikan KK | Tabungan | Ket |
|-----|----------|---------|-----------------------|-----------------------|----------------|---------------|----------|-------|
| 1 | 2 | 1 | 1 | 2 | 5 | 1 | 1 | C3 |
| 2 | 2 | 1 | 1 | 2 | 5 | 1 | 1 | C2 |
| 3 | 2 | 2 | 1 | 3 | 6 | 1 | 1 | C3 |
| 4 | 2 | 2 | 2 | 3 | 6 | 1 | 1 | C3 |
| 5 | 2 | 2 | 2 | 3 | 4 | 1 | 1 | C4 |
| 6 | 2 | 1 | 2 | 3 | 6 | 1 | 1 | C3 |
| 7 | 2 | 1 | 1 | 3 | 5 | 1 | 1 | C2 |
| 8 | 2 | 1 | 1 | 3 | 4 | 1 | 1 | C4 |
| 9 | 2 | 2 | 2 | 3 | 6 | 1 | 1 | C3 |
| 10 | 2 | 2 | 2 | 3 | 6 | 1 | 1 | C3 |
| ... | | | | | | | | |
| 107 | 2 | 2 | 2 | 3 | 5 | 1 | 1 | C2 |

Proses pengalokasikan data atau *clustering* maka dapat diketahui hasil C1 berjumlah 24, C2 berjumlah 40, C3 berjumlah 21, dan C4 berjumlah 22. Setelah didapatkan hasil *clustering* seperti pada Tabel 4.17 diatas. Selanjutnya dilakukan seleksi atribut. Skor kepentingan dalam model XGBoost dihitung berdasarkan seberapa sering sebuah fitur digunakan untuk membuat keputusan pemisahan di semua pohon keputusan dalam model ensemble, menggunakan rumus *Weight* dengan tahapan sebagai berikut:

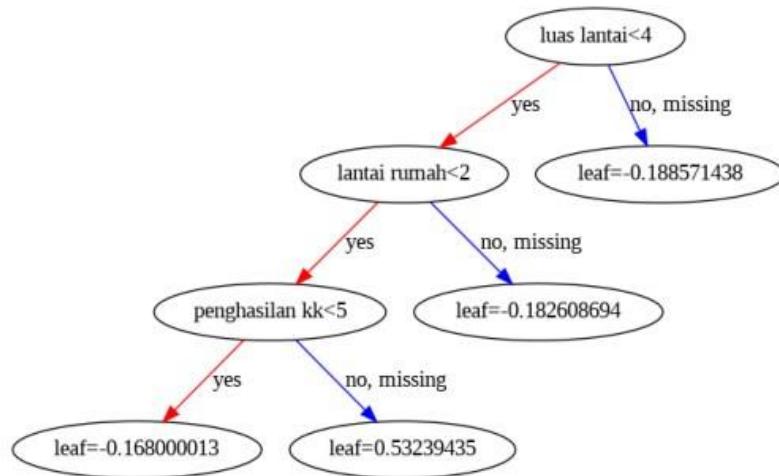
1. *Weight* (Bobot)

Bobot sebuah fitur adalah jumlah total frekunnsional dimana fitur itu muncul dalam keputusan dimana node dalam semua pohon keputusan dalam model, dihitung dengan menggunakan persamaan (2.3).

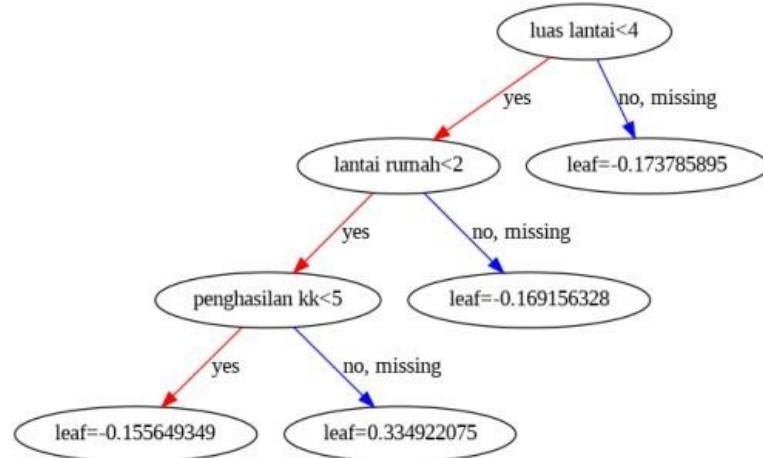
Rumus perhitungannya:

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

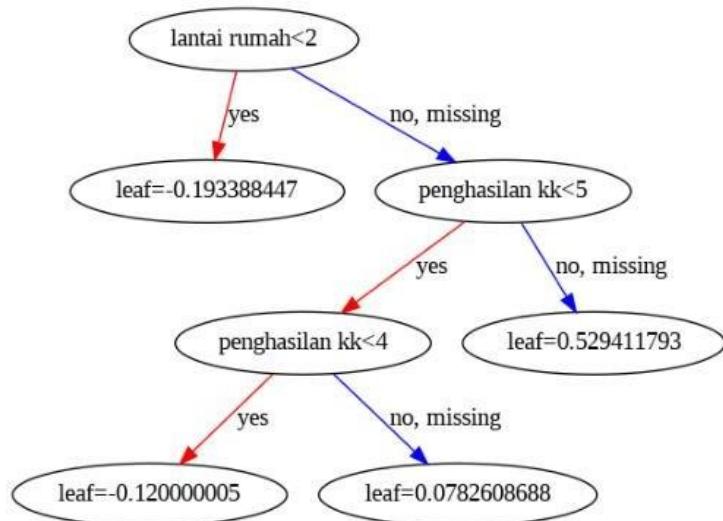
Penetuan skor kepentingan atribut terlebih dahulu harus ditentukan pohon keputusan. Pohon keputusan yang dihasilkan dari metode XGBoost dengan bantuan tools *Python* di tunjukan pada Gambar 4.1 sampai Gambar 4.5.



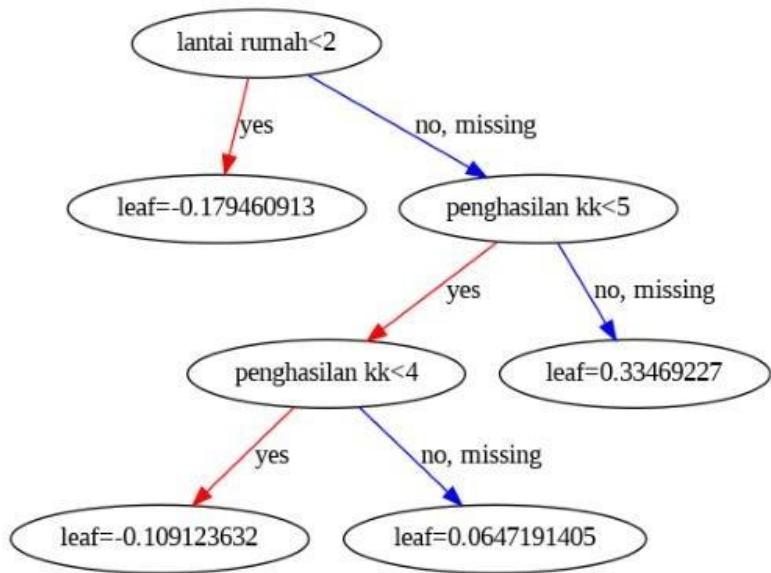
Gambar 4. 1 Sampel Pohon Keputusan 1



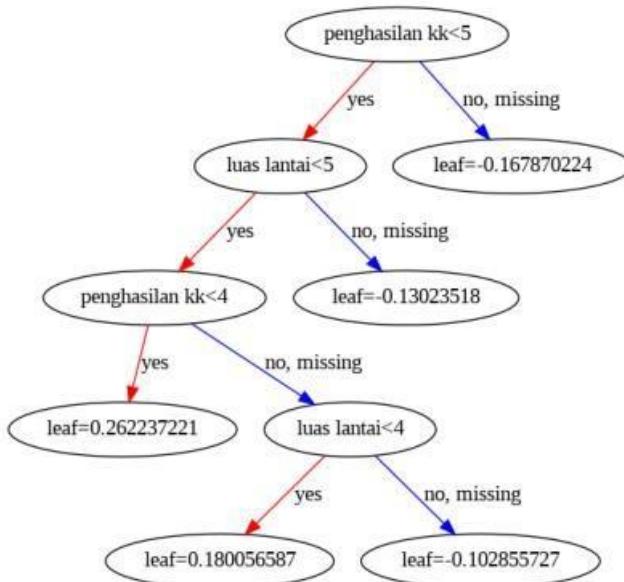
Gambar 4. 2 Sampel Pohon Keputusan 2



Gambar 4. 3 Sampel Pohon Keputusan 3



Gambar 4. 4 Sampel Pohon Keputusan 4



Gambar 4. 5 Sampel Pohon Keputusan 5

Berdasarkan pohon keputusan yang ditunjukan pada Gambar 4.1 sampai Gambar 4.5 dihitung bobot untuk setiap atribut dengan hasil perhitungan ditunjukan pada Tabel 4.18 berikutnya:

Tabel 4.18 Bobot Atribut Pohon Keputusan

| No | Atribut | Bobot Pohon keputusan | | | | | Total |
|----|----------------|-----------------------|---|---|---|---|-------|
| | | 1 | 2 | 3 | 4 | 5 | |
| 1 | Luas lantai | 1 | 1 | 0 | 0 | 2 | 4 |
| 2 | Lantai rumah | 1 | 1 | 1 | 1 | 0 | 4 |
| 3 | Penghasilan KK | 1 | 1 | 2 | 2 | 2 | 8 |

Setelah didapatkan total bobot setiap atribut berikutnya dilakukan normalisasi untuk mendapatkan skor kepentingan atribut dengan cara sebagai berikut:

$$luas\ lantai = \frac{4}{8} = 0,5$$

$$lantai\ rumah = \frac{4}{8} = 0,5$$

$$penghasilan\ KK = \frac{8}{8} = 1$$

Hasil skor kepentingan atribut diatas merupakan sampel 3 atribut dari 14 atribut dan 5 pohon keputusan dari 100 pohon keputusan yang dihasilkan oleh medel XGBoost. Adapun hasil perhitungan secara lengkap ditunjukan pada Tabel 4.19 berikut:

Tabel 4.19 Skor Kepentingan Atribut

| No | Atribut | Skor kepentingan |
|----|-----------------------|-------------------|
| 1 | Luas lantai | 0.23118228 |
| 2 | Lantai rumah | 0.5707228 |
| 3 | Dinding rumah | 0.04406862 |
| 4 | Mck | 0 |
| 5 | Sumber listrik | 0.01568607 |
| 6 | Sumber air | 0 |
| 7 | Bahan bakar | 0.00233176 |
| 8 | Konsumsi | 0 |
| 9 | Pakaian | 0.00304085 |
| 10 | Tidak sanggup berobat | 0.00196366 |
| 11 | Sumber penghasilan KK | 0.0009553 |
| 12 | Penghasilan KK | 0.13004869 |

| No | Atribut | Skor kepentingan |
|----|---------------|------------------|
| 13 | Pendidikan KK | 0 |
| 14 | Tabungan | 0 |

Berdasarkan skor kepentingan atribut pada Tabel 4.19 dipilih 3 atribut yang memiliki skor tertinggi yaitu atribut Luas lantai, Lantai rumah, dan penghasilan KK. Ketiga atribut tersebut yang akan diproses *clustering* selanjutnya.

4.2.3 Pengujian Model

Hasil *clustering* yang optimal dapat diukur menggunakan metode DBI (*davies bouldin index*) adalah metrik evaluasi yang digunakan dalam analisis klaster untuk mengukur kualitas partisi klaster. DBI digunakan untuk membandingkan hasil klaster dari berbagai algoritma klustering atau konfigurasi parameter yang berbeda. Semakin rendah nilai DBI, semakin baik kualitas partisi klaster yang dihasilkan. Oleh karena itu, DBI dapat membantu dalam memilih model klaster yang optimal untuk data yang diberikan.

4.3 Implementasi Hasil Pemodelan

Pada sub-bab ini akan membahas tentang tahapan-tahapan implementasi dari uji coba pemodelan XGBoost dan K-means untuk clustering masyarakat penerima bantuan langsung tunai.

4.3.1 Implementasi Data Preprocessing

Sebelum dilakukan pemodelan ke dalam metode XGBoost dan K-means terlebih dahulu dilakukan beberapa pengolahan data

1. Pseudocode Import Library Python

```

1 import numpy as np
2 import pandas as pd
3 from google.colab import drive
4 from sklearn.cluster import KMeans
5 import xgboost as xgb
6 import matplotlib.pyplot as plt
7 import seaborn as sns
8 import plotly.graph_objs as go
9 from sklearn.metrics import davies_bouldin_score

```

Pseudocode 4.1 Import Liblary Python

Penjelasan mengenai library diatas ditunjukan pada Tabel 4.20.

Tabel 4.20 Penjelasan Psudocode 4.1

| Nama Library | Deskripsi |
|--------------------------|---|
| Pandas | mengimpor pustaka Pandas dengan alias pd. Pandas adalah pustaka yang digunakan untuk manipulasi dan analisis data |
| Numpy | mengimpor pustaka NumPy dengan alias np. NumPy adalah pustaka yang digunakan untuk komputasi numerik dalam <i>Python</i> . |
| <i>Seabron</i> | mengimpor pustaka Seaborn dengan alias sns. Seaborn adalah pustaka yang digunakan untuk visualisasi data statistik. |
| <i>Matplotlib</i> | mengimpor modul <i>pyplot</i> dari pustaka Matplotlib dengan alias plt. <i>Matplotlib</i> adalah pustaka untuk membuat visualisasi grafik dalam <i>Python</i> . |
| <i>Sklearn.cluster</i> | digunakan untuk mengimpor modul atau kelas KMeans dari paket cluster dalam library <i>scikit-learn</i> (sklearn) di <i>Python</i> . |
| XgBoost | digunakan untuk mengimpor modul atau paket XGBoost ke dalam program <i>Python</i> , dengan menggunakan alias xgb untuk memudahkan penggunaan. |
| <i>Plotly.graph objs</i> | digunakan untuk mengimpor modul <i>graph_objs</i> dari paket <i>Plotly</i> ke dalam program <i>Python</i> , dengan menggunakan alias go untuk memudahkan penggunaan, |
| <i>Google colab</i> | mengimpor modul drive dari pustaka <i>google.colab</i> . <i>Google Colab</i> adalah platform cloud yang memungkinkan pengguna untuk menulis dan mengeksekusi kode <i>Python</i> di browser web. |

| Nama Library | Deskripsi |
|-----------------------------|---|
| <i>Davies bouldin score</i> | digunakan untuk mengimpor fungsi <i>davies_bouldin_score</i> dari modul <i>metrics</i> dalam library <i>scikit-learn</i> (<i>sklearn</i>) di <i>Python</i> , untuk mengevaluasi kualitas <i>clustering</i> yang telah Anda lakukan dengan algoritma <i>clustering</i> tertentu. |

2. Pseudocode menghubungkan *Google Colab* dengan *Google Drive*

```

1  from google.colab import drive
2  drive.mount('/content/drive')

```

Pseudocode 4.2 Import Google Colab dengan Google Drive

Kode tersebut berfungsi untuk mengakses *google drive* dari *google colab*, yang merupakan platfrom *cloud* yang memungkinkan anda menulis dan mengeksekusi kode *Python* di browser web.

3. Pseudocode Baca Dataset File Excel

```

1  dfData =
2  pd.read_excel('/content/drive/MyDrive/skripsi/dataset
3  t penelitian.xlsx')

```

Pseudocode 4.3 Baca File Excel

Kode tersebut berfungsi untuk membaca dataset yang tersimpan dalam format dari Go Excel ogle Drive menggunakan pandas, dan menyampaikannya didalam notebook *Google Colab*.

4. Pseudocode Lihat Dataset

```

1  dfData.head()

```

Pseudocode 4.4 Info dataset

Kode tersebut berfungsi untuk yang umum digunakan dalam library Pandas di *Python* untuk menampilkan sejumlah baris pertama dari DataFrame.

5. Pseudocode Konversi Dataset Ke CSV

```

1  dfData.to_csv("/content/drive/MyDrive/skripsi/dataset
2  t penelitian.xlsx", index=False)

```

Pseudocode 4.5 Konversi Dataset ke CSV

Kode tersebut berfungsi untuk mengkonversi dataset yang telah dibaca sebelumnya dari format Excel menjadi format CSV, serta memberikan informasi tentang dataset tersebut.

6. Pseudocode Cek Informasi Dataset

| | |
|---|---------------|
| 1 | dfData.info() |
|---|---------------|

Pseudocode 4.6 Info Dataset

Kode tersebut berfungsi untuk memberikan informasi tentang struktur DataFrame, yang mencakup jumlah baris dan kolom, serta tipe data dalam setiap kolom dan jumlah nilai non-null.

7. Pseudocode Menghitung Banyak Data Dalam Suatu Atribut

| | |
|---|--------------------------------------|
| 1 | dfData['luas_lantai'].value_counts() |
|---|--------------------------------------|

Pseudocode 4.7 Menghitung Banyak Data Dalam Suatu Atribut

Kode tersebut berfungsi untuk menghitung jumlah kemunculan setiap nilai unik dalam suatu kolom DataFrame, dan mengembalikan hasil dalam bentuk objek Series di mana indeksnya adalah nilai unik dari kolom tersebut dan nilainilainya adalah jumlah kemunculan setiap nilai.

8. Pseudocode Konversi Data String Ke Angka (Luas Lantai)

| | |
|---|---|
| 1 | dfData['luas_lantai'].replace(['<8m ² ', '8m ² ', '11m ² ', '12m ² ', '15m ² '], [1, 2, 3, 4, 5], inplace=True) |
|---|---|

Pseudocode 4.8 Konversi Data (Luas lantai) Ke Angka

Kode tersebut berfungsi untuk mengkonversi/mengganti nilai dalam kolom ‘Luas Lantai’ dengan nilai baru yang telah ditentukan.

9. Pseudocode Konversi Data String Ke Angka (Lantai Rumah)

| | |
|---|--|
| 1 | dfData['lantai_rumah'].replace(['tanah', 'kayu murah'], [1, 2], inplace=True) |
|---|--|

Pseudocode 4.9 Konversi Data (Lantai Rumah) Ke Angka

Kode tersebut berfungsi untuk mengonversi/mengganti nilai dalam kolom ‘Lantai Rumah’ dengan nilai baru yang telah di tentukan.

10. Pseudocode Konversi Data String Ke Angka (Dinding Rumah)

```

1 dfData['dinding rumah'].replace(['tembok tanpa
2 plester', 'kayu murah', 'bambu'], [1,2,3], 
3 inplace=True)
4 dfData.head()

```

Pseudocode 4.10 Konversi Data (Dinding Rumah) Ke Angka

Kode tersebut berfungsi untuk mengonversi/mengganti nilai dalam kolom ‘Dinding Rumah’ dengan nilai baru yang telah di tentukan.

11. Pseudocode Konversi Data String Ke Angka (MCK)

```

1 dfData['mck'].replace(['milik sendiri', 'bersama
2 orang lain'], [1,2], inplace=True)
3 dfData.head()

```

Pseudocode 4.11 Konversi Data (MCK) Ke Angka

Kode tersebut berfungsi untuk mengonversi/mengganti nilai dalam kolom ‘MCK’ dengan nilai baru yang telah di tentukan.

12. Pseudocode Konversi Data String Ke Angka (Sumber Listrik)

```

1 dfData['sumber listrik'].replace(['menumpang
2 tetangga', 'pln subsidi'], [1,2], inplace=True)
3 dfData.head()

```

Pseudocode 4.12 Konversi Data (Sumber Listrik) Ke Angka

Kode tersebut berfungsi untuk mengonversi/mengganti nilai dalam kolom ‘Sumber Listrik’ dengan nilai baru yang telah di tentukan.

13. Pseudocode Konversi Data String Ke Angka (Sumber Air)

```

1 dfData['sumber air'].replace(['sumur'], [1], 
2 inplace=True)
3 dfData.head()

```

Pseudocode 4.13 Konversi Data (Sumber Air) Ke Angka

Kode tersebut berfungsi untuk mengkonversi/mengganti nilai dalam kolom ‘Sumber Air’ dengan nilai baru yang telah di tentukan.

14. Pseudocode Konversi Data String Ke Angka (Bahan Bakar)

| | |
|---|---|
| 1 | dfData['bahan bakar'].replace(['kayu', 'arang', 'minyak tanah'], [1, 2, 3], inplace=True) |
| 2 | |
| 3 | dfData.head() |

Pseudocode 4.14 Konversi Data (Bahan Bakar) Ke Angka

Kode tersebut berfungsi untuk mengkonversi/mengganti nilai dalam kolom ‘Bahan Bakar’ dengan nilai baru yang telah di tentukan.

15. Pseudocode Konversi Data String Ke Angka (Konsumsi)

| | |
|---|--|
| 1 | dfData['komsumsi'].replace(['daging 1kali/minggu', 'ayam 1kali/minggu'], [1, 2,], inplace=True) |
| 2 | |
| 3 | dfData.head() |

Pseudocode 4.15 Konversi Data (Konsumsi) Ke Angka

Kode tersebut berfungsi untuk mengkonversi/mengganti nilai dalam kolom ‘Konsumsi’ dengan nilai baru yang telah di tentukan.

16. Pseudocode Konversi Data String Ke Angka (Pakaian)

| | |
|---|--|
| 1 | dfData['pakaian'].replace(['1stel', '>1stel'], [1, 2,], inplace=True) |
| 2 | |
| 3 | dfData.head() |

Pseudocode 4.16 Konversi Data (Pakaian) Ke Angka

Kode tersebut berfungsi untuk mengkonversi/mengganti nilai dalam kolom ‘Pakaian’ dengan nilai baru yang telah di tentukan.

17. Pseudocode Konversi Data String Ke Angka (Tidak Sanggup Berobat)

| | |
|---|--|
| 1 | dfData['tidak sanggup berobat'].replace(['puskesmas', 'poliklinik'], [1, 2,], inplace=True) |
| 2 | |
| 3 | |
| 4 | dfData.head() |

Pseudocode 4.17 Konversi Data (Tidak Sanggup Berobat) Ke Angka

Kode tersebut berfungsi untuk mengonversi/mengganti nilai dalam kolom ‘Tidak Sanggup Berobat’ dengan nilai baru yang telah di tentukan.

18. Pseudocode Konversi Data String Ke Angka (Sumber PenghasilanKK)

```

1 dfData['sumber penghasilan kk'].replace(['buru
2 perkebunan', 'buru nelayan', 'buru tani'], [1,2,3],
3 inplace=True)
4 dfData.head()

```

Pseudocode 4.18 Konversi Data (Sumber Penghasilan KK) Ke Angka

Kode tersebut berfungsi untuk mengonversi/mengganti nilai dalam kolom ‘Sumber Penghasilan KK’ dengan nilai baru yang telah di tentukan.

19. Pseudocode Konversi Data Stirng Ke Angka (Penghasilan KK)

```

1 dfData['penghasilan kk'].replace(['rp.300000',
2 'rp.800000',
3 'rp.700000', 'rp.400000', 'rp.600000', 'rp.500000'], [1
4 ,2,3,4,5,6], inplace=True)
5 dfData.head()

```

Pseudocode 4.19 Konversi Data (Penghasilan KK) Ke Angka

Kode tersebut berfungsi untuk mengonversi/mengganti nilai dalam kolom ‘Penghasilan KK’ dengan nilai baru yang telah di tentukan.

20. Pseudocode Konversi Data String Ke Angka (Pendidikan KK)

```

1 dfData['pendidikan kk'].replace(['tamat sd'], [1],
2 inplace=True)
3 dfData.head()

```

Pseudocode 4.20 Konversi Data (Pendidikan KK) Ke Angka

Kode tersebut berfungsi untuk mengonversi/mengganti nilai dalam kolom ‘Pendidikan KK’ dengan nilai baru yang telah di tentukan.

21. Pseudocode Konversi Data Stirng Ke Angka (Tabungan)

```

1 dfData['tabungan'].replace(['tidak ada'], [1],
2 inplace=True)
3 dfData.head()

```

Pseudocode 4.21 Konversi Data (Tabungan) Ke Angka

Kode tersebut berfungsi untuk mengonversi/mengganti nilai dalam kolom ‘Tabungan’ dengan nilai baru yang telah di tentukan.

4.3.2 Implementasi Pemodelan

1. Pseudocode Pemodelan K-Means Dengan Jumlah *Cluster* Yang Optimum

```

1  model = KMeans(n_clusters = 4, init = "k-means++",
2  max_iter = 300, n_init = 10, random_state = 0)
3  y_clusters = model.fit_predict(x)

```

Pseudocode 4.22 Pemodelan K-Means

Kode tersebut merupakan implementasi algoritma K-Means *clustering* menggunakan pustaka scikit-learn dalam Python. Pertama, kode mengimpor kelas K-Means dari *scikit-learn*. Kemudian, sebuah model K-Means diinisialisasi dengan beberapa parameter: ‘**n_clusters=4**’ menentukan jumlah kluster yang ingin dibentuk, ‘**init="k-means++**’ adalah metode untuk memilih titik awal *centroid* secara efisien, ‘**max_iter=300**’ menetapkan batas maksimum iterasi untuk konvergensi, ‘**n_init=10**’ menentukan jumlah kali algoritma dijalankan dengan inisialisasi berbeda untuk memilih hasil terbaik, dan ‘**random_state=0**’ memastikan hasil konsisten setiap kali kode dijalankan. Setelah itu, metode ‘**fit_predict(x)**’ diterapkan pada data ‘**x**’ untuk menyesuaikan model dan mengelompokkan data, menghasilkan label kluster untuk setiap data. Label ini disimpan dalam ‘**y_clusters**’, yang menunjukkan kluster mana setiap sampel data termasuk.

2. Pseudocode Memilih Kolom Dataset Sesuai Atribut Yang Terpenting

```

1  x = dfData[['lantai rumah','luas
2  lantai','penghasilan kk']].values
3  x

```

Pseudocode 4.23 Memilih Atribut Yang Terpenting

Kode diatas berfungsi untuk mengakses data dari sebuah DataFrame ‘**dfData**’ dan mengekstrak nilai-nilai dari kolom-kolom tertentu ke dalam sebuah *array NumPy*. Pertama, kode ‘**x = dfData [['lantai rumah','luas lantai','penghasilan kk']]**’. Values memilih tiga kolom dari DataFrame: **lantai rumah**, **luas lantai**, dan **penghasilan kk**. Kolom-kolom ini mungkin berisi

informasi tentang jumlah lantai rumah, luas lantai rumah, dan penghasilan kepala keluarga. Kemudian, dengan menggunakan ‘values’, kode ini mengonversi data yang dipilih menjadi sebuah *array NumPy*. Hasil akhirnya disimpan dalam variabel ‘x’, yang sekarang berisi data numerik dari ketiga kolom tersebut dalam bentuk *array* dua dimensi yang siap untuk digunakan dalam analisis atau pemrosesan lebih lanjut, seperti untuk algoritma *machine learning*.

3. Pseudocode Menentukan Jumlah *Cluster* Yang Optimum Dengan Teknik Elbow

```

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

```

Pseudocode 4.24 Teknik Elbow

Kode diatas berfungsi untuk menghitung dan memplot nilai WCSS (*Within-Cluster Sum of Squares*) untuk berbagai jumlah kluster menggunakan algoritma K-Means. Pertama, sebuah list kosong ‘WCSS = []’ dibuat untuk menyimpan nilai WCSS. Kemudian, dengan menggunakan loop ‘for’ dari 1 hingga 10, kode menginisialisasi dan melatih model K-Means untuk setiap jumlah kluster (**n_clusters = i**). Setelah model dilatih, nilai WCSS dari model (‘model.inertia’) ditambahkan ke list ‘WCSS’. Setelah loop selesai, kode membuat plot menggunakan *Matplotlib* untuk menggambarkan bagaimana

WCSS berubah dengan jumlah kluster. Plot ini menampilkan jumlah kluster pada sumbu x dan nilai WCSS pada sumbu y, dengan tujuan membantu menentukan jumlah kluster optimal melalui metode "elbow".

4. Pseudocode Visualisasi Hasil *Cluster* Dengan 3D *Scatterplot* Menggunakan *Matplotlib*

```

1  fig = plt.figure(figsize = (13,13))
2  ax = fig.add_subplot(111, projection='3d')
3
4
5  ax.scatter(x[y_clusters == 0,0],x[y_clusters == 0,1],x[y_clusters == 0,2], s = 50 , color = 'blue', label = "Cluster 0")
6  ax.scatter(x[y_clusters == 1,0],x[y_clusters == 1,1],x[y_clusters == 1,2], s = 50 , color = 'black', label = "Cluster 1")
7  ax.scatter(x[y_clusters == 2,0],x[y_clusters == 2,1],x[y_clusters == 2,2], s = 50 , color = 'green', label = "Cluster 2")
8  ax.scatter(x[y_clusters == 3,0],x[y_clusters == 3,1],x[y_clusters == 3,2], s = 50 , color = 'red', label = "Cluster 3")
9  #ax.scatter(x[y_clusters == 4,0],x[y_clusters == 4,1],x[y_clusters == 4,2], s = 50 , color = 'purple', label = "Cluster 4")
10 ax.scatter(model.cluster_centers_[:,0],model.cluster_centers_[:,1],model.cluster_centers_[:,2], s = 100, c = "yellow", label = "Centroids")
11 ax.set_xlabel('Lantai Rumah')
12 ax.set_ylabel('Luas Lantai')
13 ax.set_zlabel('Penghasilan KK')
14 ax.legend()
15 plt.show()

```

Pseudocode 4.25 Visualisasi 3D Hasil *Cluster*

Kode diatas merupakan visualisasi kluster data dalam plot 3D menggunakan matplotlib. Pertama, sebuah figur berukuran 13x13 inci dibuat dan subplot 3D ditambahkan. Kemudian, data dari empat kluster yang berbeda dipetakan dengan warna biru, hitam, hijau, dan merah. *Centroid* dari kluster tersebut ditampilkan dengan warna kuning dan ukuran yang lebih besar. Label sumbu

untuk "Lantai Rumah", "Luas Lantai", dan "Penghasilan KK" ditambahkan, serta legenda untuk menjelaskan warna kluster dan *centroid*. Akhirnya, plot tersebut ditampilkan dengan '`plt.show()`'.

BAB V

HASIL DAN PEMBAHASAN

5.1 Pembahasan Kinerja Model

Clustering penerima bantuan sosial langsung tunai dengan K-Means terdapat 14 atribut sehingga menyulitkan dalam proses analisis hasil *clustering*, untuk itu menentukan atribut yang paling berpengaruh dalam hasil *clustering* digunakan metode XGBoost. Sebelum dan sesudah memilih atribut yang terbaik dalam proses *clustering* perlu dilakukan evaluasi jumlah cluster yang optimum dengan menggunakan metode *Davies Bouldin Index* (DBI).

5.2 Pembahasan Hasil Pemodelan

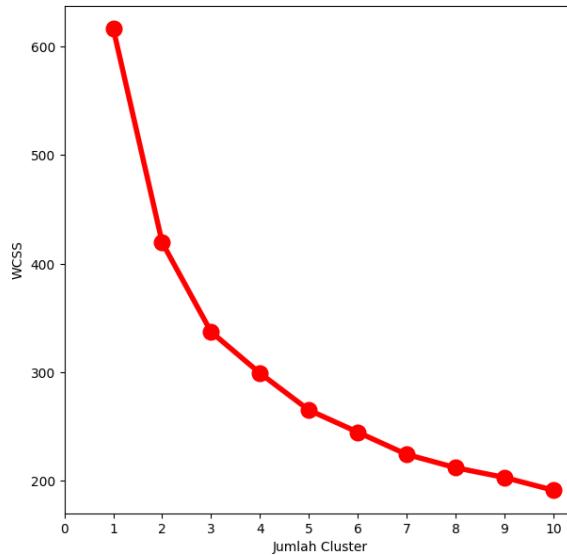
Dalam mencapai kedua tujuan penelitian yaitu menganalisis hasil seleksi atribut penting pada clustering masyarakat penerima bantuan langsung tunai dengan menggunakan XGBoost dan menerapkan XGBoost dalam seleksi atribut penting pada K-Means sehingga didapatkan hasil *clustering* yang lebih optimum, maka perlu diuraikan pembahasan hasil pemodelan.

Tabel 5.1 Perbandingan Hasil *Cluster*

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

5.2.1 Hasil Pemodelan K-Means Sebelum Seleksi Atribut

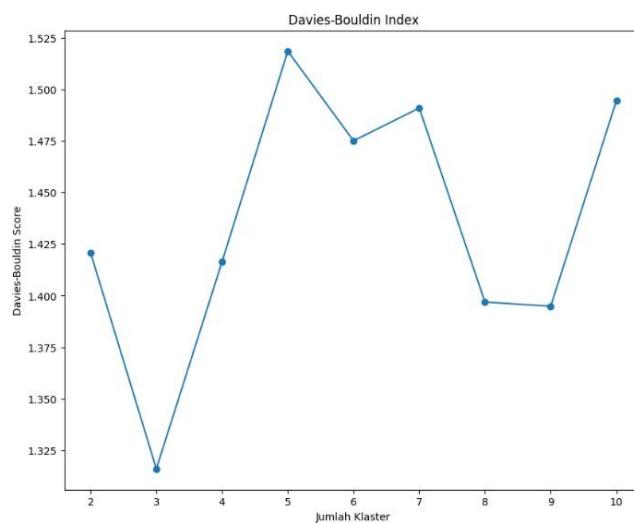
Pemodelan K-Means sebelum Seleksi Atribut perlu dilakukan supaya hasilnya dapat dibandingkan mana hasil clustering yang lebih optimum, apakah sebelum seleksi atribut atau setelah seleksi atribut. Untuk mengetahui hasil *clustering* yang optimum digunakan 2 teknik yaitu teknik *Elbow* yaitu digunakan untuk menentukan jumlah *cluster* yang optimum dalam bentuk grafik *Elbow* dan menggunakan evaluasi hasil *clustering* yaitu DBI. Hasil penentuan jumlah kluster yang optimum dengan teknik *Elbow* ditunjukkan pada Gambar 5. 1 berikut:



Gambar 5.1 Jumlah *Cluster* Dengan Teknik *Elbow*

Berdasarkan Gambar 5.1 diatas, ditunjukkan bahwa jumlah *cluster* yang optimum yaitu sebanyak 3 *cluster* dengan cara melihat posisi titik garis elbow yang penurunannya tidak signifikan lagi.

Setelah didapatkan jumlah *cluster* yang optimum selanjutnya dievaluasi apakah jumlah *cluster* tersebut merupakan jumlah cluster yang optimum dengan menggunakan teknik DBI. Hasil perhitungan DBI untuk masing-masing jumlah *cluster* ditunjukkan pada Gambar 5.2 berikut :



Gambar 5.2 Evaluasi Davies Bouldin Index (DBI) Sebelum Seleksi Atribut

Berdasarkan Gambar 5.2 di atas, ditunjukkan bahwa nilai DBI yang terendah terletak pada jumlah cluster sebanyak 3 *cluster*. Nilai DBI ini menunjukkan bahwa semakin kecil nilai DBI inya, maka hasil clustering semakin baik atau optimum.

5.2.2 Hasil Seleksi Atribut Penting Dengan XGBoost

Setelah dilakukan seleksi atribut penting pada dataset penerima bantuan sebanyak 14 atribut dengan menggunakan metode XGBoost, maka didapatkan hasil skor kepentingan untuk masing-masing atribut yang ditunjukkan pada Gambar 5.3 berikut :

```
[0.23118228 0.5707228 0.04406862 0. 0.01568607 0.
 0.00233176 0. 0.00304085 0.00196366 0.0009553 0.13004869
 0. 0. ]
Atribut paling penting: ['lantai rumah', 'luas lantai', 'penghasilan kk', 'dinding rumah',
 0.00233176 0. 0.00304085 0.00196366 0.0009553 0.13004869
 0. 0. ]
```

Gambar 5.3 Hasil Perhitungan Skor Kepentingan

Berdasarkan Gambar 5.3 diatas hasil proses melalui *tools Python*, agar memudahkan dalam analisis dalam menentukan skor kepentingan masing-masing atribut diuraikan dalam bentuk Tabel seperti tang ditunjukkan pada Tabel 5.1 berikut :

Tabel 5.2 Hasil Skor Kepentingan

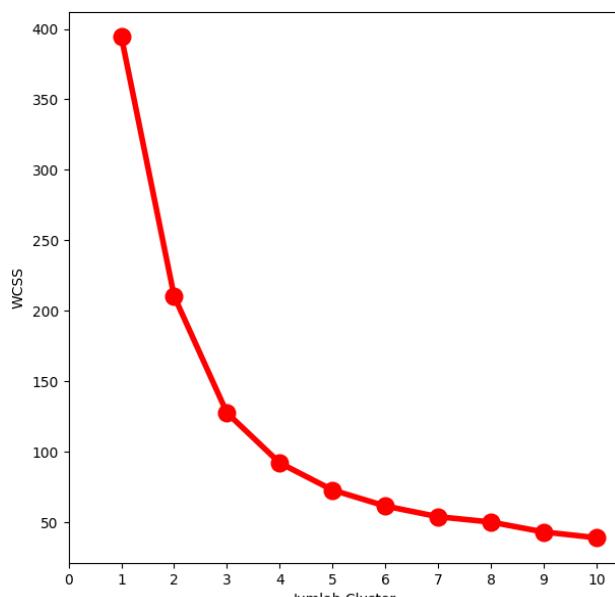
| No | Atribut | Skor Kepentingan |
|----|-----------------------|-------------------|
| 1 | Lantai Rumah | 0.5707228 |
| 2 | Luas Lantai | 0.23118228 |
| 3 | Penghasilan KK | 0.13004869 |
| 4 | Dinding Rumah | 0.04406862 |
| 5 | Sumber Listrik | 0.01568607 |
| 6 | Pakaian | 0.00304085 |
| 7 | Bahan Bakar | 0.00233176 |
| 8 | Tidak Sanggup Berobat | 0.00196366 |
| 9 | Sumber Penghasilan KK | 0.0009553 |
| 10 | Mck | 0 |

| No | Atribut | Skor Kepentingan |
|----|---------------|------------------|
| 11 | Sumber Air | 0 |
| 12 | Konsumsi | 0 |
| 13 | Pendidikan KK | 0 |
| 14 | Tabungan | 0 |

Berdasarkan Tabel 5.1 diatas, dapat dilihat bahwa terdapat 14 atribut yang memiliki pengaruh terhadap hasil *clustering*, namun hanya dipilih 3 atribut yang memiliki skor kepentingan tertinggi yaitu atribut Lantai Rumah dengan skor **0.5707228**, Luas Lantai dengan skor **0.23118228** dan Penghasilan KK dengan skor **0.13004869**. pemilihan ketiga atribut tersebut dalam proses *clustering* untuk memudahkan dalam menganalisis hasil *clustering* pada setiap *cluster*.

5.2.3 Hasil Pemodelan K-Menas Setelah Seleksi Atribut

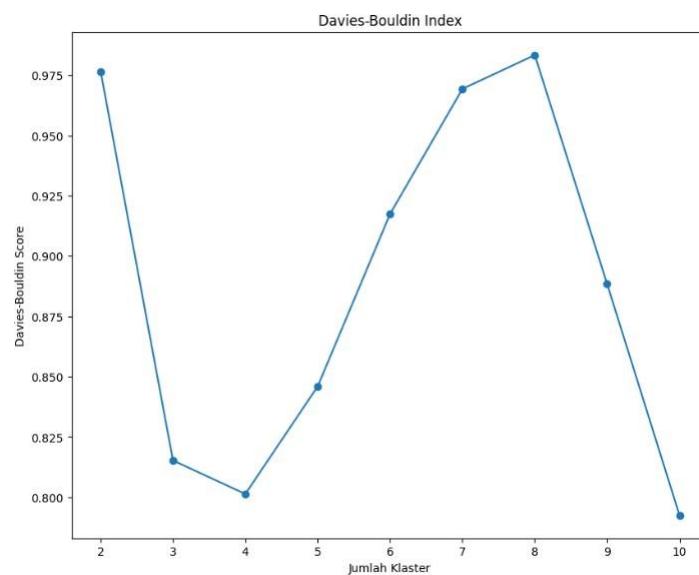
Setelah dilakukan seleksi atribut yang memiliki skor kepentingan tertinggi, dipilih sebanyak 3 atribut terpenting yaitu atribut Lantai Rumah, Luas Lantai dan Penghasilan KK untuk menentukan jumlah cluster yang optimum kembali digunakan teknik *Elbow* dan mengevaluasi hasil *clustering* digunakan teknik DBI. Hasil penentuan jumlah *cluster* setelah seleksi atribut ditunjukkan pada Gambar 5.4 berikut:



Gambar 5.4 Elbow Setelah Seleksi Atribut

Berdasarkan Gambar 5.4 diatas, ditunjukkan bahwa jumlah *cluster* yang optimum setelah seleksi atribut yaitu sebanyak 4 *cluster* dengan cara melihat posisi titik garis *elbow* yang penurunannya tidak signifikan lagi.

Selanjutnya dievaluasi apakah jumlah *cluster* tersebut merupakan jumlah *cluster* yang optimum dengan menggunakan teknik DBI. Hasil perhitungan DBI untuk masing-masing jumlah *cluster* ditunjukkan pada Gambar 5.5 berikut :



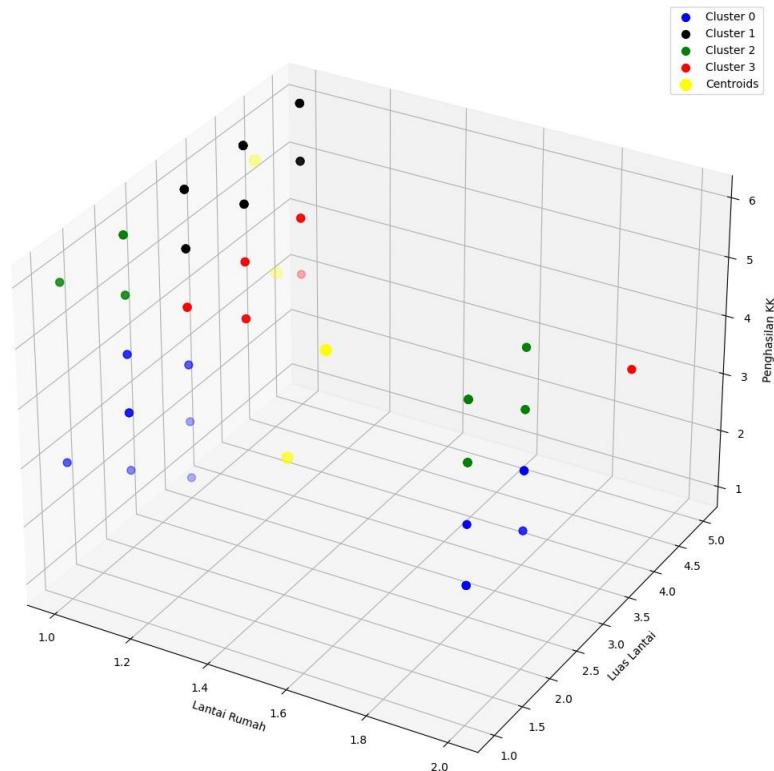
Gambar 5.5 Evaluasi *Davies Bouldin Index* (DBI) Sesudah Seleksi Atribut

Berdasarkan Gambar 5.5 diatas, ditunjukkan bahwa nilai DBI yang terendah terletak pada jumlah *cluster* sebanyak 4 *cluster*. Nilai DBI ini menunjukkan bahwa semakin kecil nilai DBI inya, maka hasil clustering semakin baik atau optimum.

Berdasarkan Gambar 5.4 dan Gambar 5.5 diatas, ditunjukkan bahwa sebelum seleksi atribut didapatkan nilai DBI sebesar 1.325 dan setelah seleksi atribut didapatkan nilai DBI sebesar 0.800 hal ini menunjukkan bahwa terjadi penurunan nilai DBI sebesar 0.525 dengan demikian hasil *clustering* setelah dilakukan seleksi atribut dengan XGBoost didapatkan hasil *clustering* lebih baik atau lebih optimum.

5.2.4 Analisis Hasil Clustering Berdasarkan Atribut Terpenting

Hasil *clustering* berdasarkan tiga atribut terpenting divisualisasikan seperti yang ditunjukkan pada Gambar 5.6 berikut :



Gambar 5.6 Grafik Hasil Visualisasi

Berdasarkan Gambar 5.6 diatas, ditunjukkan bahwa setiap *cluster* sudah dikelompokkan cukup baik namun penyembaran anggota setiap *cluster* masih kurang, hal ini disebakan nilai ketiga atribut yang digunakan nilainya hanya berkisar 1 sampai dengan 6.

Deskripsi statistik untuk masing-masing cluster ditunjukkan pada Tabel 5.2 berikut:

Tabel 5.3 Hasil Deskripsi Statistik

| Cluster | Jumlah Cluster | Lantai Rumah | | | |
|---------|----------------|--------------|----------|-----|-----|
| | | Mean | Std | Min | Max |
| 0 | 24 | 1.250 | 0.442326 | 1 | 2 |
| 1 | 40 | 1.025 | 0.158114 | 1 | 2 |
| 2 | 21 | 1.000 | 0.000000 | 1 | 1 |
| 3 | 22 | 2.000 | 0.000000 | 2 | 2 |
| | | | | | |
| Cluster | Luas Lantai | | | | |
| | Mean | Std | Min | Max | |
| 0 | 2.458333 | 0.977093 | 1 | 4 | |
| 1 | 4.525000 | 0.505736 | 4 | 5 | |
| 2 | 2.238095 | 0.624881 | 1 | 3 | |
| 3 | 1.272727 | 0.455842 | 1 | 2 | |
| | | | | | |
| Cluster | Penghasilan KK | | | | |
| | Mean | Std | Min | Max | |
| 0 | 3.125000 | 0.740887 | 1 | 4 | |
| 1 | 5.100000 | 0.900142 | 3 | 6 | |
| 2 | 5.619048 | 0.497613 | 5 | 6 | |
| 3 | 5.409091 | 0.666125 | 4 | 6 | |

Tabel 5.2 diatas merupakan hasil perhitungan deksripsi statistik dari *tools Python* yang dikonversi ke Tabel. Adapun hasil proses dari *Python* hasilnya ditunjukkan pada Gambar 5.7 berikut :

| Cluster lantai rumah | | | | luas lantai | | | | | |
|----------------------|---|-------|----------|-------------|-----|----------|----------|-----|-----|
| | | mean | std | min | max | mean | std | min | max |
| 0 | 0 | 1.250 | 0.442326 | 1 | 2 | 2.458333 | 0.977093 | 1 | 4 |
| 1 | 1 | 1.025 | 0.158114 | 1 | 2 | 4.525000 | 0.505736 | 4 | 5 |
| 2 | 2 | 1.000 | 0.000000 | 1 | 1 | 2.238095 | 0.624881 | 1 | 3 |
| 3 | 3 | 2.000 | 0.000000 | 2 | 2 | 1.272727 | 0.455842 | 1 | 2 |

| penghasilan kk | | | | count | | | |
|----------------|----------|----------|-----|-------|-----|--|--|
| | | mean | std | min | max | | |
| 0 | 3.125000 | 0.740887 | 1 | 4 | 24 | | |
| 1 | 5.100000 | 0.900142 | 3 | 6 | 40 | | |
| 2 | 5.619048 | 0.497613 | 5 | 6 | 21 | | |
| 3 | 5.409091 | 0.666125 | 4 | 6 | 22 | | |

Gambar 5.7 Deskripsi Statistik Dari *Python*

Berdasarkan hasil deskripsi statistik pada Tabel 5.2 dan Gambar 5.7 di atas dapat analisis masing-masing *cluster* yang ditunjukkan pada Tabel 5.3 berikut :

Tabel 5.4 Analisis Masing-Masing *Cluster*

| Cluster | Hasil Analisis |
|---------|--|
| 0 | <p>Lantai Rumah: Rata-rata jumlah lantai adalah 1.25 dengan standar deviasi 0.442, minimal 1 lantai, dan maksimal 2 lantai.</p> <p>Luas Lantai: Rata-rata luas lantai adalah 2.458 dengan standar deviasi 0.977, minimal 1 dan maksimal 4.</p> <p>Penghasilan KK: Rata-rata penghasilan kepala keluarga (KK) adalah 3.125 dengan standar deviasi 0.741, minimal 1 dan maksimal 4.</p> <p>Jumlah Observasi: 24.</p> |
| 1 | <p>Lantai Rumah: Rata-rata jumlah lantai adalah 1.025 dengan standar deviasi 0.158, minimal 1 lantai, dan maksimal 2 lantai.</p> <p>Luas Lantai: Rata-rata luas lantai adalah 4.525 dengan standar deviasi 0.506, minimal 4 dan maksimal 5.</p> |

| Cluster | Hasil Analisis |
|---------|---|
| | <p>Penghasilan KK: Rata-rata penghasilan kepala keluarga (KK) adalah 5.100 dengan standar deviasi 0.900, minimal 3 dan maksimal 6.</p> <p>Jumlah Observasi: 40</p> |
| 2 | <p>Lantai Rumah: Rata-rata jumlah lantai adalah 1.000 dengan standar deviasi 0.000, minimal dan maksimal 1 lantai.</p> <p>Luas Lantai: Rata-rata luas lantai adalah 2.238 dengan standar deviasi 0.625, minimal 1 dan maksimal 3.</p> <p>Penghasilan KK: Rata-rata penghasilan kepala keluarga (KK) adalah 5.619 dengan standar deviasi 0.498, minimal 5 dan maksimal 6.</p> <p>Jumlah Observasi: 21.</p> |
| 3 | <p>Lantai Rumah: Rata-rata jumlah lantai adalah 2.000 dengan standar deviasi 0.000, minimal dan maksimal 2 lantai.</p> <p>Luas Lantai: Rata-rata luas lantai adalah 1.273 dengan standar deviasi 0.456, minimal 1 dan maksimal 2.</p> <p>Penghasilan KK: Rata-rata penghasilan kepala keluarga (KK) adalah 5.409 dengan standar deviasi 0.666, minimal 4 dan maksimal 6.</p> <p>Jumlah Observasi: 22.</p> |

Hasil *cluster* penerima bantuan langsung tunai dapat dilihat pada Tabel 5.4 berikut:

Tabel 5.5 Hasil Cluster Penerima Bantuan Langsung Tunai

| No | Nama Penerima | Luas Lantai | Lantai Rumah | Dinding Rumah | MCK | Sumber Listrik | Sumber Air | Bahan Bakar |
|-----|---------------|-------------|--------------|---------------|-----|----------------|------------|-------------|
| 1 | Waser | 2 | 1 | 1 | 2 | 2 | 1 | 3 |
| 2 | Ikram | 4 | 1 | 2 | 1 | 2 | 1 | 2 |
| 3 | Hasdin | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 4 | Mawar | 1 | 2 | 3 | 1 | 2 | 1 | 3 |
| 5 | Tani | 3 | 1 | 1 | 1 | 2 | 1 | 3 |
| 6 | Rambitan | 2 | 1 | 1 | 1 | 2 | 1 | 3 |
| 7 | Siti Khodija | 5 | 1 | 1 | 1 | 2 | 1 | 3 |
| 8 | Ruiya | 4 | 1 | 1 | 1 | 1 | 1 | 3 |
| 9 | Sinayim | 2 | 1 | 1 | 1 | 1 | 1 | 2 |
| 10 | Aisyah | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| .. | | | | | .. | | | |
| 107 | Bejo prayitno | 1 | 1 | 1 | 1 | 1 | 1 | 3 |

| No | Komsumsi | Pakaian | Tidak Sanggup Berobat | Sumber Penghasilan KK | Penghasilan KK | Pendidikan kk | Tabungan | ket |
|-----|----------|---------|-----------------------|-----------------------|----------------|---------------|----------|-----|
| 1 | 2 | 1 | 1 | 2 | 5 | 1 | 1 | C3 |
| 2 | 2 | 1 | 1 | 2 | 5 | 1 | 1 | C2 |
| 3 | 2 | 2 | 1 | 3 | 6 | 1 | 1 | C3 |
| 4 | 2 | 2 | 2 | 3 | 6 | 1 | 1 | C3 |
| 5 | 2 | 2 | 2 | 3 | 4 | 1 | 1 | C4 |
| 6 | 2 | 1 | 2 | 3 | 6 | 1 | 1 | C3 |
| 7 | 2 | 1 | 1 | 3 | 5 | 1 | 1 | C2 |
| 8 | 2 | 1 | 1 | 3 | 4 | 1 | 1 | C4 |
| 9 | 2 | 2 | 2 | 3 | 6 | 1 | 1 | C3 |
| 10 | 2 | 2 | 2 | 3 | 6 | 1 | 1 | C3 |
| .. | ... | ... | | | | ... | ... | ... |
| 107 | 2 | 2 | 2 | 3 | 5 | 1 | 1 | C2 |

BAB VI

PENUTUP

6.1 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.

6.2 Saran

Berdasarkan kesimpulan di atas, saran kepada peneliti selanjutnya, yaitu:

1. Penelitian berikutnya dapat menggunakan metode seleksi atrbut yang lainnya agar bisa dibandingkan hasil seleksi atribut terpenting
2. Dapat menggunakan metode *clustering* yang lain untuk bisa dibandingkan hasilnya mana yang lebih baik.

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] Dwitri, N., Tampubolon, J. A., Prayoga, S., Zer, F. I. R., & Hartama, D. (2020). Penerapan algoritma K-Means dalam menentukan tingkat penyebaran pandemi COVID-19 di Indonesia. *JurTI (Jurnal Teknologi Informasi)*, 4(1), 128-132. <https://doi.org/10.36294/jurti.v4i1.12666>
- [3] Bengnga, Amiruddin, and Rezqiwati 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.
- [4] Sari, Dwi Astika, Budi Serasi Ginting, and Yani Maulita. "Pengelompokan Data Penduduk Penerima BSTP (Bantuan Sosial Tunai Pandemic) Menggunakan Metode Algoritma K-means Clustering (Kantor Desa Padang Brahrang)." *Seminar Nasional Informatika (SENATIKA)*. Vol. 6. No. 3. 2022.
- [5] Ufriani, Sari. "Penerapan Algoritma Clustering K-means untuk Menentukan Prioritas Penerima Bantuan Dana Sosial PKH di Kelurahan Kampung Singkep." *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)* 3.1 (2023): 342-350.
- [6] M. Jafar, "Perencanaan Pembangunan Desa," Kementerian Desa, Pembangunan Daerah Tertinggal, Dan Transmigrasi Republik Indonesia, 2015. https://www.dpr.go.id/dokblog/dokumen/F_20150410_7169.pdf (accessed Mar. 28, 2022).
- [7] Apriani, F., & Juleha, T. (2021). The Effectiveness of the Implementation of the Policy of Direct Cash Assistance–Village Funds for the Community in Anggana District, Kutai Kartanegara Regency. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, 4. <https://doi.org/10.33258/birci.v4i4.3439>

- [8] Sari, D. A., Ginting, B. S., & Maulita, Y. (2022, August). Pengelompokan Data Penduduk Penerima BSTP (Bantuan Sosial Tunai Pandemic) Menggunakan Metode Algoritma K-means Clustering (Kantor Desa Padang Brahrang). In *Seminar Nasional Informatika (SENATIKA)* (Vol. 6, No. 3, pp. 326-335).
- [9] Suyanto. (Mei 2017). *DATA MINING Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung.
- [10] Pengertian, Fungsi, Proses dan Tahapan Data Mining. (n.d.). KajianPustaka. Retrieved November 25, 2023.
from <https://www.kajianpustaka.com/2017/09/data-mining.html>
- [11] Marisa, Fitri. "Educational Data Mining (Konsep dan Penerapan)." *Jurnal Teknologi Informasi: Teori, Konsep, dan Implementasi* 4.2 (2013): 90-97.
- [12] Nurahman, Nurahman, and Jetri Susanto. "Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma K-Means." *JURIKOM (Jurnal Riset Komputer)* 10.2 (2023): 461-470.
- [13] 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.
- [14] Rozi Kesuma Dinata, Safwandi, Novia Hasdyna, dan Nur Azizah. "Analisis K-Means Clustering Pada data Sepeda Motor", *Informatics Jurnal ISSN : 2503-250X*, No.1, 2020.
- [15] Kusumawardani, Y., & Hamzah, A. (2018). Perbandingan Metode Clustering Menggunakan Hierarchial Clustering dan Partional Clustering Untuk Mengelompokan Dokumen Berita. *Jurnal Script*, 23-36.
- [16] 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).
- [17] Priya Pedamkar K- Means Clustering Algorithm. (2019, May 15). EDUCBA.
<https://www.educba.com/k-means-clustering-algorithm/>

- [18] 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.

Lampiran 1: Kode Program**Pengolahan Dataset**

```
#Import Liblary Python yang dibutuhkan 1
import pandas as pd

#koneksikan dengan Google Drive 2
from google.colab import drive
drive.mount('/content/drive')

#Baca Dataset file Excel 3
dfData =
pd.read_excel('/content/drive/MyDrive/skripsi/dataset
penelitian.xlsx')

#Lihat Datanya
dfData.head()

#Konversi dataset Excel ke CSV 4
dfData.to_csv("/content/drive/MyDrive/skripsi/dataset
penelitian.xlsx", index=False)

#Cek Informasi dataset
dfData.info()

#Menghitung Banyak Data Dalam suatu Atribut
dfData['luas lantai'].value_counts()

#Konversi data String Ke Angka 5
dfData['luas lantai'].replace(['<8m2', '8m2', '11m2', '12m2',
'15m2'],
[1,2,3,4,5], inplace=True)
dfData.head()

Dst...
```

Implementasi Metode

```

import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
import xgboost as xgb

#baca dataset file excel hasil konversi angka
dfData =
pd.read_excel('/content/drive/MyDrive/skripsi/dataset
angka.xlsx')
dfData.head()

# Menjalankan algoritma K-Means 3
kmeans = KMeans(n_clusters = 4, init = "k-means++", max_iter =
300, n_init = 10, random_state = 0)
#kmeans = KMeans(n_clusters=4)
clusters = kmeans.fit_predict(Xdata)

# Menggabungkan hasil clustering dengan data asli
dfData['Cluster'] = clusters

# Memisahkan atribut dan label kluster
attributes = ['luas lantai', 'lantai rumah', 'dinding
rumah', 'mck', 'sumber listrik',
'sumber air', 'bahan
bakar', 'komsumsi', 'pakaian', 'tidak sanggup berobat',
'sumber penghasilan kk', 'penghasilan
kk', 'pendidikan kk', 'tabungan']
clusters = dfData['Cluster']

# Membangun model XGBoost untuk memperbaiki clustering 4
xgb_model = xgb.XGBClassifier(n_estimators=100, max_depth=4)
xgb_model.fit(dfData[attributes], clusters)

# Mendapatkan skor kepentingan atribut 5
importance_scores = xgb_model.feature_importances_
print(importance_scores)
# Menampilkan atribut yang paling penting 6
important_attributes = [attributes[i] for i in
np.argsort(importance_scores)[:-1]]
print("Atribut paling penting:",
important_attributes, importance_scores)

```

Lampiran 2

 Similarity Report ID: oid:25211:60500572

| | |
|---|---|
| PAPER NAME | AUTHOR |
| SKRIPSI_T3120116_MUHAMMAD_ERDIA NSYAH1.pdf | Muhammad Erdiansyah moherdiansyah 02@gmail.com |
| WORD COUNT | CHARACTER COUNT |
| 12869 Words | 79150 Characters |
| PAGE COUNT | FILE SIZE |
| 81 Pages | 3.3MB |
| SUBMISSION DATE | REPORT DATE |
| Jun 1, 2024 11:25 PM GMT+8 | Jun 1, 2024 11:26 PM GMT+8 |

● 7% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 7% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 1% Submitted Works database

● Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less than 30 words)

Lampiran 3



SURAT REKOMENDASI BEBAS PLAGIASI
No. 104/FIKOM-UIG/R/VI/2024

Yang bertanda tangan di bawah ini :

Nama : Irvan Abraham Salihi, M.Kom
NIDN : 0928028101
Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini menerangkan bahwa :

Nama Mahasiswa : Muhammad Erdiansyah
NIM : T3120116
Program Studi : Teknik Informatika (S1)
Fakultas : Fakultas Ilmu Komputer
Judul Skripsi : Implementasi Metode XGBOOST Dalam Seleksi Atribut Pada Algoritma K-Means Untuk Clustering Masyarakat Penerima Bantuan Langsung Tunai

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi Turnitin untuk judul skripsi di atas diperoleh hasil *Similarity* sebesar 7%, berdasarkan Peraturan Rektor No. 32 Tahun 2019 tentang Pendeteksian Plagiat pada Setiap Karya Ilmiah di Lingkungan Universitas Ihsan Gorontalo dan persyaratan pemberian surat rekomendasi verifikasi calon wisudawan dari LLDIKTI Wil. XVI, bahwa batas kemiripan skripsi maksimal 30%, untuk itu skripsi tersebut di atas dinyatakan **BEBAS PLAGIASI** dan layak untuk diujiankan.

Demikian surat rekomendasi ini dibuat untuk digunakan sebagaimana mestinya.



Gorontalo, 3 Juni 2024
Tim Verifikasi,


Zulfrizanto Y. Lamasigi, M.Kom
NIDN. 0814089101

Terlampir :
Hasil Pengecekan Turnitin

Lampiran 4



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
FAKULTAS ILMU KOMPUTER
UPT. PERPUSTAKAAN FAKULTAS
SK. MENDIKNAS RI NO. 84/D/0/2001**

Jl. Achmad Nadjamuddin No.17 Telp(0435) 829975 Fax. (0435) 829976 Gorontalo

SURAT KETERANGAN BEBAS PUSTAKA

No : 004/Perpustakaan-Fikom/VI/2024

Perpustakaan Fakultas Ilmu Komputer (FIKOM) Universitas Ichsan Gorontalo dengan ini menerangkan bahwa :

Nama Anggota : Muhammad Erdiansyah
No. Induk : T3120116
No. Anggota : M202424

Terhitung mulai hari, tanggal : Senin, 03 Juni 2024, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di pergunakan sebagaimana mestinya.

Gorontalo, 03 Juni 2024

**Mengetahui,
Kepala Perpustakaan**



**Aprilyanto Alhamad, M.Kom
NIDN : 0924048601**

Lampiran 5

PEMERINTAH KABUPATEN BANGGAI
KECAMATAN BATUI SELATAN
DESA SINORANG

SURAT KETERANGAN

Nomor: 141 /29 / DS – SIN / 2024

Yang bertanda tangan dibawah ini:

Nama : Drs. MURSALIM, MA
Jaabatan : Kepala Desa Sinorang
Alamat : Desa Sinorang Kecamatan Batui Selatan

Dengan ini menerangkan bahwa:

Nama : MUHAMMAD ERDIANSYAH
Nim : T3120116
Program Pendidikan : Sarjana (S1)
Program Studi : Teknik Informatika

Bahwa benar yang bersangkutan telah selesai melaksanakan kegiatan penelitian tentang *“Implementasi Metode XGBoost Dalam Seleksi Atribut Pada Algoritma K-Means Untuk Clustering Masyarakat Penerima Bantuan Langsung Tunai”*.

Demikian Surat Keterangan ini dibuat dengan benar dan diberikan kepada yang bersangkutan agar dapat digunakan sebagaimana perlunya.

Dikeluarkan Di : Sinorang

Pada Tanggal 20 Mei 2024

KEPALA DESA SINORANG
Drs. Mursalim, MA

Lampiran 6**RIWAYAT HIDUP MAHASISWA**

Nama : Muhammad Erdiansyah



Nim : T3120116

Tempat, Tanggal Lahir : Sinorang, 01 Januari 2003

Agama : Islam

Email : moherdiansyah02@gmail.com

Riwayat Pendidikan:

1. Tahun 2014, Menyelesaikan Pendidikan Di Sekolah Dasar Negri 2 Sinorang, Kecamatan Batui Selatan Kabupaten Banggai
2. Tahun 2017, Menyelesaikan Pendidikan Di Madrasah Tsanawiyah AL-Muhajirin Sinorang, Kecamatan Batui Selatan Kabupaten Banggai
3. Tahun 2020, Menyelesaikan Pendidikan Di Madrasah Aliyah Al-Muhajirin Sinorang, Kecamatan Batui Selatan Kabupaten Banggai
4. Tahun 2020 Telah Di Terima Menjadi Mahasiswa Perguruan Tinggi Di Universitas Ichsan Gorontalo