

# Penerapan Algoritma XGBoost untuk Seleksi Atribut pada K-Means dalam *Clustering* Penerima KIP Kuliah

Amiruddin  
Program Studi Teknik Informatika  
Universitas Ichsan Gorontalo  
Gorontalo, Indonesia  
amier.76@gmail.com

Rezqiwati Ishak  
Program Studi Teknik Informatika  
Universitas Ichsan Gorontalo  
Gorontalo, Indonesia  
rezqi.uig@gmail.com

Diterima : Juni 2023  
Disetujui : Juli 2023  
Dipublikasi : Juli 2023

**Abstrak**—Pada proses clustering prioritas penerima bantuan Kartu Indonesia Pintar Kuliah dengan algoritma K-Means ada beberapa masalah yang muncul yaitu masalah seleksi atribut yang penting dan penentuan nilai K yang optimum sehingga membuat proses *clustering* tidak maksimal dan tidak ideal. Masalah pemilihan atribut yang penting akan diselesaikan dengan menggunakan algoritma XGBoost yang terbukti dapat digunakan untuk memecahkan masalah seperti pada proses *clustering* prioritas penerima bantuan KIP Kuliah. Hasil penelitian menunjukkan bahwa algoritma XGBoost dapat menentukan 3 (tiga) atribut yang paling penting yaitu Pekerjaan Ayah, Penghasilan Ibu dan Luas Bangunan dari 12 (dua belas) atribut yang ada yaitu Pekerjaan Ayah, Pekerjaan Ibu, Penghasilan Ayah, Penghasilan Ibu, Jumlah Tanggungan, Kepemilikan Rumah, Sumber Listrik, Luas Tanah, Luas Bangunan, Sumber Air, MCK, Prestasi dan metode Elbow terbukti dapat menentukan nilai K yang optimum yaitu nilai K=4. Berdasarkan penggunaan 3 (tiga) atribut terbaik dan nilai K=4 sebagai nilai K optimum berhasil didapatkan *clustering* yang paling maksimal dan ideal dengan nilai index terkecil yaitu 0.819 dengan menggunakan metode pengujian Davies-Bouldin Index.

**Kata Kunci**— XGBoost; K-Means; Elbow; Seleksi Atribut; KIP Kuliah.

**Abstract** — In the process of clustering the priority of the recipient Indonesian smart school cards with the K-Means algorithm, there are several problems that arise, namely the problem of selecting important attributes and determining the optimal value of K, so that the process is not maximum and is not ideal. Important attribute selection problems will be solved using proven XGBoost algorithm that can be used to solve problems such as in the process of clustering the priority of recipients of school KIP assistance. The results of the research showed that the XGBoost algorithm can determine the 3 (three) most important attributes, namely Father's Work, Mother's Production and Building Size from the 12 (twelve) attributes that exist: Father's Job, Mothers' Work, Fathers' Income, Mothers' Revenue, Number of Dependants, Home Ownership, Electrical Resources, Land Area, Building Area, Water Resource, MCK, Performance and Elbow Method proved to determine the optimal K value of K=4. Based on the use of the 3 (three) best attributes and the value of K=4 as the optimal K value, the maximum and ideal clustering with the smallest index value is 0.819 using the Davies-Bouldin Index test method.

**Keywords**— XGBoost; K-Means; Elbow; Attribute Selection; Class KIP.

## I. Pendahuluan

Kartu Indonesia Pintar (KIP) Kuliah merupakan program Pemerintah Indonesia yang diberikan kepada anak-anak kurang mampu melalui Program Indonesia Pintar untuk membantu mereka melanjutkan sekolah hingga Perguruan Tinggi[1]. Karena kuota bantuan yang terbatas khususnya untuk perguruan tinggi swasta, maka perguruan tinggi menghadapi kesulitan dalam menentukan penerima yang paling layak berdasarkan kriteria yang berlaku, untuk itu diperlukan suatu teknik pengelompokan prioritas penerima bantuan tersebut.

Beberapa algoritma yang banyak digunakan dalam pengelompokan atau *clustering* data untuk menyelesaikan masalah tersebut di atas yaitu algoritma *K-Means Clustering*, *Hierarchical Clustering*, *Means Shift Clustering*, DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), *Gaussian Mixture Models* (GMM) dan *Spectral Clustering*[2][3][4]. Masing-masing algoritma tersebut memiliki kelebihan dan kekurangan tergantung dari dimensi data atau atribut yang akan digunakan.

Algoritma K-Means *Clustering* dipilih dalam penelitian ini karena algoritma ini memiliki beberapa kelebihan yaitu (a). Sederhana dan cepat : Algoritma K-Means mudah digunakan dan cepat dalam mengklasifikasikan data, terutama dengan jumlah kluster yang kecil, (b). Skalabilitas : K-Means cocok untuk dataset besar dan berkinerja baik dalam hal skala data, (c). Efisien dalam ruang atribut yang tinggi : K-Means dapat menangani dataset dengan banyak atribut, (d). Interpretasi sederhana: Hasil K-Means dapat dianggap sebagai kluster-kluster terpisah secara intuitif [5]. Algoritma ini juga sudah banyak digunakan peneliti sebelumnya diantaranya dilakukan oleh [6][7][8], namun semua penelitian ini belum menggunakan seleksi atribut hanya memilih beberapa atribut secara manual dari atribut yang digunakan.

Algoritma K-Means memiliki juga kelemahan diantaranya : (a). Sensitif terhadap inisialisasi : K-Means

sangat peka terhadap inisialisasi *centroid* awal, yang dapat menyebabkan berbagai hasil *clustering*. (b). Membutuhkan jumlah kluster yang diketahui : K-Means memerlukan data sebelumnya tentang jumlah kluster yang diharapkan dalam dataset. (c). Sensitif terhadap *outlier* : kehadiran *outlier* dapat memengaruhi K-Means, yang dapat mengganggu pemisahan kluster yang ideal. (d). Hanya berlaku untuk atribut numerik: K-Means hanya dapat digunakan untuk data dengan atribut numerik; tidak dapat digunakan langsung untuk atribut teks atau kategorikal (e). K-Means tidak dapat memberikan informasi secara langsung tentang fitur atau atribut yang paling penting untuk pembentukan kluster [5].

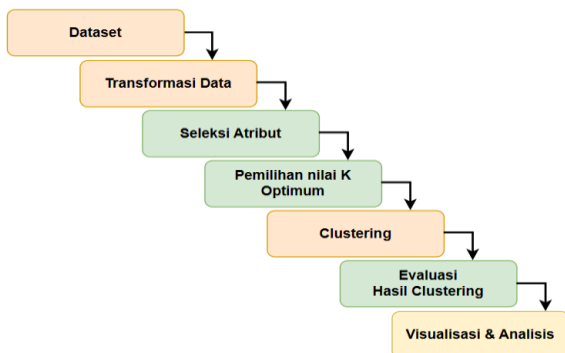
Salah satu kelemahan atau permasalahan pada algoritma K-Means adalah tidak dapat menentukan atribut yang paling penting dalam menghasilkan *clustering* yang ideal, sehingga diperlukan algoritma lain untuk mengatasi masalah tersebut. Ada beberapa algoritma yang bisa digunakan diantaranya : *Silhouette Score*, *Mean Difference* dan *Gradient Boosting (XGBoost)*. Algoritma *Silhouette Score* dan *Mean Difference* tidak dapat secara otomatis menentukan urutan atribut yang penting sedangkan pada XGBoost hal tersebut dapat dilakukan secara otomatis, sehingga pada penelitian menggunakan algoritma XGBoost dalam pemilihan atribut yang penting untuk mendapat hasil *clustering* yang ideal.

Tujuan dari penelitian ini adalah menerapkan metode XGBoost untuk seleksi atribut yang penting pada K-Means dalam *clustering* prioritas penerima Kartu Indonesia Pintar Kuliah sehingga didapatkan hasil *clustering* yang paling ideal.

## II. METODE

### A. Tahapan Penelitian

Pada penelitian ini digunakan tahapan penelitian seperti pada Gambar 1.



Gambar 1. Tahapan Penelitian

1. Dataset : pada tahapan ini dilakukan pengambilan Dataset yaitu Dataset pemohon KIP Kuliah pada Universitas Ichsan Gorontalo Tahun 2022 sebanyak 134 Data
2. Transformasi Data : pada tahapan dilakukan konversi atribut bernilai kategori menjadi numerik, karena algoritma K-Means hanya dapat mengelola data yang bertipe Numerik
3. Seleksi Atribut : pada tahapan ini dilakukan seleksi atribut paling penting dari 12 atribut yang ada menggunakan algoritma XGBoost.

4. Pemilihan nilai K Optimum : pada tahap ini ditentukan nilai K atau jumlah kluster optimum yang akan digunakan pada algoritma K-Means
5. *Clustering* : pada tahap ini dilakukan proses *Clustering* berdasarkan atribut yang paling penting
6. Evaluasi Hasil *Clustering* : pada tahap ini dilakukan Evaluasi hasil *clustering* dengan membanding jumlah atribut dan jumlah kluster yang digunakan, pengujian menggunakan algoritma Davies-Bouldin Index.
7. Visualisasi dan Analisis : pada tahap ini dilakukan visualisasi hasil *clustering* dan analisis untuk menyimpulkan hasil setiap kluster.

### B. Algoritma XGBoost

XGBoost Sebenarnya adalah algoritma yang lebih banyak digunakan dalam pemodelan prediktif, seperti regresi dan klasifikasi. Meskipun XGBoost dapat menghasilkan skor kepentingan atribut, biasanya digunakan untuk menentukan atribut yang paling penting dalam pemodelan prediktif, bukan untuk seleksi atribut untuk *clustering* K-Means[9]. Meskipun demikian, XGBoost dapat digunakan sebagai alat tambahan untuk analisis atribut pada *clustering* K-Means. Berikut tahapan penerapannya[10] :

1. Melakukan *clustering* menggunakan algoritma K-Means untuk membagi data menjadi kluster-kluster tertentu.
2. Menggabungkan hasil *clustering* dengan dataset asli
3. Memisahkan atribut dan label kluster
4. Membangun model XGBoost untuk memperbaiki *clustering*
5. Menghitung skor kepentingan setiap atribut
6. Menampilkan atribut yang paling penting berdasarkan skor kepentingan tertinggi

### C. Metode Elbow

Metode Elbow (siku) adalah salah satu metode yang digunakan untuk menentukan jumlah kluster yang optimal dalam algoritma *clustering* seperti K-Means. Metode ini didasarkan pada visualisasi grafik jumlah kluster dibandingkan variabilitas (*inertia*) dalam *clustering*[11][12][13].

Berikut adalah langkah-langkah dalam metode Elbow:

1. Lakukan *clustering* pada dataset dengan berbagai jumlah kluster yang berbeda.
2. Hitung variabilitas (*inertia*) dari setiap *clustering* yang dihasilkan..
3. Plot grafik jumlah kluster dibandingkan variabilitas.
4. Pada grafik cari titik siku yang menunjukkan titik di mana penurunan variabilitas menurun secara signifikan ketika jumlah kluster meningkat.

### D. K-Means Clustering

K-Means adalah algoritma *clustering* yang populer dan sederhana yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok atau kluster berdasarkan kesamaan atribut. Tujuan utama dari algoritma K-Means adalah untuk meminimalkan jarak antara setiap titik data dengan *centroid* (pusat) kluster yang sesuai[5][2]. Secara umum, langkah-langkah dalam algoritma K-Means adalah sebagai berikut:

Langkah 1: Inisialisasi

- a. Pilih jumlah kluster yang diinginkan

- b. Pilih titik-titik awal sebagai *centroid* kluster secara acak atau menggunakan metode inisialisasi tertentu.

Langkah 2: Assignment

- a. Hitung jarak antara setiap titik data dan *centroid* kluster menggunakan metrik jarak tertentu (misalnya, jarak *Euclidean* atau jarak *Manhattan*).

$$d_{ik} = \sqrt{\sum_{i=1}^n (C_{ij} - C_{kj})^2} \quad (1)$$

Dimana:

$C_{ij}$  = pusat *cluster*

$C_{kj}$  = data

- b. Tentukan kluster terdekat untuk setiap titik dengan memilih *centroid* dengan jarak terkecil.  
c. Setiap titik akan di-assign ke kluster yang sesuai.

Langkah 3: Update Centroid

- a. Hitung rata-rata dari semua titik dalam setiap kluster yang dihasilkan pada langkah sebelumnya.  
b. Gunakan rata-rata tersebut sebagai *centroid* baru untuk masing-masing kluster.

Langkah 4: Iterasi,

Ulangi langkah 2 dan 3 hingga konvergensi tercapai. Konvergensi biasanya terjadi ketika tidak ada perubahan dalam *assignment* kluster atau ketika jumlah iterasi maksimum tercapai

#### E. Davies-Bouldin Index

Davies–Bouldin index (DBI) yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979, adalah metrik evaluasi yang digunakan untuk mengukur kualitas *clustering*. Metrik ini menggabungkan ukuran seberapa baik kluster yang dibentuk dan seberapa kompak kluster tersebut. Semakin rendah nilai DBInya, semakin baik *clustering* yang dihasilkan[11].

- Lakukan *clustering* pada dataset dengan K-Means.
- Hitung *centroid* (pusat) dari setiap kluster.
- Hitung jarak antara setiap pasangan *centroid* dan simpan hasilnya dalam matriks jarak antar-kluster.
- Untuk setiap kluster, hitung nilai  $R(i)$ , yang merupakan rata-rata jarak antara *centroid* kluster tersebut dan semua titik dalam kluster itu sendiri dengan rumus :

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (2)$$

dimana:

- $R_{ij}$  adalah nilai R untuk kluster i dan j.
  - $d_{ij}$  adalah jarak antara *centroid* kluster i dan j.
  - $S_i$  dan  $S_j$  adalah jarak rata-rata antara setiap titik kluster i dan *centroid* dari kluster j
5. Hitung nilai DB, yang merupakan rata-rata dari semua nilai DB untuk setiap kluster. Rumusnya adalah:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{ij,i \neq j} \quad (3)$$

### III. HASIL DAN PEMBAHASAN

Proses *clustering* dilakukan sesuai dengan tahapan penelitian yang diuraikan di atas, mulai dari pengolahan

dataset hingga visualisasi. Ini dilakukan menggunakan *tools Google Colaboratory* dengan bahasa pemrograman Python, dan *library* seperti Numpy, Pandas, Matplotlib, Seaborn, dan Scikit Learn[14]. Dataset yang digunakan sebanyak 134 data sesuai dengan jumlah pemohon KIP Kuliah tahun 2022 pada Universitas Ichsan Gorontalo.

#### A. Pengumpulan Dataset

Dataset yang digunakan merupakan dataset *Private* yaitu data pemohon KIP Kuliah pada Universitas Ichsan Gorontalo[15], Adapun atribut yang digunakan seperti pada Tabel 1 berikut ini :

TABEL 1. ATRIBUT DATASET KIP KULIAH

No	Atribut	Nilai Atribut
1	Pekerjaan Ayah	- Tidak Bekerja - Petani - Nelayan - Peg. Swasta - Wirausaha - Lainnya
2	Pekerjaan Ibu	- Tidak Bekerja - Petani - Wirausaha - Lainnya
3	Penghasilan Ayah	- 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	- 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	- 1-2 Orang - 3-4 Orang - >4 Orang
6	Kepemilikan Rumah	- Tidak Memiliki - Menumpang - Sendiri
7	Sumber Listrik	- Tidak Ada - Menumpang tetangga - PLN - PLN dan Genset - Tenaga Surya
8	Luas Tanah	- < 25 M2 - 25-50 M2 - 50-99 M2 - 100-200 M2 - >200 M2
9	Luas Bangunan	- < 25 M2 - 25-50 M2 - 50-99 M2 - 100-200 M2 - >200 M2
10	Sumber Air	- Sungai/Mata Air - Sumur - PDAM - Kemasan
11	MCK	- Berbagi Pakai - Kepemilikan Sendiri Diluar - Kepemilikan Sendiri Didalam
12	Prestasi	- Tidak Ada - Tingkat Kabupaten/Kota - Tingkat Nasional - Tingkat Provinsi

## B. Transformasi Data

Sebelum dilakukan transformasi data atribut pada Tabel 1 di atas, maka dataset sudah harus bersih data kosong, tidak konsisten dan lainnya. Proses transformasi yaitu melakukan konversi nilai atribut dari nilai kategori ke nilai numerik karena pada algoritma K-Means hanya bisa mengelola data berjenis numerik. Berikut pada Tabel 2 contoh transformasi nilai atribut pada salah satu atribut yang digunakan.

TABEL 2. SAMPEL TRANSFORMASI NILAI ATRIBUT

Nilai Atribut	Nilai Konversi
Tidak Bekerja	1
Petani	2
Nelayan	3
Peg. Swasta	4
Wirasaha	5
Lainnya	6

Semua atribut pada Tabel 1 di atas dilakukan konversi nilai seperti pada Tabel 2 di atas agar bisa diproses dengan algoritma K-Means *Clustering*.

## C. Seleksi Atribut

Setelah semua atribut dikonversi ke nilai numerik, tahap selanjutnya dilakukan seleksi atribut yang paling penting dengan menggunakan Algoritma XGBoost. Adapun nilai skor untuk masing-masing atribut setelah dilakukan proses seleksi atribut melalui *tools* Python didapatkan hasil pada Tabel 3.

TABEL 3. HASIL SELEKSI ATRIBUT

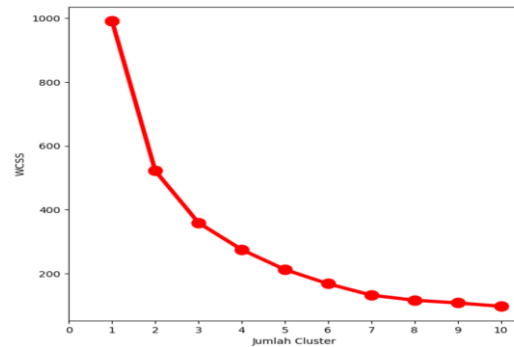
No	Atribut	Skor Nilai
1	<b>Pekerjaan Ayah</b>	<b>0.24280505</b>
2	Pekerjaan Ibu	0.04341467
3	Penghasilan Ayah	0.04936758
4	<b>Penghasilan Ibu</b>	<b>0.193563</b>
5	Jumlah Tanggungan	0.02889493
6	Kepemilikan Rumah	0.01300715
7	Sumber Listrik	0
8	Luas Tanah	0.12334969
9	<b>Luas Bangunan</b>	<b>0.21694902</b>
10	Sumber Air	0.02042621
11	MCK	0.01916789
12	Prestasi	0.04905481

Berdasarkan skor Nilai pada Tabel 3, dipilih 3 (tiga) atribut yang paling tinggi skornya yaitu : Pekerjaan Ayah, Penghasilan Ibu dan Luas Bangunan sebagai atribut yang penting dalam menghasilkan klasster yang ideal. Pemilihan hanya 3 (tiga) atribut yang digunakan supaya memudahkan dalam melakukan analisis dan visualisasi hasil *clustering*.

## D. Pemilihan Nilai K Optimum

Sebelum dilakukan proses pemodelan dengan K-Means, terlebih dahulu dilakukan perhitungan untuk menentukan

nilai K atau jumlah cluster yang optimum dengan menggunakan metode Elbow. Berikut grafik Elbow yang dihasilkan setelah diproses melalui *tools* Python seperti pada Gambar 2.



Gambar 2. Visualisasi Hasil Metode Elbow

Berdasarkan gambar 2 di atas, dapat dilihat bahwa posisi jumlah cluster yang optimum berada pada nilai 4 (empat) cluster karena dengan melihat titik pada graph dimana penurunan inersia tidak lagi signifikan setelah titik 4 (empat), sehingga ini dijadikan sebagai nilai K yang optimum dalam pemodelan K-Means.

## E. Proses Clustering dengan K-Means

Setelah ditentukan atribut dan nilai K (jumlah klaster) yang akan digunakan, maka tahap selanjutnya adalah melakukan pemodelan dengan algoritma K-Means. Proses pemodelan ini menggunakan *tools* Python dengan menggunakan *library* KMeans dari Sklearn. Adapun hasil pemodelan dengan menggunakan 4 (empat) kluster dapat dilihat hasilnya pada Tabel 4.

TABEL 4. HASIL CLUSTERING

Cluster	Titik Centroids			Jumlah Data
	Pekerjaan Ayah	Penghasilan Ibu	Luas Bangunan	
1	2.42	5.33	3.91	12
2	2.03	1.79	1.29	65
3	5.80	1.42	2.62	42
4	2.20	1.46	4.20	15

## F. Evaluasi Hasil Clustering

Hasil *clustering* dengan menggunakan algoritma K-Means dapat diukur atau dievaluasi hasil jumlah klaster yang dihasilkan. Adapun Teknik evaluasi yang digunakan adalah menggunakan Davies-Bouldin Index, dimana pada metode ini akan mengukur sejauh mana klaster berbeda satu sama lain. Semakin rendah nilai indeksnya, maka hasil klaster semakin baik. Hasil Eksprimen dapat dilihat pada Tabel 5.

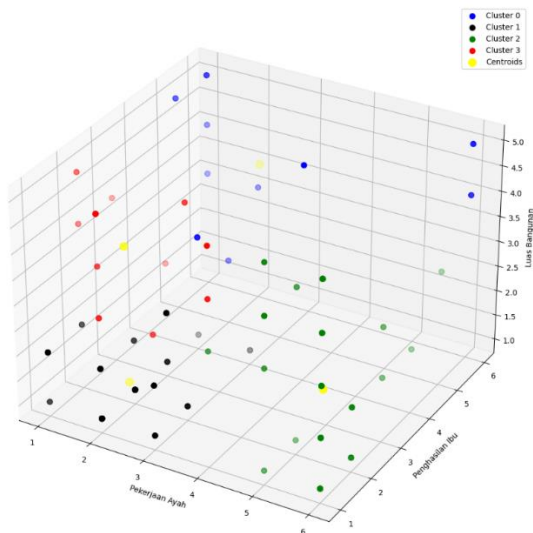
TABEL 5. HASIL PENGUJIAN

Jumlah Atribut	Jumlah Cluster			
	2	3	4	5
12	1.638	1.538	1.461	1.499
6	1.205	1.334	1.209	1.345
3 terbaik	0.891	0.959	<b>0.819</b>	0.836

Berdasarkan Tabel 5, dapat dilihat bahwa nilai index terkecil yaitu 0,819 yang didapat dari penggunaan 3 (tiga) terbaik berdasarkan hasil seleksi atribut terpenting dan jumlah cluster optimum yaitu 4 (empat) berdasarkan perhitungan nilai K optimum.

### G. Visualisasi dan Analisis Hasil Clustering

Setelah didapatkan hasil pengukuran *clustering* yang paling ideal, maka selanjutnya dibuatkan visualisasi dalam bentuk grafik 3 (tiga) dimensi agar dapat dilakukan analisis setiap kelompok klaster yang dihasilkan. Hasil visualisasi dapat dilihat pada Gambar 3.



Gambar 3. Visualisasi Hasil Clustering

Berdasarkan Gambar 3, dapat dilihat bahwa pembagian kelompok atau klaster dengan jelas bisa dibedakan antara cluster satu dengan cluster lainnya.

Berdasarkan Gambar 3 dan Tabel 4, dapat dilakukan analisis hasil *clustering* yaitu (a). klaster 1 : merupakan kelompok prioritas ke empat penerima bantuan KIP Kuliah, (b). klaster 2 : merupakan kelompok prioritas pertama penerima bantuan KIP Kuliah, (c). klaster 3 : merupakan kelompok prioritas ke dua penerima bantuan KIP Kuliah dan ((d). klaster 4 : merupakan kelompok prioritas ke tiga penerima bantuan KIP Kuliah.

## IV. KESIMPULAN

Berdasarkan hasil penelitian yang sudah diuraikan di atas, maka dapat disimpulkan bahwa dengan menerapkan algoritma 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 pemilihan dari 12 (dua belas) atribut yang ada dipilih 3 (tiga) atribut yang paling penting yaitu (1). Pekerjaan Ayah, (2). Penghasilan Ibu, (3). Luas Bangunan dan berdasarkan pengukuran evaluasi hasil *clustering* didapatkan nilai index terkecil yaitu 0.819 dari

penggunaan 3 (tiga) atribut terbaik dan nilai K (jumlah klaster)=4 sebagai nilai K optimum.

## REFERENSI

- [1] Puslapdik, *Pedoman Pendaftaran Kartu Indonesia Pintar Kuliah (KIP Kuliah)*. Jakarta: Puslapdik, 2021.
- [2] "Clustering," *scikit-learn developers*. <https://scikit-learn.org/stable/modules/clustering.html#> (accessed May 20, 2023).
- [3] "Algoritma Clustering Data Science Terupdate 2022." <https://dqlab.id/algoritma-clustering-data-science-terupdate-2022> (accessed May 25, 2023).
- [4] D. T. Larose, *Data Mining Methods and Models*. 2006.
- [5] R. Primartha, *Algoritma Machine Learning*. Bandung: Informatika, 2021.
- [6] D. T. Yuliana, M. I. A. Fathoni, and N. Kurniawati, "Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah dengan Menggunakan Metode K-Means Clustering," *Focus ACTION Res. Math.*, vol. 5, no. 1, pp. 127–141, 2022, doi: 10.30762/f.
- [7] M. S. Sompia and R. Ishak, "Clustering Tingkat Ekonomi Mahasiswa Calon Penerima Kartu Indonesia Pintar ( KIP ) Kuliah Metode K-Means," *J. BALOK*, vol. 1, no. 2, pp. 65–71, 2022.
- [8] F. Nuraeni, D. Kurniadi, and G. F. Dermawan, "Pemetaan Karakteristik Mahasiswa Penerima Kartu Indonesia Pintar Kuliah ( KIP-K ) menggunakan Algoritma K-Means ++," *J. SISFOKOM*, vol. 11, pp. 437–443, 2023.
- [9] T. Chen and C. Guestrin, "XGBoost : A Scalable Tree Boosting System," 2016.
- [10] "XGBoost Documentation," *xgboost developers*. <https://xgboost.readthedocs.io/en/latest/> (accessed May 15, 2023).
- [11] "Clustering Performance Evaluation," *scikit-learn developers*. <https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation> (accessed May 20, 2023).
- [12] A. B. H. Kiat, Y. Azhar, and V. Rahmayanti, "Penerapan Metode K-Means Dengan Metode Elbow Untuk Segmentasi Pelanggan Menggunakan Model RFM (Recency, Frequency & Monetary)," *Repository*, vol. 2, no. 7, pp. 945–952, 2020.
- [13] R. Ishak and Amiruddin, "Clustering Tingkat Pemahaman Dasar Mahasiswa Pada Pra-Perkuliahan Probabilitas Statistika Dengan Metode K-Means," *Jambura J. Electr. Electron. Eng.*, vol. 4, pp. 65–69, 2022, doi: 10.37905/jjee.v4i1.11997.
- [14] T. Wahyono, *Fundamental of Python for Machine Learning*. Yogyakarta: Gava Media, 2018.
- [15] BAAK UNISAN, "Dataset Pemohon KIP Kuliah," Gorontalo, 2023.