



SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi,
dan Teknik Informatika

<https://ejurnal.itats.ac.id/snestik> dan <https://snestik.itats.ac.id>



Informasi Pelaksanaan :

SNESTIK V - Surabaya, 26 April 2025

Fakultas Teknik Elektro Dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:

DOI : 10.31284/p.snestik.2025.7644

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Penerapan Algoritma Agglomerative Hierarchical Untuk Clustering Peringkat Prioritas Penerima Bantuan Rumah

Katrina Yuliana Kartika, J.B. Budi Darmawan

Universitas Sanata Dharma

e-mail: inekartika6@gmail.com

ABSTRACT

Poverty remains a serious issue in Indonesia, particularly in rural and underdeveloped areas. Southwest Sumba Regency is one of the regions that continues to face poverty-related challenges. To address this issue, the Indonesian government has implemented several social assistance programs, including housing assistance. However, determining the priority of housing assistance recipients remains a challenge, as the process is still manual and may result in families who should be prioritized not receiving the aid. Using data of proposed housing assistance recipients, the Agglomerative Hierarchical Clustering algorithm will be applied to group recipients based on priority. This approach aims to help the Southwest Sumba Regency government improve the efficiency of housing assistance distribution, ensuring that support is accurately targeted to those in real need. The results of this study show that the Agglomerative Hierarchical Clustering method provides effective outcomes, supported by a good silhouette coefficient. The best validation result of this method on 2022 data, normalized using the Z-score method, occurs when $k=2$, with a silhouette coefficient value of 0.79 using the Average Linkage approach. The findings show that cluster 0, containing 1,116 objects, should be prioritized first, while cluster 1, with 95 objects, becomes the second priority.

Keywords: clustering; PCA; silhouette coefficient; agglomerative hierarchical clustering.

ABSTRAK

Kemiskinan masih menjadi masalah serius di Indonesia, terutama di daerah pedesaan dan tertinggal. Salah satu daerah yang masih menghadapi permasalahan terkait kemiskinan adalah Kabupaten Sumba Barat Daya.

Untuk mengatasi masalah tersebut, pemerintah Indonesia telah menerapkan berbagai program bantuan sosial, salah satunya adalah program bantuan rumah. Namun, dalam pelaksanaannya, penentuan prioritas penerima bantuan rumah sering kali menjadi tantangan karena masih dilakukan secara manual, sehingga keluarga yang seharusnya diprioritaskan terkadang tidak mendapatkan bantuan. Dengan memanfaatkan data usulan calon penerima bantuan rumah, algoritma Agglomerative Hierarchical Clustering akan diterapkan untuk mengelompokkan peringkat prioritas penerima bantuan. Pendekatan ini bertujuan membantu Pemerintah Kabupaten Sumba Barat Daya meningkatkan efisiensi penyaluran program bantuan rumah agar lebih tepat sasaran dan menysasar masyarakat yang benar-benar membutuhkan. Hasil penelitian menunjukkan bahwa metode Agglomerative Hierarchical Clustering menghasilkan keluaran yang efektif dengan nilai silhouette coefficient yang baik. Hasil validasi terbaik dari metode ini terhadap data tahun 2022 yang telah dinormalisasi menggunakan metode Z-score terjadi saat $k = 2$, dengan nilai silhouette coefficient sebesar 0,79 menggunakan pendekatan Average Linkage. Penelitian ini menunjukkan bahwa cluster 0, yang terdiri dari 1.116 obyek, menjadi prioritas pertama, sedangkan cluster 1, yang berjumlah 95 obyek, ditetapkan sebagai prioritas kedua.

Kata kunci: clustering; PCA; silhouette coefficient; agglomerative hierarchical clustering.

PENDAHULUAN

Salah satu permasalahan yang masih dihadapi oleh negara Indonesia adalah kemiskinan. Masalah kemiskinan merupakan masalah yang kompleks dan bersifat tidak hanya dapat diukur dari satu aspek tunggal sehingga menjadi prioritas pembangunan [1]. Kemiskinan di Indonesia menghantui sebagian besar penduduk, khususnya di daerah pedesaan dan tertinggal. Provinsi Nusa Tenggara Timur (NTT) adalah salah satu contoh wilayah yang masih menghadapi tantangan serius terkait kemiskinan. Tingkat kemiskinan yang tinggi menyebabkan wilayah ini terus menderita. Daerah tandus dan gersang di NTT menyebabkan kekeringan dan kekurangan pangan menjadi masalah rutin, ditambah dengan isu putus sekolah, gizi buruk, dan kurangnya fasilitas kesehatan serta air bersih pada periode 2016-2018 [2].

Salah satu kabupaten di NTT yang masih menghadapi permasalahan serius terkait kemiskinan adalah Kabupaten Sumba Barat Daya. Pada tahun 2022, Badan Pusat Statistik Nusa Tenggara Timur melaporkan bahwa jumlah penduduk miskin di kabupaten ini menempati peringkat kedua dengan jumlah penduduk miskin sebanyak 98.500 orang. Untuk mengatasi masalah kemiskinan, pemerintah daerah telah menerapkan berbagai program bantuan sosial. Salah satunya adalah program Bantuan Stimulan Perumahan Swadaya (BSPS) yang ditujukan bagi masyarakat berpenghasilan rendah [3]. Program ini bertujuan membantu keluarga kurang mampu yang tidak memiliki rumah layak huni, dengan menyediakan tempat tinggal yang layak untuk meningkatkan kualitas hidup dan melindungi dari lingkungan tidak sehat.

Program bantuan rumah adalah strategi pemerintah untuk mengurangi kemiskinan dan meningkatkan kesejahteraan. Namun, penentuan prioritas penerima sering menjadi tantangan karena dilakukan secara manual. Untuk mengatasi ini, diperlukan pengelompokan dengan bantuan Machine Learning. Terdapat penelitian terkait *clustering* yang menggunakan algoritma Agglomerative Hierarchical Clustering untuk pengelompokan titik api di provinsi Jambi memperoleh *cluster* terbaik dengan nilai *silhouette coefficient* sebesar 0,585644 dengan 2 *cluster* [4]. Penelitian lain menunjukkan bahwa algoritma Hierarchical dengan dua *cluster* dan nilai *silhouette* 0.944473 adalah yang terbaik untuk data *marketplace* elektronik [5].

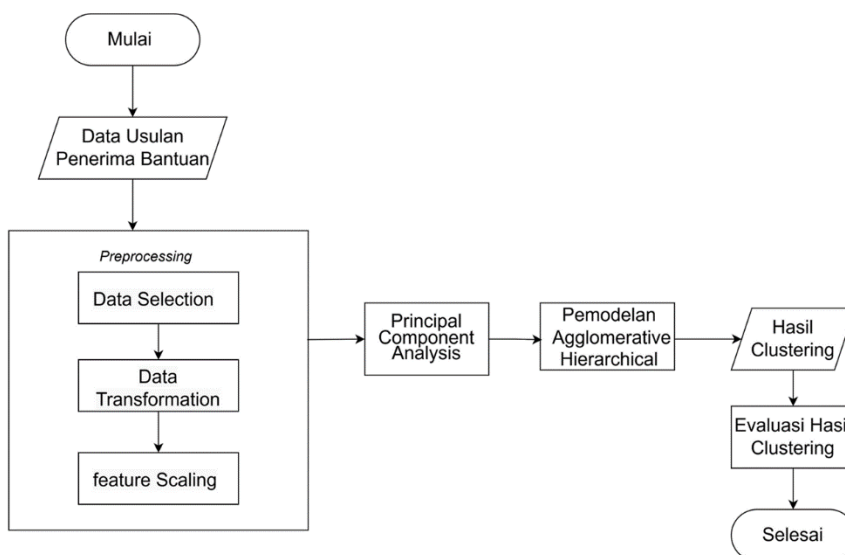
Berdasarkan uraian di atas, akan dilakukan pengelompokan menggunakan algoritma Agglomerative Hierarchical Clustering (AHC) untuk mengelompokkan prioritas penerima bantuan rumah. AHC adalah *clustering* dengan pendekatan hirarki, dimana mengelompokkan data yang mirip dalam hirarki yang sama dan yang tidak mirip di hirarki yang agak jauh. *Agglomerative* melakukan *clustering* dari N *cluster* menjadi satu kesatuan *cluster*, dimana N adalah jumlah data, sebaliknya *divisive* melakukan proses *clustering* dari satu *cluster* menjadi N *cluster* [6]. Pada AHC, akan menggunakan tiga pendekatan yang berbeda, yakni *single linkage*,

average linkage, dan *complete linkage*, untuk *cluster* yang dapat dibandingkan. Melalui hasil data yang dikelompokkan dan jumlah *cluster* yang sudah optimal, diharapkan AHC dapat membantu dan mempermudah pemerintah kabupaten dalam mengelompokkan prioritas penerima bantuan rumah. Hasil algoritma ini akan dievaluasi menggunakan *silhouette coefficient* untuk menentukan efektivitas pengelompokan.

METODE

Alur Penelitian

Proses alur kerja keseluruhan yang dilakukan dalam penelitian ini dapat dibagi menjadi beberapa tahap seperti yang diilustrasikan pada Gambar 1 di bawah ini.



Gambar 1. Alur penelitian.

Diagram pada Gambar 1 dimulai dengan tahap memasukkan data mengenai usulan nama-nama penerima bantuan rumah layak huni. Setelah data dimasukkan, akan dilakukan tahap *preprocessing* untuk mengubah data sehingga siap digunakan. Langkah pertama adalah *data selection* untuk memilih fitur yang relevan dari total 40 atribut yang ada. Pemilihan atribut ini berdasarkan pertimbangan bahwa tidak semua atribut memiliki kontribusi penting dalam pengambilan keputusan, atribut yang dipilih berfokus pada kondisi rumah. Sebagian atribut memiliki nilai yang sama sehingga tidak memberikan kontribusi yang signifikan dalam proses *clustering*. Setelah atribut dipilih, dilakukan *data transformation* dengan teknik One-Hot Encoding untuk mengkonversi data kategorikal menjadi format numerik. Proses selanjutnya adalah melakukan *feature scaling* data untuk mengubah data ke dalam skala yang seragam menggunakan metode Z-score dan Min-Max. Min-Max adalah teknik yang menormalkan data dengan menggunakan nilai minimum dan maksimum dalam data sedangkan Z-score menggunakan rata-rata dan deviasi standar data.

Tahap selanjutnya, data yang telah diproses dan distandarisasi dilakukan ekstraksi ciri menggunakan Principal Component Analysis (PCA) untuk mengurangi dimensi dan mempertahankan varians data. PCA adalah metode deterministik yang data dari bentuk aslinya ke

bentuk baru. Variabel baru yang dihasilkan adalah kombinasi linier dari variabel-variabel asli. Tujuan utama PCA adalah mereduksi dimensi daigunakan untuk mengurangi dimensi suatu kumpulan data dengan mentransformasikan dta sambil mempertahankan sebanyak mungkin informasi varians dari data asli. Dengan cara ini, PCA menyederhanakan data dengan menghasilkan variabel baru, yang dikenal sebagai komponen utama, yang menggantikan variabel-variabel asli. Komponen utama ini tidak berkorelasi satu sama lain dan menyimpan informasi yang paling signifikan dari data, sehingga memudahkan analisis lebih lanjut dan mengurangi kompleksitas data [7].

Tahap ketiga adalah penerapan *clustering* menggunakan metode *agglomerative clustering*. Dalam penelitian ini, proses *clustering* dilakukan menggunakan Google Collab. Data hasil ekstraksi ciri kemudian dikenakan tiga pendekatan *linkage* untuk menentukan jarak antara *cluster*, yaitu *single linkage*, *linkage average*, dan *complete linkage*.

Tahap keempat adalah evaluasi hasil clustering yang diperoleh menggunakan *silhouette coefficient*. *Silhouette coefficient* adalah metode yang digunakan untuk mengevaluasi kualitas sebuah cluster dengan menggabungkan nilai kohesi dan nilai separasi. *Silhouette coefficient* memiliki skala dari -1 hingga 1. Ketika nilai *silhouette coefficient* mendekati 1, itu menandakan bahwa data cenderung sangat sesuai dengan *cluster* yang ditetapkan. Di sisi lain, nilai *silhouette coefficient* yang negatif menunjukkan bahwa data tersebut tidak cocok dengan *cluster* yang telah ditentukan [8]. Berikut adalah interpretasi dari nilai *silhouette coefficient* disajikan dalam Tabel 1 [9].

Tabel 1. Struktur nilai *silhouette coefficient*

<i>Silhouette coefficient</i> (skala -1 sampai 1)	Variabel A (kualitatif)
0.71 – 1.00	Struktur yang dihasilkan kuat
0.51 – 0.70	Struktur yang dihasilkan baik
0.26 – 0.50	Struktur yang dihasilkan lemah
≤ 0.25	Cluster yang terbentuk tidak teratur

Data

Penelitian ini menggunakan data usulan nama-nama penerima bantuan rumah layak huni yang berada di Kabupaten Sumba Barat Daya pada tahun 2022 yang berbentuk file Ms. Excel berekstensi xlsx. Pada penelitian ini, total data yang digunakan pada tahun 2022 adalah 1.211 record. Terdapat 7 atribut yang digunakan dalam penelitian ini, yaitu: Sumber Penerangan, Material Lantai, Material Dinding, Material Atap, Jarak Tpa Tinja, KM dan Jamban, dan Jenis Kloset. Berikut ini ditampilkan beberapa data yang digunakan dalam penelitian ini.

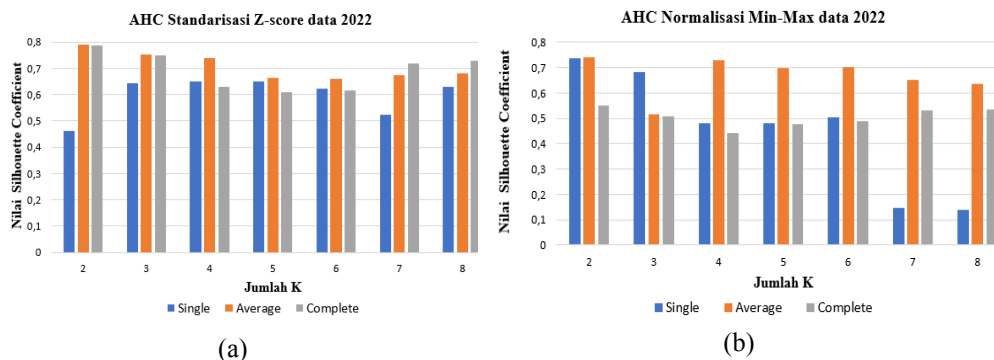
Tabel 2. Atribut data

Sumber penerangan	Material Lantai	Material Dinding	Material Atap	Jarak-T pa Tinja	Km dan Jamban	Jenis Kloset
Listrik Non PIn	Bambu	Bambu	Jerami	> 10 M	Sendiri	Cemplung /Cubluk
Listrik Non PIn	Bambu	Bambu	Daun-Dau nan	> 10 M	Sendiri	Cemplung /Cubluk
0.26 – 0.50	Plesteran	Plesteran Anyaman Bambu	Jerami	< 10 M	Bersama /Mck Komunal	Leher Angsa
Listrik Non PIn	Bambu	Bambu	Daun-Dau nan	> 10 M	Sendiri	Cemplung /Cubluk

Listrik Non Pln	Tanah	Bambu	Daun-Dau nan	> 10 M	Sendiri	Cemplung /Cubluk
--------------------	-------	-------	-----------------	--------	---------	---------------------

HASIL DAN PEMBAHASAN

Hasil validasi *cluster* menggunakan *silhouette coefficient* pada metode AHC dilakukan dengan menggunakan nilai K sebanyak 8.



Gambar 2. a) Grafik hasil nilai *silhouette coefficient* algoritma AHC Menggunakan Z-score, b) Grafik hasil nilai *silhouette coefficient* algoritma AHC menggunakan Min-Max.

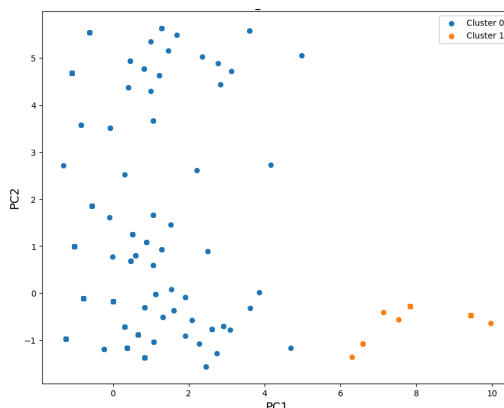
Pada Gambar 2 (a), terlihat bahwa penerapan algoritma AHC dengan standarisasi Z-score K = 2 memiliki nilai *silhouette coefficient* tertinggi sebesar 0,79 menggunakan pendekatan average linkage. Sedangkan pada Gambar 2 (b), AHC dengan normalisasi Min-Max K = 2 mencapai nilai *silhouette coefficient* tertinggi sebesar 0,741 dengan pendekatan yang sama. Dengan demikian, untuk penerapan AHC pada data tahun 2022, data yang menggunakan standarisasi Z-score lebih baik dalam validasi *silhouette coefficient* dibandingkan dengan normalisasi min-max. Hal ini menunjukkan bahwa metode standarisasi Z-score menghasilkan *clustering* yang lebih baik dengan struktur *cluster* yang lebih bagus dengan struktur *cluster* yang kuat. Penerapan AHC pada data tahun 2022 menunjukkan struktur yang baik dari kedua pendekatan, yaitu Z-score dan Min-Max, seperti yang terlihat pada Tabel 3.

Tabel 3. Perbandingan hasil AHC dan DBSCAN

<i>Normalization / Standardization</i>	<i>Silhouette coefficient</i>	<i>Cluster</i>
Z-Score	0,79	2
Min-Max	0,741	2

Pada Tabel 3, terlihat bahwa metode AHC dengan standarisasi Z-score dan normalisasi Min-Max sama-sama menghasilkan 2 *cluster*. Namun, AHC dengan standarisasi Z-score menggunakan pendekatan average linkage lebih unggul karena nilai *silhouette coefficient* yang lebih tinggi dibandingkan dengan normalisasi Min-Max. Oleh karena itu, pada data tahun 2022, AHC dengan Z-score menghasilkan *clustering* yang lebih baik.

Pada Gambar 3 terlihat bahwa persebaran *cluster* tidak merata, dengan jumlah *cluster* 0 lebih banyak daripada *cluster* 1. Untuk memperjelas jumlah obyek dalam setiap *cluster*, hasil pembagian tersebut dirangkum dalam bentuk Tabel 4.



Gambar 3. Visualisasi hasil AHC dengan standarisasi Z-score.

Tabel 4. Visualisasi hasil AHC dengan standarisasi Z-score.

Cluster	Obyek
0	1.116
1	95

Pada Tabel 4, terlihat bahwa *cluster* 0 memiliki jumlah obyek terbanyak, yaitu sebanyak 1.116 obyek. Analisis berdasarkan data menunjukkan bahwa struktur *cluster* 0 ditandai dengan penggunaan sumber penerangan non-listrik, lantai rumah yang mayoritas terbuat dari bambu dan tanah, dinding rumah yang umumnya terbuat dari bambu, serta penggunaan kloset bertipe cubluk. Sementara itu, struktur *cluster* 1 menunjukkan ciri-ciri seperti penggunaan listrik PLN tanpa meteran, lantai rumah yang berplester, dinding setengah plesteran dan setengah anyaman bambu, serta penggunaan kloset bertipe leher angsa. Berdasarkan hasil analisis diatas, *cluster* 0 ditetapkan sebagai prioritas pertama karena kondisi rumah dan fasilitas yang paling memerlukan bantuan. Sedangkan *cluster* 1 ditetapkan sebagai prioritas kedua dalam penerimaan bantuan. Dengan menentukan prioritas berdasarkan karakteristik *cluster* yang terbentuk, alokasi bantuan dapat dilakukan secara lebih tepat sasaran.

KESIMPULAN

Penerapan Agglomerative Hierarchical Clustering untuk mengelompokkan prioritas penerima bantuan rumah pada data tahun 2022 dengan standarisasi Z-score menghasilkan hasil terbaik, di mana dua *cluster* terbentuk dengan nilai *silhouette coefficient* sebesar 0,79. Nilai ini menunjukkan bahwa metode standarisasi Z-score lebih efektif dalam mengelompokkan data dibandingkan dengan normalisasi Min-Max, meskipun kedua metode tersebut berhasil membentuk struktur *cluster* yang kuat. Keunggulan Z-score dalam menghasilkan nilai *silhouette coefficient* yang lebih tinggi menunjukkan bahwa pendekatan ini mampu memberikan pemisahan *cluster* yang lebih jelas dan lebih tepat.

DAFTAR PUSTAKA

- [1] D. V. Ferezagia, "Analisis tingkat kemiskinan di Indonesia," *Jurnal Sosial Humaniora Terapan*, vol. 1, no. 1, p. 1, 2018.

-
- [2] T. T. WUNGO, “PENYEBAB DAN DAMPAK KEMISKINAN PADA MASYARAKAT DESA (Studi Kasus di Desa Waikarara, Kecamatan Kodi Balaghar, Kabupaten Sumba Barat Daya),” 2021.
 - [3] D. Handayani and S. Safrida, “Implementasi Program Bantuan Stimulan Perumahan Swadaya di Masa Pandemi COVID-19 dalam Meningkatkan Kesejahteraan Masyarakat,” *Jurnal Ilmiah Ilmu Sosial*, vol. 8, no. 1, pp. 34–45, 2022.
 - [4] K. P. Simanjuntak and U. Khaira, “Pengelompokan titik api di provinsi jambi dengan algoritma agglomerative hierarchical clustering: Hotspot clustering in jambi province using agglomerative hierarchical clustering algorithm,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 7–16, 2021.
 - [5] I. Sufairoh, A. C. Rani, K. Amalia, and D. Rolliawati, “Perbandingan Hasil Analisis Clustering Metode K-Means, DBSCAN Dan Hierarchical Pada Data Marketplace Electronic Phone,” *JOINS (Journal of Information System)*, vol. 8, no. 1, pp. 97–105, 2023.
 - [6] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, “Cluster analysis,” 2011.
 - [7] A. Arab, “Ekstraksi Ciri Orde Pertama dan Metode Principal Component Analysis untuk Mengidentifikasi Jenis Telur”.
 - [8] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.
 - [9] A. Struyf, M. Hubert, and P. Rousseeuw, “Clustering in an object-oriented environment,” *Journal of Statistical Software*, vol. 1, pp. 1–30, 1997.

Halaman ini sengaja dikosongkan