

PERBANDINGAN ANALISIS *K-MEDIAN* DAN *FUZZY C-MEANS CLUSTERING* PADA ASURANSI KESEHATAN

WINDA MERTORINI



**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Analisis *K-median* dan *Fuzzy C-means Clustering* pada Asuransi Kesehatan” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Desember 2024

Winda Mertorini
G5402201030

ABSTRAK

WINDA MERTORINI. Perbandingan Analisis *K-median* dan *Fuzzy C-means Clustering* pada Asuransi Kesehatan. Dibimbing oleh RETNO BUDIARTI dan NGAKAN KOMANG KUTHA ARDANA.

Penelitian ini bertujuan untuk membandingkan metode *k-median* dan *fuzzy c-means* (FCM) *clustering* dalam mengelompokkan data pelanggan asuransi kesehatan. Data yang digunakan meliputi usia, BMI, gaya hidup, dan biaya kesehatan pelanggan di Amerika Serikat. Uji multikolinearitas menunjukkan tidak adanya masalah multikolinearitas, sehingga semua variabel dapat digunakan dalam analisis *clustering*. Penentuan jumlah *cluster* optimal dilakukan menggunakan indeks *silhouette coefficient*, yang menunjukkan bahwa dua *cluster* adalah pilihan terbaik untuk kedua metode. Berdasarkan evaluasi *Davies-Bouldin Index* (DBI), metode *k-median* terbukti lebih unggul dibandingkan FCM karena menghasilkan *cluster* yang lebih kompak dan terpisah dengan baik. Hasil *clustering k-median* juga menunjukkan pemisahan yang lebih jelas berdasarkan variabel kategorikal seperti gaya hidup. Keunggulan ini membuktikan bahwa metode *k-median* memiliki karakteristik *cluster* yang lebih relevan dengan perbedaan signifikan pada usia, BMI, kebiasaan merokok, dan biaya kesehatan.

Kata kunci: *k-median clustering*, *fuzzy c-means clustering*, analisis *cluster*, *Davies-Bouldin Index* (DBI)

ABSTRACT

WINDA MERTORINI. Comparison of *K-median* and *Fuzzy C-means Clustering* Analysis in Health Insurance. Supervised by RETNO BUDIARTI and NGAKAN KOMANG KUTHA ARDANA.

This study aims to compare the *k-median* and *fuzzy c-means* (FCM) clustering methods in segmenting health insurance customer data. The data includes age, BMI, lifestyle, and healthcare costs of customers in the United States. Multicollinearity tests showed no multicollinearity issues, allowing all variables to be used in the clustering analysis. The optimal number of clusters was determined using the *silhouette coefficient* index, indicating that two clusters were the best choice for both methods. Based on the evaluation using the *Davies-Bouldin Index* (DBI), the *k-median* method proved superior to FCM by producing more compact and well-separated clusters. *K-median* clustering results also displayed clearer separation based on categorical variables such as lifestyle. This superiority demonstrates that the *k-median* method provides more relevant cluster characteristics with significant differences in age, BMI, smoking habits, and healthcare.

Keywords: *k-median clustering*, *fuzzy c-means clustering*, cluster analysis, *Davies-Bouldin Index* (DBI)



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2024
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

PERBANDINGAN ANALISIS *K-MEDIAN* DAN *FUZZY C-MEANS CLUSTERING* PADA ASURANSI KESEHATAN

WINDA MERTORINI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Aktuaria pada
Program Studi Aktuaria

**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Skripsi : Perbandingan Analisis *K-median* dan *Fuzzy C-means Clustering* pada Asuransi Kesehatan

Nama : Winda Mertorini
NIM : G5402201030

Disetujui oleh

Pembimbing 1:

Dr. Ir. Retno Budiarti, M.S.

Pembimbing 2:

Ir. Ngakan Komang Kutha Ardana, M.Sc.

Diketahui oleh

Ketua Departemen Matematika:

Prof. Ir. Endar Hasafah Nugrahani, MS.
NIP 19631228 198903 2 001

Tanggal Ujian: 11 November 2024

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan dengan judul “Perbandingan Analisis *K-median* dan *Fuzzy C-means Clustering* pada Asuransi Kesehatan”. Penyusunan karya ilmiah ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, ungkapan terima kasih sebesar-besarnya penulis sampaikan kepada:

1. Ibu Dr. Ir. Retno Budiarti, M.S dan Bapak Ir. Ngakan Komang Kutha Ardana, M.Sc. selaku dosen pembimbing yang telah membimbing dan banyak memberi saran.
2. Ibu Poni Astutik dan Bapak Tohirin selaku orang tua serta Mila Oktorini selaku kakak penulis yang telah memberikan banyak doa, dukungan, dan motivasi kepada penulis.
3. Seluruh dosen dan staf Departemen Matematika IPB yang telah membantu penulis selama masa perkuliahan dan penyusunan skripsi.
4. Keluarga Besar Lawalata IPB serta teman-teman Departemen Matematika 57 yang telah mendampingi selama penulis menempuh pendidikan di IPB University.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Desember 2024

Winda Mertorini

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	x
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Asuransi	3
2.2 Analisis <i>Cluster</i>	3
2.3 Metode <i>Clustering</i>	4
2.4 Banyak <i>Cluster</i> yang Optimum	7
2.5 Uji Multikolinearitas	8
2.6 Perbandingan Metode Terbaik	9
III METODE PENELITIAN	11
3.1 Data	11
3.2 Variabel	11
3.3 Perangkat Lunak	12
3.4 Tahapan Analisis Data	12
IV HASIL DAN PEMBAHASAN	14
4.1 Deskripsi Data	14
4.2 Uji Multikolinearitas	18
4.3 Banyak <i>Cluster</i> yang Optimum	19
4.4 Hasil <i>Clustering</i>	22
4.5 Perbandingan Metode <i>K-median</i> dan <i>Fuzzy C-means Clustering</i>	27
4.6 Analisis Hasil <i>Clustering</i>	27
V SIMPULAN DAN SARAN	30
5.1 Simpulan	30
5.2 Saran	30
DAFTAR PUSTAKA	31
LAMPIRAN	33
RIWAYAT HIDUP	37

DAFTAR TABEL

1	Kriteria <i>silhouette coefficient</i>	8
2	Deskripsi variabel dataset	11
3	Deskripsi data	14
4	Matriks korelasi	15
5	Nilai VIF dan <i>tolerance</i>	18
6	Indeks <i>silhouette coefficient k-median clustering</i>	20
7	Indeks <i>silhouette coefficient fuzzy c-means clustering</i>	21
8	Pusat <i>cluster k-median clustering</i>	23
9	Jumlah anggota pada setiap <i>cluster k-median</i>	23
10	Pusat <i>cluster fuzzy c-means clustering</i>	25
11	Jumlah anggota pada setiap <i>cluster fuzzy c-means</i>	26
12	Nilai metrik evaluasi hasil <i>k-median</i> dan <i>fuzzy c-means clustering</i>	27
13	Nilai rata-rata hasil <i>k-median clustering</i>	28

DAFTAR GAMBAR

1	Tahapan penelitian	13
2	Histogram dan boxplot usia	15
3	Histogram dan boxplot BMI	16
4	Histogram gaya hidup	17
5	Histogram dan boxplot <i>charges</i> (biaya total)	17
6	Plot <i>silhouette coefficient index k-median clustering</i>	20
7	Plot <i>silhouette coefficient index fuzzy c-means clustering</i>	22
8	Hasil <i>k-median clustering</i>	24
9	Hasil <i>fuzzy c-means clustering</i>	26

DAFTAR LAMPIRAN

1	Hasil <i>Clustering</i> pada 30 Data Pengguna Asuransi Kesehatan dalam Penelitian	33
2	Syntax Program R	33