

Deteksi Outlier Hasil Clustering Algoritma K-Medoids Menggunakan Metode Boxplot Pada Data KIP Kuliah

Elsya Sabrina Asmita Simorangkir*, Andysah Putera Utama Siahaan, Leni Marlina, Darmeli Nasution, Zulham Sitorus

Pascasarjana, Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

Email: ^{1,*}elsyasabrinaas@gmail.com, ²andiesiahaan@gmail.com, ³lheny@pancabudi.ac.id,

⁴darmelinasution@dosen.pancabudi.ac.id, ⁵zulhamsitorus@dosen.pancabudi.ac.id

Email Penulis Korespondensi: elsyasabrinaas@gmail.com

Submitted: 03/07/2024; Accepted: 14/08/2024; Published: 15/08/2024

Abstrak—Dalam proses pembentukan cluster dengan algoritma K-Medoids sering terjadi anomali hasil cluster seperti outlier. Nilai ini muncul sebagai pengecualian dalam pola data yang ada. Outlier terjadi karena kesalahan pengukuran, kejadian langka, atau karena faktor lain yang tidak terduga. Pada penelitian ini, dataset yang digunakan adalah data calon mahasiswa penerima KIP di Universitas Budi Darma, dimana tingginya peminat KIP Kuliah sementara kuota terbatas yang mengakibatkan pengelola KIP Kuliah kadang kesulitan menentukan mahasiswa yang layak memperoleh KIP Kuliah. Untuk itu, teknik clustering K-Medoids digunakan untuk clustering data sebanyak 54 orang calon mahasiswa yang layak menerima KIP Kuliah Merdeka dan yang tidak layak. Dari hasil cluster tersebut dilakukan deteksi outlier dengan metode box plot dengan tujuan mengetahui apakah setiap anggota cluster benar-benar berada pada cluster yang sesuai atau yang tidak. Hasilnya cluster data dibagi menjadi 2(K-2). Pada pemilihan centroid secara max min, cluster I terdiri atas 52 anggota dan cluster II terdiri atas 2 anggota, dimana data outliernya terdiri atas 3 data, sedangkan pemilihan centroid acak (python), cluster I terdiri atas 36 anggota dan cluster II 18 anggota dengan data outliernya terdiri atas 4 anggota. Akurasi hasil clustering antara pemilihan centroid secara max min dengan acak memiliki akurasi 64,81%, dan akurasi outliernya 75%.

Kata Kunci: KIP Kuliah; Outlier; K-Medoids; Box Plot

Abstract—In the process of forming clusters with the K-Medoids algorithm, cluster result anomalies often occur, such as outliers. This value appears as a revelation in existing data patterns. Outliers occur due to measurement errors, rare events, or due to other unexpected factors. In this research, the dataset used is data on prospective KIP recipient students at Budi Darma University, where there is a high level of interest in KIP Kuliah while the quota is limited, which means that KIP Kuliah administrators sometimes have difficulty determining which students are eligible to receive KIP Kuliah. For this reason, the K-Medoids clustering technique was used to cluster data on 54 prospective students who were eligible to receive KIP Kuliah Merdeka and those who were not eligible. From the cluster results, outlier detection was carried out using the box plot method with the aim of finding out whether each cluster member was actually in the appropriate cluster or not. The result is that the data cluster is divided into 2 (K-2). In the max min centroid selection, cluster I consists of 52 members and cluster II consists of 2 members, where the outlier data consists of 3 data, while in random centroid selection (python), cluster I consists of 36 members and cluster II 18 members with data The outlier consists of 4 members. The accuracy of the clustering results between max min and random centroid selection has an accuracy of 64.81%, and the outlier accuracy is 75%.

Keywords: KIP Kuliah; Outlier; K-Medoids; Box Plot

1. PENDAHULUAN

Beasiswa KIP (Kartu Indonesia Pintar) kuliah merdeka merupakan program pemerintah untuk membantu calon mahasiswa yang memiliki potensi akademik baik tetapi memiliki keterbatasan ekonomi dengan tujuan untuk meningkatkan akses dan kesempatan belajar bagi keluarga yang kurang mampu. Calon mahasiswa yang layak menerima beasiswa KIP Kuliah adalah mahasiswa yang memenuhi kriteria ataupun ketentuan yang ditetapkan oleh pemerintah yang tertulis pada Buku Panduan KIP Kuliah. Penyelenggara dan pengelola penerima beasiswa KIP Kuliah bukan hanya Perguruan Tinggi Negeri tetapi juga Perguruan Tinggi Swasta yang memenuhi persyaratan yang telah ditentukan[1].

Calon mahasiswa penerima beasiswa KIP kuliah merdeka di Universitas Budi Darma, selain harus memenuhi kriteria dan ketentuan yang telah ditetapkan juga harus mengikuti tes seleksi tertulis dan wawancara melalui seleksi mandiri. Tingginya peminat melalui jalur beasiswa menjadi kendala pada bagian pengelola beasiswa KIP kuliah merdeka dikarenakan jumlah ataupun kuota yang diberikan kepada masing-masing perguruan tinggi terbatas. Selain itu yang sering terjadi di lapangan adalah calon mahasiswa yang mendapatkan beasiswa KIP kuliah merdeka ternyata mampu secara ekonomi dan juga adanya calon mahasiswa yang memiliki kriteria nilai yang hampir sama satu dengan lainnya. Dengan banyaknya data penerima KIP kuliah merdeka yang harus dikelola, ada tantangan yang signifikan dalam memastikan agar bantuan ini benar-benar sampai kepada calon mahasiswa penerima yang memenuhi kriteria. Salah satu tantangan utama adalah adanya data yang tidak konsisten atau anomali, yang dikenal sebagai outlier. Outlier dalam data penerima beasiswa KIP Kuliah merdeka memiliki banyak implikasi penting karena dapat menunjukkan kesalahan data. Dengan mendeteksi dan menganalisis outlier, program dapat lebih efektif dalam memberikan bantuan, memastikan keadilan, dan meningkatkan keberhasilan penerima beasiswa.

Outlier adalah data yang mempunyai karakteristik berbeda secara signifikan dengan data lain dalam kelompoknya. Pencilan dapat terjadi karena berbagai alasan, seperti kesalahan pengukuran, kejadian langka, atau faktor tidak terduga lainnya[2]. Mendeteksi dan menangani outlier memastikan analisis data lebih akurat dan representatif, serta memberikan informasi berguna seperti mengidentifikasi kesalahan data, menemukan peristiwa luar biasa, dan menemukan tren yang tidak biasa. Ada beberapa cara untuk mengidentifikasi outlier, salah satunya adalah box plot[3][4].

Penelitian yang dilakukan oleh Moh Yamin Darsyah tentang penggunaan Stem and Leaf dan Boxplot untuk analisis data, menggunakan metode grafik stem and leaf dan boxplot untuk menganalisis dan membandingkan data nilai mahasiswa dari empat fakultas pada salah satu perguruan tinggi. Tujuan untuk mengetahui penyebaran data, kesimetrisan, dan adanya data outlier pada nilai mahasiswa dari tujuh mata pelajaran ujian nasional. Hasil analisis menunjukkan bahwa metode ini efektif dalam menggambarkan distribusi data dan mengidentifikasi outlier[5]. Penelitian yang dilakukan oleh Pardomuan Robinson Sihombing dkk tahun 2022 tentang identifikasi data outlier (pencilan) dan kenormalan data pada data univariat serta alternatif penyelesaiannya. Penelitian bertujuan untuk mengidentifikasi data outlier dan memeriksa kenormalan data pada data univariat. Studi ini menggunakan data persentase kemiskinan di Indonesia untuk tahun 2022 yang diperoleh dari Badan Pusat Statistik (BPS). Hasil penelitiannya adalah menunjukkan bahwa identifikasi dan penanganan outlier sangat penting dalam analisis data statistik untuk memastikan validitas hasil[6].

Pada penelitian ini berbeda dengan penelitian sebelumnya, dimana pada penelitian ini melakukan deteksi outlier hasil dari clustering data dengan algoritma K-Medoids dengan metode box plot. Selain itu, data yang digunakan sebagai sampel dan uji coba dinormalisasikan terlebih dahulu dengan metode max min dengan tujuan untuk menyederhanakan dan menghilangkan ketimbangan nilai antar kriteria dalam data. Pengujian data menggunakan tools anaconda, editor jupyter notebook, bahasa pemrograman pyhton serta library yang digunakan adalah sklearn_extra.cluster import Kmedoids.

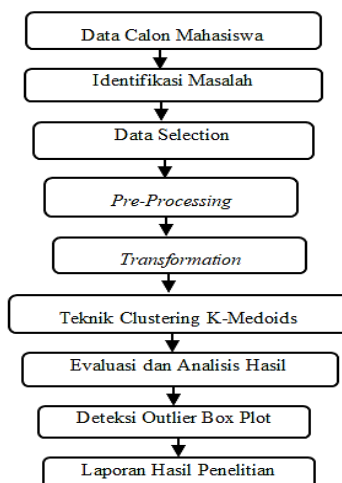
Penelitian ini tidak terlepas dari peran big data. Big data telah menjadi salah satu topik utama dalam teknologi informasi dan data analytics dalam beberapa dekade terakhir. Istilah ini mengacu pada kumpulan data yang sangat besar dan kompleks[7]. Volume data yang besar ini memerlukan metode yang efektif untuk mengekstraksi informasi berharga. Salah satu metode yang sering digunakan adalah clustering, yang bertujuan untuk mengelompokkan data ke dalam kelompok-kelompok homogen berdasarkan kesamaan tertentu. Salah satu algoritma yang digunakan untuk masalah clustering adalah Algoritma K-Medoids[8]. K-Medoids adalah algoritma clustering yang bertujuan untuk meminimalkan jarak total antara titik data dan pusat cluster[9][10]. Algoritma ini lebih tahan terhadap outlier dan data dengan distribusi yang tidak normal[11]. Namun, dalam proses clustering, seringkali ada data yang sangat berbeda dari data lainnya dalam kelompok yang sama, yang dikenal sebagai outlier[12].

Penelitian mengenai deteksi outlier pada hasil clustering data dengan menggunakan metode boxplot menunjukkan bahwa pendekatan ini efektif untuk mengidentifikasi anomali. Integrasi metode clustering dan boxplot dalam deteksi ini memungkinkan identifikasi dan penanganan outlier yang lebih akurat, sehingga mendukung tujuan utama program beasiswa yaitu memastikan penerima yang tepat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian ini ada langkah-langkah yang harus dilakukan untuk menyelesaikan penelitian ini, adapun tahapan yang dilakukan seperti pada gambar 1 berikut:



Gambar 1. Tahapan Penelitian

a. Data Calon Mahasiswa

Penelitian diawali dengan adanya data calon mahasiswa penerima KIP Kuliah pada bagian pengelola KIP Kuliah di Universitas Budi Darma. Data tersebut belum dianalisis dengan baik untuk digunakan sebagai salah satu alternatif dalam pengambilan keputusan khususnya dalam penentuan mahasiswa yang layak dan yang tidak layak menerima KIP Kuliah Merdeka di Universitas Budi Darma.

b. Identifikasi Masalah

Salah satu masalah yang dihadapi oleh pengelola KIP Kuliah di Universitas Budi Darma adalah jumlah calon mahasiswa penerima tidak sebanding dengan kuota yang tersedia, penentuan mahasiswa yang layak atau tidak layak kadang kurang tepat, nilai kriteria mahasiswa yang hampir sama dan lain sebagainya sehingga kadang pengelola KIP Kuliah kesulitan dalam menentukan mahasiswa yang layak dan tidak layak menerima KIP Kuliah.

c. Data Selektion

Data KIP Kuliah yang sudah ada sebelumnya yaitu dari tahun 2018 sampai tahun 2023 dilakukan seleksi untuk menentukan data yang layak digunakan sebagai data sample dan data uji pada penelitian ini.

d. Pre-processing

Dari data yang sudah dilakukan dilakukan pre-processing yang terdiri atas kegiatan cleaning data, integration data, menghilangkan miss value data, mengisi data kosong dan menglekapai data kosong.

e. Transformation

Selanjutnya dilakukan proses perubahan data sesuai dengan kebutuhan dari algoritma clustering k-Medoids, dimana data terdiri atas attribute objek dan beberapa attribute kriteria. Kemudian merubah data dalam bentuk kategori menjadi data dalam bentuk numerik, serta melakukan normalisasi data dengan metode max min untuk menyederhanakan data dan menghilangkan ketimpangan antar kriteria data. Berikut adalah rumus untuk normalisasi data dengan max min[13].

$$X_{baru} = \frac{X_{lama} - X_{min}}{X_{max} - X_{min}} \quad (1)$$

f. Teknik Clustering K-Medoids

Setelah data ditransformasikan maka selanjutnya data diclusteringkan dengan algoritma K-Medoids. Pada penelitian ini, cluster data dibagi menjadi 2 (dua) yaitu cluster layak dan yang tidak layak ($K=2$). Langkah awal yang dilakukan untuk cluster data adalah menentukan K, kemudian menentukan nilai centroid, menghitung jarak dengan menggunakan model Euclidean distance, menentukan cluster data dan menghitung cost nilai cluster[14][4]. Pada penelitian ini khusus menentukan nilai centroid dilakukan dengan dua cara yaitu secara max min yaitu berdasarkan total nilai attribute data, dan secara acak dengan menggunakan bahasa python[15][16].

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (2)$$

g. Evaluasi dan Analisis Hasil

Setelah diperoleh hasil clustering data, maka dilakukan evaluasi hasil dan analisis hasil clustering berupa penamaan clustering dan kesesuaian hasil clustering dengan datanya. Pada penelitian ini untuk penamaan clustering yaitu cluster I untuk calon mahasiswa yang layak menerima KIP Kuliah dan cluster II untuk mahasiswa yang tidak layak menerima KIP Kuliah.

h. Deteksi Outlier dengan Box Plot

Setelah memperoleh hasil dan evaluasi hasil clustering, selanjutnya dilakukan deteksi outlier hasil cluster dengan metode Box Plot. Deteksi outlier berdasarkan Sum of Squared Errors (SSE), dan dari nilai SSE tersebut data diurutkan dari nilai SSE terkecil yang kenilai yang lebih besar. Selanjutnya dilakukan pencarian nilai Q1, Q2, dan Q3, kemudian dapat nilai jarak batas bawah dan nilai jarak batas atas[17][18][19]. Nilai yang diluar dari jarak batas bawah dan jarak batas atas, itu yang menjadi data outlier. Persamaan untuk mencari nilai SSE[20]:

$$SSE = \sum_{i=1}^n (X_i - Y_i)^2 \quad (3)$$

Rumus metode box plot:

$$IQR = Q3 - Q1 \quad (4)$$

$$Jarak\ batas\ bawah = Q1 - 1,5 * IQR \quad (5)$$

$$Jarak\ batas\ atas = Q3 + 1,5 * IQR \quad (6)$$

i. Laporan Hasil Penelitian

Langkah terakhir adalah menulis hasil penelitian sebagai bentuk pertanggungjawaban penulis terhadap penelitian yang dilakukan, kemudian dipublish dalam bentuk artikel pada jurnal terakreditasi sehingga dapat bermanfaat menjadi salah satu referensi bagi penelitian selanjutnya.

3. HASIL DAN PEMBAHASAN

Dalam penentuan calon mahasiswa yang layak memperoleh KIP Kuliah Merdeka yang diperlukan adalah data calon mahasiswa yang mengikuti seleksi calon mahasiswa KIP Kuliah sebelumnya. Data yang digunakan harus mempunyai attribute kriteria. Adapun datanya seperti pada tabel 1 berikut:

Tabel 1. Data Seleksi KIP Kuliah Merdeka

No	Nama	Prestasi Sekolah	Pekerjaan Orang Tua	Penghasilan Orang Tua	Jumlah Tanggungan	Hasil Ujian
1	Agus Minta Riang Zega	Kosong	Petani	1,000,000	5	85
2	Desika Marbun	Kosong	Petani	1,500,000	1	91
3	Amran Saleh Harahap	Kosong	Karyawan Swasta	1,000,000	2	90
4	Anro Tamba	Kosong	Petani	1,000,000	3	86
5	Kristina Lenti Kristie Dakhi	Kosong	Petani	1,000,000	3	83
6	Kevin Maulana Putra	Kota	Lainnya	1,000,000	8	82
7	Ibnu Prinanda	Kosong	Lainnya	1,000,000	1	81
8	Bobo Firman Zega	Kosong	Petani	1,000,000	3	81
9	Patar Situmorang	Kab	Petani	1,000,000	5	80
10	Susi Mardiana Giawa	Kab	Petani	1,200,000	5	80
11	Ritrit Yanto Zai	Kosong	Petani	1,000,000	4	80
12	Rico Albert Andika Saragih	Kab	Petani	1,000,000	4	78
13	Mei Fides Purnawati Zega	Prov	Petani	1,000,000	4	77
14	Miko Putra Haposan Tinambunan	Kab	Lainnya	500,000	3	77
15	Elhanan Kris Daniel Simanjuntak	Kosong	Lainnya	1,000,000	3	77
16	Putri Patrisia Purba	Kosong	Wiraswasta	2,000,000	3	76
17	Justari Zega	Kosong	Petani	1,000,000	4	76
18	Enzel Febrianti Telaumbanua	Kosong	Petani	1,000,000	5	76
19	Friska Agusnayanti Zebua	Kosong	Petani	500,000	3	74
20	Devi Ritonga	Kab	Petani	1,000,000	4	74
21	Hardi Giat Sofian Manalu	Kosong	Petani	1,000,000	4	73
22	Walderina Situmeang	Kosong	Petani	2,000,000	4	73
23	Rani Selvira	Kosong	Petani	1,000,000	3	72
24	Ribka Tasya Simarmata	Kosong	Petani	2,000,000	5	71
25	Athillah Putra Windra	Provinsi	Lainnya	2,250,000	5	70
26	Najwa Saufa Silmi	Kosong	Petani	750,000	6	69
27	Otomosi Gulo	Kosong	Petani	1,000,000	5	69
28	Tetikariani Halawa	Kosong	Petani	1,000,000	3	66
29	Ampuni Waruwu	Kosong	Petani	1,000,000	5	65
30	Edarni Waruwu	Kosong	Petani	750,000	3	64
31	Lasria Mandalahi	Kosong	Petani	500,000	4	63
32	Elgayuliana Wau	Kosong	Petani	1,000,000	2	63
33	April Defi Ndruru	Kosong	Petani	1,000,000	7	60
34	Dika Sartika	Kosong	Petani	500,000	2	60
35	Agusliaman Telaumbanua	Kab	Petani	1,000,000	5	58
36	Fitriana Nasution	Kosong	Lainnya	2,000,000	4	55
37	Muhammad Febryanto	Kosong	Karyawan Swasta	4,500,000	6	54
38	Andrean Harahap	Kosong	Petani	1,000,000	3	54
39	Lasri Everyday Sihotang	Kosong	Petani	1,000,000	5	34
40	Rizky Maulana	Kosong	Petani	1,000,000	4	29
41	Anggy Editya Pratama	Kab	Wiraswasta	4,000,000	3	23
42	Wiyuda Pratama	Kosong	Pensiunan	1,000,000	1	43

No	Nama	Prestasi Sekolah	Pekerjaan Orang Tua	Penghasilan Orang Tua	Jumlah Tanggungan	Hasil Ujian
	Mahararikan		BUMN			
43	Cici Alfiani Pradika Dita	Kab	Karyawan Swasta	2,500,000	2	37
44	Tantri Cepti Buana	Kosong	Buruh Harian Lepas	1,500,000	1	20
45	Tiara Syahri Hsb	Kosong	Buruh Harian Lepas	1,000,000	3	20
46	Rasiddin Rambe	Kab	Petani	1,000,000	6	38
47	Miranda Putri	Kab	Karyawan Swasta	2,400,000	1	24
48	Khairunnisah	Kosong	Petani	1,000,000	2	25
49	Ari Pradana	Kosong	Petani	700,000	3	40
50	Diana Prata Yanti Simangungsong	Kosong	Wiraswasta	1,000,000	1	43
51	Tri Rizky Ananda	Kosong	Tidak Bekerja	1,000,000	5	47
52	M. Andika HSb	Kab	Karyawan Swasta	2,500,000	2	41
53	Fathul Fauzan	Kosong	Karyawan Swasta	10,000,000	5	43
54	Nicolas Cage Sitanggang	Kosong	Petani	1,500,000	3	54

Tabel 1 diatas terdiri atas beberapa attribute yaitu nama menjelaskan tentang nama calon mahasiswa, prestasi sekolah menjelaskan tentang prestasi calon mahasiswa pada saat sekolah, pekerjaan orang tua menjelaskan tentang pekerjaan orang tua calon mahasiswa, penghasilan orang tua menjelaskan tentang jumlah penghasilan orang tua mahasiswa, jumlah tanggungan menjelaskan jumlah tanggungan yang ditanggung oleh orang tua calon mahasiwa dan hasil ujian menjelaskan nilai yang diperoleh calon mahasiswa pada saat mengikuti seleksi mandiri.

Berhubung pada algoritma K-Medoids data yang diproses hanya dalam bentuk numerik, maka data dalam bentuk kategori harus dirubah dalam bentuk numerik seperti pada tabel 2 dan tabel 3 berikut:

Tabel 2. Tabel Prestasi Sekolah

Prestasi	Numerik
Kosong	1
Kab/ Kota	2
Provinsi	3

Tabel 3. Tabel Prestasi Sekolah

Pekerjaan Orang Tua	Numerik
Lainnya	1
Tidak Bekerja	2
Buruh Harian Lepas	3
Petani	4
Karyawan Swasta	5
Wiraswasta	6
Pensiunan BUMN	7

Berdasarkan ketentuan pada tabel 2 dan tabel 3 untuk perubahan data dari kategori ke numerik, maka datanya berubah seperti pada tabel 4 berikut:

Tabel 4. Data Dalam Bentuk Numerik

No	Nama	Prestasi Sekolah	Pekerjaan Orang Tua	Penghasilan Orang Tua	Jumlah Tanggungan	Hasil Ujian
1	Agus Minta Riang Zega	1	4	1,000,000	5	85
2	Desika Marbun	1	4	1,500,000	1	91
3	Amran Saleh Harahap	1	5	1,000,000	2	90
4	Anro Tamba	1	4	1,000,000	3	86
5	Kristina Lenti Kristie Dakhi	1	4	1,000,000	3	83
6	Kevin Maulana Putra	2	1	1,000,000	8	82
7	Ibnu Prinanda	1	1	1,000,000	1	81

No	Nama	Prestasi Sekolah	Pekerjaan Orang Tua	Penghasilan Orang Tua	Jumlah Tanggungan	Hasil Ujian
8	Bobo Firman Zega	1	4	1,000,000	3	81
9	Patar Situmorang	2	4	1,000,000	5	80
10	Susi Mardiana Giawa	2	4	1,200,000	5	80
11	Ritrit Yanto Zai	1	4	1,000,000	4	80
12	Rico Albert Andika Saragih	2	4	1,000,000	4	78
13	Mei Fides Purnawati Zega	3	4	1,000,000	4	77
14	Miko Putra Haposan Tinambunan	2	1	500,000	3	77
15	Elhanan Kris Daniel Simanjuntak	1	1	1,000,000	3	77
16	Putri Patrisia Purba	1	6	2,000,000	3	76
17	Justari Zega	1	4	1,000,000	4	76
18	Enzel Febrianti Telaumbanua	1	4	1,000,000	5	76
19	Friska Agusnayanti Zebua	1	4	500,000	3	74
20	Devi Ritonga	2	4	1,000,000	4	74
21	Hardi Giat Sofian Manalu	1	4	1,000,000	4	73
22	Walderina Situmeang	1	4	2,000,000	4	73
23	Rani Selvira	1	4	1,000,000	3	72
24	Ribka Tasya Simarmata	1	4	2,000,000	5	71
25	Athillah Putra Windra	3	1	2,250,000	5	70
26	Najwa Saufa Silmi	1	4	750,000	6	69
27	Otomosi Gulo	1	4	1,000,000	5	69
28	Tetikariani Halawa	1	4	1,000,000	3	66
29	Ampuni Waruwu	1	4	1,000,000	5	65
30	Edarni Waruwu	1	4	750,000	3	64
31	Lasria Mandalahi	1	4	500,000	4	63
32	Elgayuliana Wau	1	4	1,000,000	2	63
33	April Defi Ndruru	1	4	1,000,000	7	60
34	Dika Sartika	1	4	500,000	2	60
35	Agusliaman Telaumbanua	2	4	1,000,000	5	58
36	Fitriana Nasution	1	1	2,000,000	4	55
37	Muhammad Febryanto	1	5	4,500,000	6	54
38	Andrean Harahap	1	4	1,000,000	3	54
39	Lasri Everyday Sihotang	1	4	1,000,000	5	34
40	Rizky Maulana	1	4	1,000,000	4	29
41	Anggy Editya Pratama	2	6	4,000,000	3	23
42	Wiyuda Pratama Maharaikan	1	7	1,000,000	1	43
43	Cici Alfiani Pradika Dita	2	5	2,500,000	2	37
44	Tantri Cepti Buana	1	3	1,500,000	1	20
45	Tiara Syahri Hsb	1	3	1,000,000	3	20
46	Rasiddin Rambe	2	4	1,000,000	6	38
47	Miranda Putri	2	5	2,400,000	1	24
48	Khairunnisah	1	4	1,000,000	2	25
49	Ari Pradana	1	4	700,000	3	40
50	Diana Prata Yanti Simangungsong	1	6	1,000,000	1	43
51	Tri Rizky Ananda	1	2	1,000,000	5	47
52	M. Andika HSb	2	5	2,500,000	2	41
53	Fathul Fauzan	1	5	10,000,000	5	43
54	Nicolas Cage Sitanggung	1	4	1,500,000	3	54

Langkah selanjutnya adalah melakukan normalisasi data tersebut diatas dengan model max min dengan tujuan untuk menghindari ketimpangan pengaruh antar kriteria attribute pada data. Setelah data dinormalisasikan maka terjadi perubahan pada nilai attribute seperti pada tabel 5 berikut:

Tabel 5. Data Seleksi KIP Kuliah Merdeka Setelah Normalisasi

No	Nama	Prestasi Sekolah	Pekerjaan Orang Tua	Penghasilan Orang Tua	Jumlah Tanggungan	Hasil Ujian	Total
1	Agus Minta Riang Zega	0	0.5	0.05	0.57	0.92	2,04
2	Desika Marbun	0	0.5	0.11	0.00	1.00	1,61
3	Amran Saleh Harahap	0	0.67	0.05	0.14	0.99	1,85
4	Anro Tamba	0	0.5	0.05	0.29	0.93	1,77
5	Kristina Lenti Kristie Dakhi	0	0.5	0.05	0.29	0.89	1,73
6	Kevin Maulana Putra	0.5	0	0.05	1.00	0.87	2,43
7	Ibnu Prinanda	0	0	0.05	0.00	0.86	0,91
8	Bobo Firman Zega	0	0.5	0.05	0.29	0.86	1,70
9	Patar Situmorang	0.5	0.5	0.05	0.57	0.85	2,47
10	Susi Mardiana Giawa	0.5	0.5	0.07	0.57	0.85	2,49
11	Ritrit Yanto Zai	0	0.5	0.05	0.43	0.85	1,83
12	Rico Albert Andika Saragih	0.5	0.5	0.05	0.43	0.82	2,30
13	Mei Fides Purnawati Zega	1	0.5	0.05	0.43	0.80	2,78
14	Miko Putra Haposan Tinambunan	0.5	0	0	0.29	0.80	1,59
15	Elhanan Kris Daniel Simanjuntak	0	0	0.05	0.29	0.80	1,14
16	Putri Patrisia Purba	0	0.83	0.16	0.29	0.79	2,07
17	Justari Zega	0	0.5	0.05	0.43	0.79	1,77
18	Enzel Febrianti Telaumbanua	0	0.5	0.05	0.57	0.79	1,91
19	Friska Agusnayanti Zebua	0	0.5	0	0.29	0.76	1,55
20	Devi Ritonga	0.5	0.5	0.05	0.43	0.76	2,24
21	Hardi Giat Sofian Manalu	0	0.5	0.05	0.43	0.75	1,73
22	Walderina Situmeang	0	0.5	0.16	0.43	0.75	1,83
23	Rani Selvira	0	0.5	0.05	0.29	0.73	1,57
24	Ribka Tasya Simarmata	0	0.5	0.16	0.57	0.72	1,95
25	Athillah Putra Windra	1	0	0.18	0.57	0.70	2,46
26	Najwa Saufa Silmi	0	0.5	0.03	0.71	0.69	1,93
27	Otomosi Gulo	0	0.5	0.05	0.57	0.69	1,81
28	Tetikariani Halawa	0	0.5	0.05	0.29	0.65	1,49
29	Ampuni Waruwu	0	0.5	0.05	0.57	0.63	1,76
30	Edarni Waruwu	0	0.5	0.03	0.29	0.62	1,43
31	Lasria Mandalahi	0	0.5	0	0.43	0.61	1,53
32	Elgayuliana Wau	0	0.5	0.05	0.14	0.61	1,30
33	April Defi Ndruru	0	0.5	0.05	0.86	0.56	1,97
34	Dika Sartika	0	0.5	0	0.14	0.56	1,21
35	Agusliaman Telaumbanua	0.5	0.5	0.05	0.57	0.54	2,16
36	Fitriana Nasution	0	0	0.16	0.43	0.49	1,08
37	Muhammad Febryanto	0	0.67	0.42	0.71	0.48	2,28
38	Andrean Harahap	0	0.5	0.05	0.29	0.48	1,32
39	Lasri Everyday Sihotang	0	0.5	0.05	0.57	0.20	1,32
40	Rizky Maulana	0	0.5	0.05	0.43	0.13	1,11
41	Anggy Editya Pratama	0.5	0.83	0.37	0.29	0.04	2,03
42	Wiyuda Pratama Maharaikan	0	1.00	0.05	0.00	0.32	1,38
43	Cici Alfiani Pradika	0.5	0.67	0.21	0.14	0.24	1,76

No	Nama	Prestasi Sekolah	Pekerjaan Orang Tua	Penghasilan Orang Tua	Jumlah Tanggungan	Hasil Ujian	Total
	Dita						
44	Tantri Cepti Buana	0	0.33	0.11	0.00	0.00	0,44
45	Tiara Syahri Hsb	0	0.33	0.05	0.29	0.00	0,67
46	Rasiddin Rambe	0.5	0.5	0.05	0.71	0.25	2,02
47	Miranda Putri	0.5	0.67	0.2	0.00	0.06	1,42
48	Khairunnisah	0	0.5	0.05	0.14	0.07	0,77
49	Ari Pradana	0	0.5	0.02	0.29	0.28	1,09
50	Diana Prata Yanti Simangungsong	0	0.83	0.05	0.00	0.32	1,21
51	Tri Rizky Ananda	0	0.17	0.05	0.57	0.38	1,17
52	M. Andika HSb	0.5	0.67	0.21	0.14	0.30	1,82
53	Fathul Fauzan	0	0.67	1	0.57	0.32	2,56
54	Nicolas Cage Sitanggung	0	0.5	0.11	0.29	0.48	1,37

Tabel 5 merupakan data yang telah berubah, dimana nilai setiap kriteria sudah lebih sederhana yaitu minimal 0 dan maksimal 1 sehingga lebih mudah untuk dilakukan proses komputasi.

3.1 K-Medoids dengan Centroid Max Min

K-Medoids dengan Centroid Max Min merupakan penentuan nilai centroid berdasarkan nilai paling tinggi (max) dan nilai paling rendah (min) dari total penjumlahan nilai attribite kriteria data. Attribute kriteria terdiri atas 2 (dua) cluster yaitu data mahasiswa calon penerima KIP Kuliah merdeka yang layak menerima atau tidak layak menerima KIP Kuliah Merdeka, maka jumlah cluster (K)=2.

a. Iterasi I

Nilai centroid (medoids) atau titik pusat diambil max min

Pusat medoids 1 = 0, 0.33, 0.11, 0, 0 (44.Tantri Cepti Buana)

Pusat medoids 2 = 1, 0.5, 0.05, 0.43, 0.80 (13.Mei Fides Purnawati Zega)

Selanjutnya dilakukan perhitungan jarak dengan menggunakan rumus Euclidean Distance, dengan hasil sebagai berikut:

Tabel 6. Hasil Cluster Iterasi I

Cluster	Jumlah Anggota	Anggota (No)
I	43	2,3,4,5,7,8,11,15,16,17,18,19,21,22,23,24,26,27,28,29,30,31,32,33,34,36,37,38,39,40,41,42,43,44,45,47,48,49,50,51,52,53,54
II	11	1,6,9,10,12,13,14,20,25,35,46

Dengan total nilai cost adalah 39,81

b. Iterasi II

Pusat medoids 1 = 0, 0.33, 0.05, 0.29, 0.00 (45.Tiara Syahri Hsb)

Pusat medoids 2 = 0, 0.67, 1.00, 0.57, 0.32 (53.Fathul Fauzan)

Dilakukan perhitungan jarak dengan menggunakan rumus Euclidean Distance. Adapun hasil clusternya sebagai berikut:

Tabel 7. Hasil Cluster Iterasi II

Cluster	Jumlah Anggota	Anggota
I	52	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,38,39,40,41,42,43,44,45,46,47,48,49, 50,51,52,54
II	2	37,53

Dengan total nilai cost adalah 40,31. Karena nilai cost iterasi II lebih besar dari pada iterasi I maka proses iterasi berhenti.

3.2 K-Medoids dengan Centroid Acak

Untuk percobaan selanjutnya, mengclusterkan data calon mahasiswa yang layak menerima dan yang tidak layak dengan nilai centroid dipilih secara acak dengan menggunakan bahasa Python. Adapun nilai centroid yang diperoleh secara acak / random adalah sebagai berikut:

Pusat medoids 1 = 0, 0.5, 0.05, 0.43, 0.75 (21.Hardi Giat Sofian Manalu.)

Pusat medoids 2 = 0, 0.5, 0.02, 0.29, 0.28 (49.Ari Pradana)

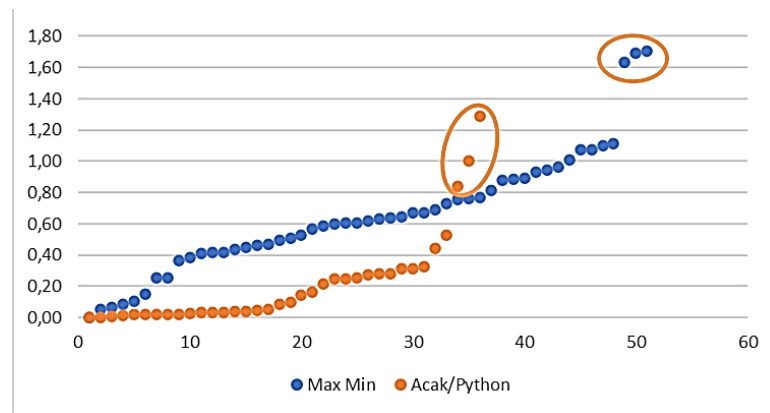
Adapun hasil clusternya adalah sebagai berikut:

Tabel 8. Hasil Cluster dengan Centroid Acak

Cluster	Jumlah Anggota	Anggota
I	36	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,35,36,37
II	18	34,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54

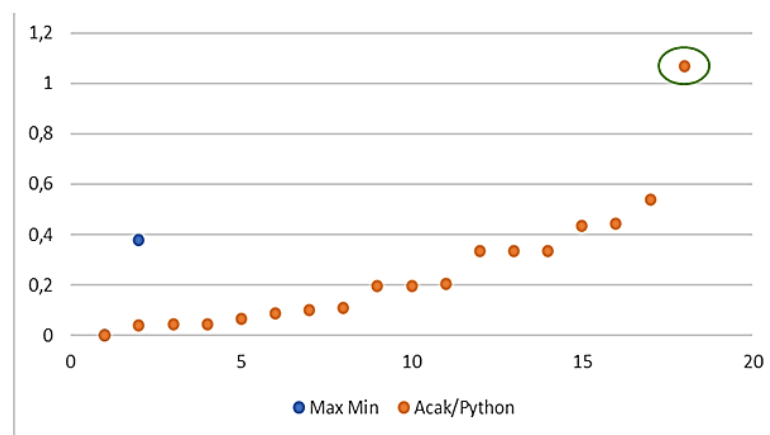
3.3 Deteksi Outlier Dengan Box Plot

Setelah diperoleh hasil clustering pengelompokkan calon mahasiswa yang layak dan tidak layak menerima KIP kuliah Merdeka yang terdiri atas 2 cluster maka selanjutnya dilakukan perhitungan nilai SSE. Setelah memperoleh nilai SSE dari masing-masing cluster maka dilakukan deteksi outlier data dari masing-masing cluster yang terbentuk berdasarkan nilai SSE nya dengan menggunakan metode box plot. Berikut hasil outlier dalam bentuk grafik nilai SSE cluster 1 :



Gambar 2. Grafik Outlier Cluster 1 Secara Max Min dan Acak

Gambar 2 merupakan grafik outlier cluster 1 secara max min dan acak dimana terdapat 3 data yang outlier yaitu Kevin Maulana Putra (No.6), Mei Fides Purnawati Zega (No.13), dan Athillah Putra Windra (No.25), dikarenakan data tersebut melampaui batas bawah dan batas atas yaitu -0,295 sampai dengan 1,585 sedangkan jarak antar kuartil adalah 0,47. Berikut hasil outlier dalam bentuk grafik nilai SSE cluster 2 :



Gambar 3. Grafik Outlier Cluster 2 Secara Max Min dan Acak

Gambar 3 merupakan grafik outlier cluster 2 secara max min dan acak / Python dimana terdapat satu outlier yaitu yaitu Fathul Fauzan (No.53), karena pada data tersebut khususnya pada clustering data dengan pemilihan nilai centroid secara acak.

Hasil clustering data dengan pemilihan nilai centroid max min dan secara acak memiliki kesamaan hasil sebanyak 35 data, sehingga nilai akurasi antara kedua teknik pemilihan centroid tersebut $(35/54) \times 100\%$ yaitu 64,81%. Untuk data outliernya pada pemilihan centroid secara max min memiliki 3 data outlier, sedangkan secara acak / python terdiri atas 4 data yang outlier, sehingga akurasi outliernya adalah $3/4 \times 100\% = 75\%$.

4. KESIMPULAN

Kesimpulan yang dapat diambil pada penelitian ini adalah algoritma K-Medoids berhasil dalam mengelompokkan data mahasiswa ke dalam 2 cluster yaitu cluster I merupakan cluster yang layak dan cluster II merupakan cluster yang tidak layak menerima beasiswa KIP kuliah merdeka, hasil dari pengelompokkan data calon mahasiswa penerima beasiswa KIP kuliah merdeka dengan algoritma k-Medoids adalah cluster pertama memiliki 52 anggota mahasiswa, dan cluster kedua memiliki 2 anggota mahasiswa, penerapan metode boxplot untuk mendeteksi outlier hasil cluster pengelompokkan calon mahasiswa yaitu pada cluster pertama terdapat 3 data outlier dan pada cluster kedua tidak terdapat data outlier, outlier yang teridentifikasi adalah data yang berada jauh dari medoid cluster, menunjukkan anomali atau karakteristik unik yang berbeda dari mayoritas data dalam cluster tersebut, perbandingan hasil cluster dengan pemilihan centroid secara max min dan secara acak memiliki kesamaan hasil cluster / akurasi hasil sebesar 64,81%, dan akurasi outliernya sebesar 75%.

REFERENCES

- [1] Puslapdik, *PEDOMAN PENDAFTARAN KARTU INDONESIA PINTAR KULIAH - KIP KULIAH MERDEKA 2022*. Jakarta: Sekretaris Jenderal Kemendikbudristek, 2022. [Online]. Available: https://kip-kuliah.kemdikbud.go.id/uploads/Pedoman-Pendaftaran-KIP-K-2022-ver-20220202---final_cd9b5e.pdf
- [2] G. Pu, L. Wang, J. Shen, and F. Dong, "A hybrid unsupervised clustering-based anomaly detection method," *Tsinghua Sci. Technol.*, vol. 26, no. 2, pp. 146–153, 2021, doi: 10.26599/TST.2019.9010051.
- [3] M. Menhendry, A. Alfian, E. Adril, J. Junaidi, and Z. Zuhendri, "Penerapan Quality Control pada Souvenir Logo PNP dengan Metode Box Plot dan Six Sigma," *J. Tek. Mesin*, vol. 14, no. 2, pp. 112–117, 2021, doi: 10.30630/jtm.14.2.674.
- [4] D. F. Pramesti, Lahan, M. Tanzil Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 723–732, 2017, doi: 10.1109/EUMC.2008.4751704.
- [5] M. Y. Darsyah, "Penggunaan Stem and Leaf dan Boxplot untuk Analisis Data," *J. Pendidik. Mat.*, vol. 1, no. 1, pp. 55–67, 2014, [Online]. Available: <http://103.97.100.145/index.php/JPMat/article/view/1045/1093>
- [6] P. R. Sihombing, S. Suryadiningrat, D. A. Sunarjo, and Y. P. A. C. Yuda, "Identifikasi Data Outlier (Pencilan) dan Kenormalan Data Pada Data Univariat serta Alternatif Penyelesaiannya," *J. Ekon. Dan Stat. Indones.*, vol. 2, no. 3, pp. 307–316, 2023, doi: 10.11594/jesi.02.03.07.
- [7] S. Bu'ulolo;Efori, Mesran, Hasibuan;Nelly Astuti, Utomo;Aripin;Soeb, Putro Utomo, *Big Data Analysis dengan Python untuk Perguruan Tinggi*, I. Yogyakarta, 2023.
- [8] Buulolo;Efori, *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: deepublish, 2020.
- [9] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.
- [10] B. Wira, A. E. Budianto, and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang," *RAINSTEK J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 53–68, 2019, doi: 10.21067/jtst.v1i3.3046.
- [11] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm for Big Data," in *Physics Procedia*, 2016. doi: 10.1016/j.procs.2016.02.095.
- [12] E. Buulolo and R. Syahputra, "Implementasi Algoritma Clustering K-Means Untuk Mengelompokkan Mahasiswa Baru Yang Berpotensi (Studi Kasus : Stmik Budi Darma)," vol. 2, no. September, pp. 17–24, 2019.
- [13] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [14] I. Kamila *et al.*, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," vol. 5, no. 1, pp. 119–125, 2019.
- [15] A. Bhat, "K-Medoids Clustering Using Partitioning Around Medoids for Performing Face Recognition," *Int. J. Soft Comput. Math. Control*, 2014, doi: 10.14810/ijscmc.2014.3301.
- [16] W. A. Triyanto, "ALGORITMA K-MEDOIDS UNTUK PENENTUAN STRATEGI PEMASARAN PRODUK," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, 2015, doi: 10.24176/simet.v6i1.254.
- [17] C. Zhang *et al.*, "Unsupervised Anomaly Detection Based on Deep Autoencoding and Clustering," *Secur. Commun. Networks*, vol. 2021, 2021, doi: 10.1155/2021/7389943.
- [18] B. Tang and H. He, "A local density-based approach for outlier detection," *Neurocomputing*, vol. 241, pp. 171–180, 2017, doi: 10.1016/j.neucom.2017.02.039.
- [19] I. S. Rufiana, W. Wahyudi, and D. A. Nurhidayah, "Misinterpretasi Mahasiswa pada Representasi Histogram dan Box Plot," *ANARGYA J. Ilm. Pendidik. Mat.*, vol. 3, no. 2, pp. 82–87, 2020, doi: 10.24176/anargya.v3i2.5324.
- [20] N. Puspitasari, G. Lempas, H. Hamdani, H. Haviuddin, and A. Septiarini, "Perbandingan Algoritma K-Means dan Algoritma K-Medoids Pada Kasus Covid-19 di Indonesia," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 2015–2027, 2023, doi: 10.47065/bits.v4i4.2994.