

PENERAPAN METODE *K-MEDOIDS* UNTUK PENGELOMPOKAN MAHASISWA BERPOTENSI *DROP OUT*

Syamsul Bahri^{*1}, Dwi Marisa Midyanti²

^{1,2}Universitas Tanjungpura, Pontianak

Email: ¹syamsul.bahri@siskom.untan.ac.id, ²dwi.marisa@siskom.untan.ac.id

^{*}Penulis Korespondensi

(Naskah masuk: 26 Oktober 2022, diterima untuk diterbitkan: 28 Februari 2023)

Abstrak

Drop out merupakan penghentian atau pemutusan hubungan studi mahasiswa di perguruan tinggi, hal ini disebabkan oleh beberapa hal yang telah ditentukan oleh universitas. Perguruan tinggi dapat membuat kebijakan guna meminimalkan jumlah mahasiswa *drop out* dengan mengidentifikasi mahasiswa yang berisiko di tahap awal pendidikan. Mahasiswa *drop out* dapat diprediksi melalui beberapa proses memperoleh pola atau pengetahuan dari kumpulan data yang disebut *data mining*. *Data mining* melakukan analisis data yang telah ada dalam basis data guna penyelesaian masalah. Analisis yang dilakukan salah satunya dengan metode *clustering*. Cara kerja *clustering* dengan mengelompokkan data atau objek ke dalam *cluster* (kelompok). Penelitian ini menerapkan K-Medoids untuk melakukan pengelompokan kategori mahasiswa berpotensi *drop out*. K-Medoids merupakan salah satu metode *data mining* yang dapat menyelesaikan permasalahan *clustering*. Pada algoritma ini representasi sebuah *cluster* menggunakan objek pada sekumpulan objek. Pengukuran hasil K-Medoids menggunakan *Silhouette Coefficient* yang berfungsi untuk mengevaluasi jarak kedekatan antar data dalam satu *cluster*. Setelah dilakukan *clustering* menggunakan algoritma K-Medoids didapatkan hasil evaluasi *Silhouette Coefficient* terbaik sebesar 0.39415227406014575 dengan jumlah *cluster* 2. Hasil *cluster* dari 389 data, didapat 3 atribut yang memiliki rentang nilai yang berbeda antar *cluster* yaitu IPK, IP Semester 1, dan Status Beasiswa, ketiga atribut ini menjadi pembeda yang membedakan antar *cluster*.

Kata kunci: Mahasiswa, Drop out, Data Mining, Clustering, K-Medoids, Silhouette Coefficient

APPLICATION OF K-MEDOIDS METHOD FOR DROPOUT POTENTIAL STUDENT GROUPING

Abstract

Dropout is the termination of student studies in college, this is caused by several things that have been determined by the university. Universities can make policies to minimize the number of students dropping out by identifying students at risk in the early stages of education. Dropout students can be predicted through several processes of obtaining patterns or knowledge from data sets called data mining. Data mining analyzes data that already exists in the database to solve problems. One of the analyzes carried out is the clustering method. How clustering works by grouping data or objects into clusters (groups). This study applies K-Medoids to classify categories of students who have the potential to drop out. K-Medoids is a data mining method that can solve clustering problems. In this algorithm, the representation of a cluster uses objects in a set of objects. Measurement of the results of K-Medoids using the Silhouette Coefficient which serves to evaluate the proximity between data in one cluster. After clustering using the K-Medoids algorithm, the best Silhouette Coefficient evaluation results are 0.39415227406014575 with the number of clusters 2. The cluster results from 389 data, obtained 3 attributes that have different value ranges between clusters, namely GPA, IP Semester 1, and Scholarship Status, all three attributes are distinguishing feature between clusters.

Keywords: Student, Dropout, Data Mining, Clustering, K-Medoids, Silhouette Coefficient

1. PENDAHULUAN

Perguruan tinggi merupakan penyelenggara pendidikan akademik bagi mahasiswa. Perguruan tinggi diharapkan dapat menyelenggarakan

pendidikan yang berkualitas agar menghasilkan lulusan yang cerdas dan kreatif guna mendukung terlaksananya pembangunan nasional yang berkualitas. Pendidikan pada perguruan tinggi dapat

dinyatakan berhasil jika telah mencapai tujuan pendidikan nasional, dimana proses belajar mengajar yang terjadi secara efektif dan efisien. Kesuksesan atau prestasi belajar mahasiswa seringkali dilihat sebagai keunggulan dan kesuksesan perguruan tinggi dalam menyelenggarakan proses pendidikan. Sebaliknya kegagalan atau rendahnya kualitas mahasiswa sering dianggap bahwa perguruan tinggi tidak mampu menyelenggarakan proses pendidikan (Nurhayati, Kusrini and Luthfi, 2015).

Drop out (DO) merupakan penghentian atau pemutusan hubungan studi mahasiswa di perguruan tinggi, hal ini disebabkan oleh beberapa hal yang telah ditentukan oleh universitas. Terdapat beberapa alasan dilakukan *drop out* terhadap mahasiswa, pertama karena tidak aktif selama 2 tahun akademik secara berturut-turut, kedua karena prestasi akademiknya rendah (belum menyelesaikan 45 sks di empat semester pertama serta IPK kurang dari 2.00), dan ketiga dikarenakan melewati masa maksimal studi (UNTAN, 2021).

Program Studi Rekayasa Sistem Komputer (Resiskom) merupakan salah satu program studi yang ada di Fakultas MIPA Universitas Tanjungpura (Untan). Setiap tahunnya prodi Resiskom menerima banyak mahasiswa, bahkan beberapa tahun terakhir lebih dari seratus mahasiswa masuk ke prodi Resiskom tiap angkatan. Akan tetapi seiring tingginya minat mahasiswa yang bergabung di prodi Resiskom, diikuti pula dengan tingginya mahasiswa yang berhenti ataupun *drop out* di tengah masa perkuliahan.

Banyaknya jumlah mahasiswa yang *drop out* di perguruan tinggi dapat diminimalkan melalui kebijakan pimpinan untuk memberikan arahan dan mencegah mahasiswa dari kemungkinan *drop out*. Hal ini dilakukan dengan cara mendeteksi mahasiswa yang berisiko di tahap awal pendidikan. Langkah ini sangat penting untuk dilakukan guna menghindarkan mahasiswa dari *drop out* (Dekker, Pechenizkiy and Vleeshouwers, 2009). Jurusan atau departemen sebagai penyelenggara pendidikan dapat memberikan pembinaan dan pengarahan kepada mahasiswa yang membutuhkan.

Oleh karena itu, kajian ataupun prediksi mahasiswa yang memiliki kemungkinan *drop out* perlu dilakukan, sehingga dapat dijadikan sebagai informasi yang bermanfaat untuk memprediksi jumlah mahasiswa *drop out* pada tahun-tahun berikutnya dan tentunya dapat mengurangi jumlah *drop out* mahasiswa. Prediksi *drop out* mahasiswa juga dapat memberikan opsi bagi pihak institusi dalam pengambilan kebijakan. Faktor penyebab *drop out* mahasiswa dapat dianalisis menggunakan algoritma tertentu untuk mendapatkan parameter terbaik guna memprediksi mahasiswa yang berpotensi *drop out*.

Proses mengelompokkan mahasiswa dengan perhitungan manual membutuhkan waktu yang tidak sedikit. *Drop out* dapat diprediksi melalui beberapa

proses memperoleh pola atau pengetahuan dari kumpulan data yang disebut *data mining*. *Data mining* melakukan analisis data yang telah ada dalam basis data guna penyelesaian masalah. Analisis yang dilakukan salah satunya dengan metode *clustering*. *Clustering* mengelompokkan data atau objek ke dalam beberapa *cluster*. Data pada setiap *cluster* dibuat semirip mungkin, sementara jarak antar *cluster* dibuat sejauh mungkin (Han and Kamber, 2006).

Penelitian prediksi mahasiswa *drop out* pernah diangkat sebelumnya, misalnya penelitian oleh Nurhayati, Kusrini and Luthfi (2015) yang menggunakan algoritma Support Vector Machine. Prediksi dilakukan menggunakan variabel masukan berupa data individu, evaluasi IP serta IPK mahasiswa. Faktor yang menjadi penyebab mahasiswa *drop out* yaitu akademik dan non akademik. Penelitian lain oleh Muhamad, Windarto and Suhada (2019) yaitu klasifikasi potensi siswa *drop out* menerapkan metode C4.5 menghasilkan 9 pola dan tingkat akurasi sebesar 96,15% serta nilai *Area Under the ROC Curve* sebesar 0,998. Penelitian oleh Hidayat, Purwitasari and Ginardi (2013) tentang analisis prediksi DO mahasiswa dalam *educational data mining* menggunakan jaringan saraf tiruan. Hasil penelitian yaitu model yang diusulkan dapat digunakan untuk memprediksi kemungkinan *drop out* dan sebesar 98,91% serta tingkat signifikan sensitivitas terbesar yaitu perilaku sosial sebesar 4,737. Sedangkan penelitian ini menerapkan K-Medoids untuk melakukan pengelompokan kategori mahasiswa berpotensi *drop out*.

K-Medoids merupakan salah satu metode *data mining* yang dapat menyelesaikan permasalahan *clustering* atau pengelompokan. Pada algoritma ini representasi sebuah *cluster* menggunakan objek pada sekumpulan objek. Algoritma *K-Medoids* dikenal cukup efisien pada dataset kecil. Hasil proses *clustering K-Medoids* juga tidak bergantung pada urutan masuk dataset. *K-Medoids* juga dapat mengatasi kekurangan metode *K-means* yang sensitive terhadap *outlier* (Dwilestari et al., 2021).

K-Medoids telah banyak digunakan dalam penelitian, diantaranya digunakan untuk pengelompokan sebaran data mahasiswa baru (Irawan et al., 2020), strategi promosi di Politeknik (Anggerini, 2019), *clustering data penduduk miskin Indonesia* (Dwilestari et al., 2021), *cluster mahasiswa pelamar beasiswa* (Defiyanti, Jajuli and Rohmawati, 2017), pengelompokan desa yang memiliki fasilitas sekolah (Damanik et al., 2019).

Pengukuran hasil K-Medoids menggunakan *Silhouette Coefficient* yang berfungsi untuk mengevaluasi jarak kedekatan antar data dalam satu *cluster*.

2. METODE PENELITIAN

Penelitian ini dimulai dengan identifikasi masalah yang dilakukan melalui pengamatan dan

survey pada instansi terkait, dilanjutkan dengan pengumpulan data, pengolahan data dan pembahasan. Pengolahan data menerapkan konsep *preprocessing* data sebelum data di *clustering*. Tujuan *preprocessing* adalah mengubah data ke suatu format yang menjadikan proses data *mining* lebih mudah dan efektif sesuai kebutuhan. *Preprocessing* juga membantu dalam memperoleh hasil yang lebih akurat, mengurangi waktu untuk penghitungan dalam skala yang besar, dan menghasilkan nilai data yang lebih kecil tanpa merubah informasi di dalamnya.

Penelitian ini merupakan penelitian model eksperimen menggunakan bahasa pemrograman Python. Penelitian ini bertujuan mengamati atribut yang menjadi pembeda antar kluster yang terbentuk, dimana pembeda ini menjadi ciri sebagai indikator potensi mahasiswa yang DO atau tidak. Penelitian ini menggunakan data mahasiswa prodi Rekayasa Sistem Komputer FMIPA Universitas Tanjungpura angkatan 2014-2018 yang berasal dari sistem Siakad, Biro Akademik dan Kemahasiswaan (BAK), serta dari Sub Bagian Akademik Fakultas MIPA Universitas Tanjungpura. Dataset yang dihimpun terdiri dari 390 record data, terdiri dari nim, nama mahasiswa, jalur masuk mahasiswa, ipk, ip semester 1, ip semester 2, status beasiswa, jumlah penghasilan orang tua mahasiswa, dan status pekerjaan mahasiswa.

2.1. Preprocessing Data

Data yang telah dihimpun dari beberapa sumber sebelum dapat digunakan dalam program menggunakan algoritma K-Medoids harus dilakukan *preprocessing* terlebih dahulu. Data tersebut merupakan data seluruh mahasiswa program studi (prodi) Rekayasa Sistem Komputer FMIPA Universitas Tanjungpura angkatan 2014 hingga angkatan 2018 dengan total sebanyak 390 mahasiswa.

Preprocessing data yang dilakukan diantaranya seperti data *cleaning*, data *selection*, dan data *transformation*. Tahapan-tahapan ini dilakukan agar data tersebut dapat diproses oleh algoritma K-Medoids.

2.2. Data Cleaning

Tujuan data *cleaning* yaitu untuk menghilangkan *noise* data maupun data yang tidak konsisten maupun data yang duplikat. Pada penelitian ini hanya terdapat satu data yang perlu dilakukan proses *cleaning*. Terdapat satu orang mahasiswa yang datanya duplikat yaitu mahasiswa atas nama Aris Fajrianto angkatan 2015 yang datanya ada dua. Oleh karena itu proses *cleaning* yang dilakukan adalah menghapus satu data yang keliru.

2.3. Data Selection

Tujuan data *selection* yaitu memilih atribut apa saja yang relevan dengan penelitian. Pada penelitian ini atribut yang dikumpulkan yaitu nim,

nama mahasiswa, jalur masuk mahasiswa, ipk, ip semester 1, ip semester 2, status beasiswa, jumlah penghasilan orang tua mahasiswa, dan status pekerjaan mahasiswa. Dalam proses ini terdapat beberapa atribut yang perlu dihilangkan dan ada atribut yang perlu diolah untuk menghasilkan atribut yang lain, diantaranya :

- Atribut nim dan nama mahasiswa tidak diperlukan dalam proses *clustering* dikarenakan kedua atribut ini adalah atribut yang hanya menggambarkan identitas tiap record data dan tidak relevan serta tidak sesuai untuk menjadi indikator potensi mahasiswa DO atau tidak. Oleh karena itu, kedua atribut tersebut dihapus atau tidak disertakan dalam proses metode K-Medoids.
- Berdasarkan atribut ip semester 1 dan ip semester 2, dihasilkan sebuah atribut baru yaitu tren ip semester 1 ke semester 2. Atribut baru ini bernilai status progres atau perubahan ip mahasiswa dari semester 1 ke semester 2, apakah naik, turun, atau tetap. Atribut ini cukup penting digunakan dalam memberikan gambaran potensi mahasiswa DO atau tidak. Hal ini dikarenakan apa yang terjadi selama ini di prodi Rekayasa Sistem Komputer menunjukkan mahasiswa yang DO memiliki kecenderungan IP di semester awal lebih rendah atau menurun. Selanjutnya atribut semester 2 dihapuskan atau tidak disertakan dalam proses metode K-Medoids, karena dari awal atribut ini memang disertakan hanya untuk membentuk atribut tren ip, sedangkan atribut ip semester 2 sendiri tidak relevan dan sesuai sebagai indikator potensi mahasiswa DO atau tidak.

2.4. Data Transformation

Tujuan data *transformation* yaitu mengubah data menjadi bentuk yang sesuai agar dapat dilakukan proses *mining* data. Pada penelitian ini, proses *transformation* yang dilakukan yaitu membuat inialisasi data yang berjenis data nominal pada beberapa atribut ke dalam bentuk numerik. Adapun atribut yang dilakukan proses *transformation* yaitu seperti atribut jalur masuk mahasiswa, tren ip semester 1 ke semester 2, status beasiswa, penghasilan orang tua, dan status pekerjaan mahasiswa. Sementara atribut ipk dan ip semester 1 digunakan data asli tanpa perlu dilakukan *transformation* karena sudah dalam bentuk numerik.

Atribut jalur masuk mahasiswa, tren ip semester 1 ke semester 2, status beasiswa, penghasilan orang tua, dan status pekerjaan mahasiswa yang telah dilakukan inialisasi dapat dilihat berturut-turut pada Tabel 1, Tabel 2, Tabel 3, Tabel 4, dan Tabel 5.

Tabel 1. Inialisasi Atribut Jalur Masuk

Inialisasi	Jalur Masuk
0	snmptn
1	sbmptn
2	selok/mandiri

Tabel 2. Inisialisasi Atribut Tren IP 1-2

Inisialisasi	Tren IP 1-2
0	Naik
1	Turun
2	Tetap

Tabel 3. Inisialisasi Atribut Beasiswa

Inisialisasi	Beasiswa
0	Ya
1	Tidak

Tabel 4. Inisialisasi Atribut Penghasilan Orang Tua

Inisialisasi	Penghasilan Ortu
0	Penghasilan \leq 500.000
1	500.000 < Penghasilan \leq 2.000.000
2	2.000.000 < Penghasilan \leq 4.000.000
3	Penghasilan > 4.000.000

Tabel 5. Inisialisasi Atribut Bekerja

Inisialisasi	Bekerja
0	Ya
1	Tidak

2.5. Clustering

Clustering mengelompokkan sekumpulan objek data menjadi beberapa *cluster* atau kelompok, proses ini membuat objek pada suatu *cluster* memiliki persamaan atau kemiripan yang tinggi, akan tetapi sangat berbeda dengan objek dalam *cluster* yang lain. Kemiripan ataupun tidaknya dinilai berdasarkan atribut yang menggambarkan objek, dan sering menyertakan pengukuran jarak atau *measures distance*. Dalam konteks ini, metode *clustering* yang berbeda dapat menghasilkan *cluster* yang berbeda pada kumpulan data yang sama.

Proses *clustering* tidak dikerjakan oleh manusia, namun dengan algoritma *clustering* dan tanpa mengetahui target kelasnya terlebih dulu. Proses ini termasuk dalam kategori *unsupervised learning* yang tidak memiliki data latih. Oleh karena itu data yang ada dikelompokkan menjadi dua bagian atau tiga bagian dan selanjutnya.

Terdapat berbagai pendekatan yang digunakan dalam pengembangan algoritma *clustering*. Dua pendekatan utama yaitu pendekatan hirarki dan partisi. Pendekatan partisi atau *partition-based clustering* mengelompokkan data dengan cara memilah data yang akan dianalisis ke dalam *cluster* yang ada. Salah satu algoritma yang termasuk dalam pendekatan ini yaitu algoritma K-Medoids.

2.6. K-Medoids

K-Medoids atau Partitioning Around Medoids (PAM) atau K-Medians merupakan algoritma clustering sebagai varian dari algoritma K-Means. Kelemahan K-Means yang sensitif terhadap outlier dianggap dapat diatasi oleh K-Medoids, dimana suatu objek dengan nilai yang besar mungkin saja secara

substansi menyimpang dari distribusi data (Han and Kamber, 2006).

Penggunaan medoids bukan merupakan hasil pengamatan mean yang dimiliki oleh setiap cluster. Tujuannya yaitu mengurangi sensitivitas partisi, hal ini terkait nilai ekstrim yang ada pada dataset (Vercellis, 2009).

Objek medoid letaknya terpusat di dalam suatu cluster sehingga kuat terhadap outlier. Cluster dibangun dengan memperhitungkan kedekatan yang dimiliki antara medoid dengan objek non medoid (Setiyawati, 2017).

Seperti algoritma K-Means, algoritma K-Medoids lebih baik diterapkan pada data yang sangat besar. Untuk Data yang berbeda parameternya, sebelum dihitung jaraknya terlebih dahulu lakukan normalisasi data agar data seimbang.

Cara kerja K-Medoids adalah sebagai berikut (Han and Kamber, 2006) :

1. Tentukan jumlah pusat *cluster* sebanyak k (jumlah *cluster*)
2. Tentukan medoid awal secara acak sebanyak k dari n data
3. Hitung jarak objek ke tiap medoid yang telah dipilih, gunakan *Euclidean Distance*. Hitung Total jarak terdekat pada medoid awal (a) menggunakan Persamaan (1).

$$\sqrt{(x_i - X)^2 + \dots + (y_n - Y)^2} \quad (1)$$

4. Pilih secara acak medoid baru, lalu hitung jarak objek ke tiap medoid yang telah dipilih, gunakan *Euclidean Distance*. Hitung Total jarak terdekat pada medoid baru (b)
5. Hitung total simpangan (S) menggunakan Persamaan (2).

$$S = b - a \quad (2)$$

a : jumlah jarak terdekat antara objek ke medoid awal,

b : jumlah jarak terdekat antara objek ke medoid baru,

Jika $S < 0$, maka tukar objek dengan data untuk menghasilkan sekumpulan k baru sebagai medoid

Ulangi 3-5, sampai tidak terjadi lagi perubahan *medoids*, sehingga dihasilkan sejumlah *cluster* beserta anggotanya.

2.7. Silhouette Coefficient

Silhouette coefficient adalah metode untuk melakukan evaluasi sebuah *cluster* agar dapat dilihat kedekatan antar objek dalam satu *cluster*, serta kedekatan antar objek dalam *cluster* yang berbeda. Adapun nilai *silhouette coefficient* dalam rentang -1 sampai 1. Nilai mendekati 1 mengartikan hasil *clustering* yang didapat lebih baik. Sebaliknya jika nilai yang didapat mendekati nilai -1, maka hasil yang didapat kurang baik (Rousseeuw, 1987). Adapun tahapan untuk menghitung nilai *silhouette coefficient* adalah sebagai berikut:

1. Hitung rata-rata jarak dari suatu data i dengan semua data lain yang berada dalam satu cluster menggunakan Persamaan (3).

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (3)$$

Dimana,

j : data lain dalam satu cluster A

$d(i, j)$: jarak antara data i dengan j

2. Hitung rata-rata jarak dari suatu data i dengan semua data lain yang berada dalam satu cluster menggunakan Persamaan (4).

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (4)$$

Dimana,

$d(i, C)$: jarak rata-rata data i dengan semua objek pada cluster lain C , dimana $A \neq C$. Lalu hitung $b(i)$ yaitu nilai minimum $d(i, C)$ menggunakan Persamaan (5).

$$b(i) = \min_{C \neq A} d(i, C) \quad (5)$$

3. Hitung nilai *silhouette coefficient* menggunakan Persamaan (6).

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$

Dimana,

$s(i)$: nilai *silhouette* data ke i

Tabel 6 menunjukkan kriteria pengukuran nilai *silhouette coefficient* (Rousseeuw, 1987).

Tabel 6. Kriteria Silhouette Coefficient	
Silhouette Coefficient	Kriteria Penilaian
$0.7 < SC \leq 1.0$	Stronge Structure
$0.5 < SC \leq 0.7$	Medium Structure
$0.25 < SC \leq 0.5$	Weak Structure
$SC \leq 0.25$	No Structure

3. HASIL DAN PEMBAHASAN

Penelitian ini mengamati *range* nilai atribut yang akan menjadi pembeda tiap *cluster* yang terbentuk. Pembeda ini menjadi ciri sebagai indikator potensi mahasiswa yang DO atau tidak.

Proses implementasi yang dilakukan pertama yaitu menggunakan library *scikit-learn* untuk membangun model K-Medoids. Data yang diolah oleh metode K-Medoids diantaranya dapat dilihat pada Tabel 7.

Dalam pengolahan program *Clustering* pada data mahasiswa prodi Rekayasa Sistem Komputer untuk menentukan potensi *Drop Out* atau tidak, menggunakan bahasa pemrograman Python. Setelah dataset sudah disiapkan langkah pertama yang dilakukan yaitu mengimport library yang dibutuhkan.

Dataset yang disimpan dalam sebuah file txt diupload ke aplikasi. Dataset tersebut disimpan ke dalam variabel menggunakan fungsi *getfromtxt* yang tersedia dalam library *Numpy*.

Tabel 7. Dataset Penelitian

No	X1	X2	X3	X4	X5	X6	X7
1	1	3.24	0	0	0	1	1
2	1	2.76	2.14	0	0	1	1
3	1	2.97	2.82	1	0	0	0
4	1	3.15	2.95	1	0	1	1
5	1	2.87	2.45	0	0	0	0
6	1	3.27	3.5	1	0	1	1
7	1	3.82	4	1	0	1	1
8	1	2.65	2.14	0	0	1	1
...
383	2	3.28	2.84	1	1	3	1
384	2	0.58	1.18	1	1	1	1
385	2	0.85	1.74	1	1	1	1
386	2	3.48	1.58	1	1	2	1
387	2	2.22	2.58	1	1	3	1
388	2	1.84	1.84	1	1	1	1
389	2	3.46	2.18	1	1	2	1

Data yang telah diupload dan disimpan dalam variabel Python masih bertipe string, sehingga perlu dilakukan proses konversi ke tipe bilangan integer. Selanjutnya dalam rangka efisiensi, nilai tiap atribut pada dataset dilakukan normalisasi dengan nilai antara 0 sampai 1. Tampilan data nilai atribut yang telah dinormalisasi terdapat pada Gambar 1.

```

Normalisasi Data Rentang 0-1
[[0.5 0.81 0. ... 0. 0.33333333 1. ]
 [0.5 0.69 0.535 ... 0. 0.33333333 1. ]
 [0.5 0.7425 0.705 ... 0. 0. 0. ]
 ...
 [1. 0.555 0.645 ... 1. 1. 1. ]
 [1. 0.46 0.46 ... 1. 0.33333333 1. ]
 [1. 0.865 0.545 ... 1. 0.66666667 1. ]

```

Gambar 1. Data Hasil Normalisasi

Proses utama dalam penelitian ini adalah membangun model K-Medoids. Inisialisasi nilai parameter yang digunakan dalam membangun model K-Medoids diantaranya yaitu :

- a. jumlah kluster = 2 (berpotensi DO, tidak berpotensi DO)
- b. jumlah data = 389 (pada Tabel 5.1)
- c. jumlah atribut = 7 (X1 sampai X7)

Terdapat beberapa fungsi yang digunakan dalam membangun model K-Medoids, diantaranya yaitu :

- a. *random_state* yang digunakan untuk menentukan keadaan acak untuk generator angka acak dalam rangka menginisialisasi medoid ketika *init=random*
- b. *labels_* sebagai label dari setiap titik atau cluster
- c. *cluster_centers_* merupakan pusat cluster, yaitu medoid (elemen dari dataset asli)
- d. *inertia_* merupakan jumlah jarak sampel ke pusat cluster terdekat

Pada Gambar 2 menunjukkan 389 label kluster yang terbentuk dari proses yang dilakukan oleh metode K-Medoids. 389 label ini menjadi kolom baru

yang menjadi output untuk setiap *record* dataset mahasiswa prodi Rekayasa Sistem Komputer. Hanya terdapat 2 label kluster yaitu 0 dan 1 sesuai jumlah k yang diatur yaitu menggambarkan kluster berpotensi DO dan kluster tidak berpotensi DO.

[illegible]

Gambar 2. Hasil Kluster Label

Pada Gambar 3 menunjukkan titik pusat cluster, yaitu medoid (elemen dari dataset asli). Jadi nilai tersebut merupakan titik pusat yang secara acak dipilih sebagai medoid pada setiap iterasi yang terjadi. Titik pusat yang ditampilkan merupakan medoid terakhir. Selanjutnya terdapat hasil inerti sebesar 268.782 yang menunjukkan jumlah jarak sampel ke pusat cluster terdekat. Terakhir *silhouette coefficient* sebesar 0.39415227406014575 yang menunjukkan kedekatan kemiripan data yang dikelompokkan di dalam suatu kluster. Berdasarkan klasifikasi pada Tabel 6, hasil ini menunjukkan *Weak Structure*.

```
Pusat Kluster
[[0.5      0.78      0.7375    0.5      0.      0.33333333
 1.      ]
 [1.      0.3875    0.4025    0.5      1.      0.66666667
 1.      ]]
```

Inertia : 268.7820812464123

Silhouette coefficient = 0.39415227406014575

Gambar 3. Hasil Pusat Kluster, Inertia, dan Silhouette Coefficient

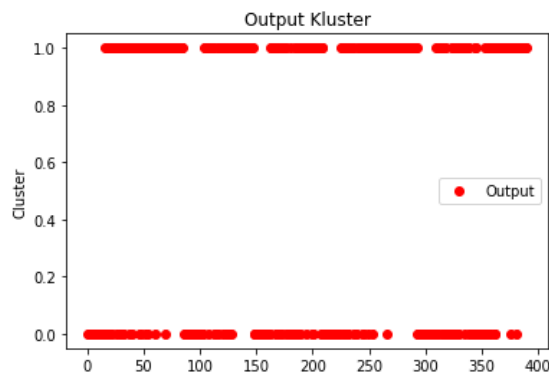
Adapun nilai *Silhouette Coefficient* yang dihasilkan sudah merupakan nilai tertinggi yaitu dengan 2 kluster jika dibandingkan jumlah cluster yang lebih banyak seperti yang terlihat pada Tabel 8.

Tabel 8. Perbandingan *Silhouette Coefficient*

No	Jumlah Cluster	<i>Silhouette Coefficient</i>
1	2	0.3941522740601458
2	3	0.2993778427665128
3	4	0.3035838986915357
4	5	0.2948000892567961

Langkah terakhir yang dilakukan yaitu membuat plot yang menggambarkan sebaran data tiap kluster yang terbentuk. Selain itu label kluster yang dihasilkan disimpan ke dalam sebuah file txt untuk selanjutnya diolah untuk dilakukan analisa lebih lanjut.

Pada Gambar 4 menunjukkan plot yang dihasilkan sebagai gambaran sebaran data pada setiap kluster yang terbentuk. Pada penelitian ini hanya terdapat 2 kluster yang menunjukkan mahasiswa yang berpotensi DO dan tidak berpotensi DO.



Gambar 4. Plot Output Kluster

Label kluster yang telah disimpan pada sebuah file txt selanjutnya digabungkan dengan dataset untuk menunjukkan setiap *record* data mahasiswa didapatkan hasil klusternya masing-masing.

Pada Gambar 5 menunjukkan rangkuman hasil klustering. Rekap ini menunjukkan terdapat 167 data yang masuk kategori kluster 0, dan 222 data kluster 1. Nilai tiap atribut pada setiap kluster juga dihitung nilai *min* maupun *max* nya untuk memberikan gambaran batasan nilai tiap atribut yang membedakan untuk setiap kluster.

Kluster	Jumlah Data	min/ max	X1	X2	X3	X4	X5	X6	X7
			Jalur Masuk	IPK	IP Sem 1	Tren IP 1-2	Beasiswa	Penghasilan Ortu	Bekerja
Kluster 0	167	min	0	0	0	0	0	0	0
		max	2	4	4	2	1	3	1
Kluster 1	222	min	0	0.00	0.00	0	1	0	0
		max	2	3.48	3.60	2	1	3	1
Total	389								

Gambar 5. Rekap Hasil Klustering

Hasil analisa kluster seluruh data menunjukkan terdapat beberapa perbedaan nilai pada setiap kluster yang terbentuk. Kluster 0 terdiri dari 167 mahasiswa yang memiliki IPK dengan rentang nilai 0 sampai 4, IP Semester 1 dengan rentang nilai 0 sampai 4, dan status beasiswa ada yang menerima beasiswa dan ada juga yang tidak menerima beasiswa. Sedangkan kluster 1 terdiri dari 222 mahasiswa yang memiliki IPK dengan rentang nilai 0 sampai 3.48, IP Semester 1 dengan rentang nilai 0 hingga 3.6, serta status beasiswa dimana semua berstatus tidak menerima beasiswa.

Hasil analisa tersebut menunjukkan ketiga atribut IPK, IP Semester 1, dan Status Beasiswa menjadi atribut yang membedakan nilai antar kluster yang ada. Dari perbandingan ciri kedua kluster tersebut, maka dapat diasumsikan kluster 0 merupakan kluster mahasiswa yang tidak berpotensi DO, sedangkan kluster 1 merupakan kluster mahasiswa yang berpotensi DO.

Analisa berikutnya terkait evaluasi dari proses klustering yang dilakukan dengan metode *Silhouette*. Dari hasil pengujian yang dilakukan didapatkan hasil 0.39415227406014575 dimana masuk kategori *Weak Structure*. Walaupun dianggap lemah, akan tetapi nilai ini masih positif, dimana jika nilai yang didapat

mendekati nilai 1, maka hasil yang didapat lebih baik dibandingkan nilai yang mendekati -1. Terlebih hasil yang didapat ini merupakan nilai tertinggi dibandingkan jika jumlah kluster nya lebih dari 2. Artinya jumlah sebanyak 2 kluster sudah merupakan jumlah yang paling ideal dalam kasus potensi DO ini berdasarkan data yang ada.

4. KESIMPULAN

Kesimpulan dari penelitian ini yaitu setelah dilakukan *clustering* menggunakan algoritma K-Medoids didapatkan hasil evaluasi *Silhouette Coefficient* terbaik sebesar 0.39415227406014575 dengan jumlah *cluster* 2. Hasil *cluster* dari 389 data, didapat 3 atribut yang memiliki rentang nilai yang berbeda antar *cluster* yaitu IPK, IP Semester 1, dan Status Beasiswa, ketiga atribut ini menjadi penciri yang membedakan antar *cluster*.

UCAPAN TERIMA KASIH

Penelitian ini didanai oleh DIPA FMIPA UNTAN tahun 2022 (SP DIPA-023.17.2677517/2022).

DAFTAR PUSTAKA

- ANGGERINI, N.L., 2019. Teknik Clustering Dengan Algoritma K-Medoids Untuk Menangani Strategi Promosi Di Politeknik TEDC Bandung. *Jurnal Teknologi Informasi dan Pendidikan*, 12(2), pp.1–7.
- DAMANIK, I.I.P., SOLIKHUN, SARAGIH, I.S., PARLINA, I., SUHENDRO, D. AND WANTO, A., 2019. Algoritma K-Medoids untuk Mengelompokkan Desa yang Memiliki Fasilitas Sekolah di Indonesia. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), p.520. <https://doi.org/10.30645/senaris.v1i0.58>.
- DEFIYANTI, S., JAJULI, M. AND ROHMAWATI, N., 2017. Optimalisasi K-MEDOID dalam Pengklasteran Mahasiswa Pelamar Beasiswa dengan CUBIC CLUSTERING CRITERION. *Jurnal Nasional Teknologi dan Sistem Informasi*, 3(1), pp.211–218. <https://doi.org/10.25077/teknosi.v3i1.2017.211-218>.
- DEKKER, G.W., PECHENIZKIY, M. AND VLEESHOUWERS, J.M., 2009. Predicting students drop out: A case study. *EDM'09 - Educational Data Mining 2009: 2nd International Conference on Educational Data Mining*, (January), pp.41–50.
- DWILESTARI, G., MULYAWAN, MARTANTO AND ALI, I., 2021. Analisis Clustering menggunakan K-Medoid pada Data Penduduk Miskin Indonesia. *JURSIMA: Jurnal Sistem Informasi dan Manajemen*, 9(3), pp.282–290.
- HAN, J. AND KAMBER, M., 2006. *Data Mining: Concepts and Techniques*. Soft Computing, San Francisco: Morgan Kaufmann. <https://doi.org/10.1007/978-3-642-19721-5>.
- HIDAYAT, M., PURWITASARI, D. AND GINARDI, H., 2013. Analisis Prediksi Drop Out Berdasarkan Perilaku Sosial Educational Data Mining Menggunakan Jaringan Syaraf Tiruan. *Jurnal IPTEK*, 17(2), pp.1–10.
- IRAWAN, E., SIREGAR, S.P., DAMANIK, I.S. AND SARAGIH, I.S., 2020. Implementasi Algoritma K-Medoids untuk Pengelompokkan Sebaran Mahasiswa Baru. *Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika)*, 5(2), p.275. <https://doi.org/10.30645/jurasik.v5i2.213>.
- MUHAMAD, WINDARTO, A.P. AND SUHADA, 2019. Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3(1), pp.1–8. <https://doi.org/10.30865/komik.v3i1.1688>.
- NURHAYATI, S., KUSRINI AND LUTHFI, E.T., 2015. Prediksi Mahasiswa Drop Out Menggunakan Metode Support Vector. *Jurnal Ilmiah SISFOTENIKA*, 3(6), pp.82–93.
- ROUSSEEUW, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), pp.53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- SETIYAWATI, A.W., 2017. *Implementasi Algoritma Partitioning Around Medoid (PAM) untuk Pengelompokan Sekolah Menengah Atas di DIY Berdasarkan Nilai Daya Serap Ujian Nasional*. [online] Universitas Sanata Dharma. Available at: <https://repository.usd.ac.id/9188/2/125314076_full.pdf>.
- UNTAN, T.F.M., 2021. *Buku Pedoman Akademik Fakultas MIPA Universitas Tanjungpura Tahun Ajaran 2021/2022*.
- VERCELLIS, C., 2009. *Business Intelligence: Data Mining and Optimization for Decision Making*. Business Intelligence: Data Mining and Optimization for Decision Making. John Wiley & Sons. <https://doi.org/10.1002/9780470753866>.

Halaman ini sengaja dikosongkan