

## PENERAPAN *K-MEANS++* UNTUK PENGELOMPOKAN MAHASISWA BERPOTENSI *DROP OUT* (STUDI KASUS: UNIVERSITAS IBN KHALDUN BOGOR)

Raka Putra Nugraha, Gibtha Fitri Laxmi, Freza Riana  
Teknik Informatika, Universitas Ibn Khaldun Bogor  
Jl. Sholeh Iskandar Kedungbadak, Kota Bogor, Indonesia  
Rakaptran78@gmail.com

### ABSTRAK

*Drop out* (DO) adalah fenomena penghentian atau pemutusan hubungan studi mahasiswa pada suatu perguruan tinggi yang disebabkan oleh beberapa faktor yang ditentukan oleh universitas tersebut. Fenomena ini merupakan masalah serius karena dapat berdampak pada kualitas perguruan tinggi. Penelitian ini bertujuan untuk mengidentifikasi dan mengelompokkan mahasiswa yang berpotensi *drop out* (DO) di Universitas Ibn Khaldun Bogor menggunakan algoritma *K-Means++*. Data yang digunakan merupakan data akademik dari 842 mahasiswa Teknik Informatika angkatan 2016 – 2022, dengan 84 diantaranya berstatus berpotensi *drop out*. Metode yang digunakan dalam penelitian ini adalah *K-Means++ clustering*, yang merupakan pengembangan dari algoritma *K-Means clustering*. *K-Means++ clustering* digunakan untuk mengatasi kekurangan *k-means* dalam hal efisiensi waktu dengan menentukan nilai awal *centroid* secara lebih cerdas, sehingga dapat mengurangi waktu pemrosesan. Hasil penelitian menunjukkan tiga *cluster* berpotensi DO: (T) tinggi, (S) sedang, dan (R) rendah, dengan mempertimbangkan faktor seperti IPK, jumlah SKS, aktivitas mahasiswa, penghasilan orang tua, dan jalur biaya. Hasil menunjukkan faktor berpengaruh ialah IPK, Jumlah Aktif, dan SKS, sedangkan faktor yang tidak berpengaruh ialah Jalur Biaya dan Penghasilan Orang Tua.

**Kata kunci :** *Clustering, Drop Out, K-Means++*, Perguruan Tinggi

### 1. PENDAHULUAN

Perguruan tinggi, sebagai penyelenggara pendidikan akademik, memiliki peran penting dalam membentuk lulusan yang cerdas dan kreatif, mendukung terlaksananya pembangunan nasional yang berkualitas[1]. Pada setiap perguruan tinggi pasti mempunyai database yang bertujuan untuk menyimpan data akademik, administrasi, dan data mahasiswa. Data tersebut apabila digali dengan benar maka dapat diketahui pola atau pengetahuan untuk mengambil keputusan. Salah satu data yang digali adalah pemahaman informasi mahasiswa yang potensial *drop out*[2]. *Drop out* (DO) merupakan penghentian atau pemutusan hubungan studi mahasiswa di perguruan tinggi karena beberapa faktor yang di tentukan universitas. Terdapat beberapa alasan dilakukan *drop out* terhadap mahasiswa[1]. Untuk menghindari terjadinya pemberhentian mahasiswa, maka perlu pendataan mahasiswa yang berpeluang *drop out* ataupun dilanjutkan studinya. Variabel yang mendukung pendataan mahasiswa berpotensi *drop out* ialah Induk Prestasi Semester (IPS), Induk Prestasi Kumulatif (IPK), status, penghasilan orang tua, total SKS, jumlah nonaktif, jumlah aktif, dan jalur biaya. Proses mencari pola atau informasi menarik dalam data yang telah terpilih dilakukan menggunakan teknik atau metode tertentu yang dikenal sebagai *data mining*[3]. *Data mining* adalah proses menemukan pola dalam data. Proses ini harus otomatis atau semi otomatis. Pola yang dihasilkan harus signifikan bahwa pola tersebut memberikan beberapa keunggulan. Polanya antara lain : deskripsi, estimasi, prediksi,

klasifikasi, pengklasteran, dan asosiasi. Salah satu teknik yang sering digunakan dalam *data mining* adalah teknik pengklasteran[4]. *Clustering* merupakan metode yang digunakan untuk mengelompokkan objek atau data ke dalam *cluster* (kelompok)[5]. Metode *clustering* yang biasa digunakan baik dalam dunia akademik, industri akademik, dan bisnis adalah algoritma *K-Means*. Algoritma *K-Means* adalah membagi data menjadi beberapa kluster yang nantinya akan di analisis tingkat kesamaan dan ketidaksamaan dari data yang terkumpul. Namun *K-Means* memiliki kelemahan dalam hal kecepatan dalam menentukan pusat *cluster*. Oleh sebab itu diperlukan metode khusus mampu mengatasi kelemahan dari metode *K-Means*. Kelemahan ini ternyata bisa disempurnakan menggunakan metode *K-Means++*[6]. Algoritma *K-Means++* menggunakan proses inialisasi nilai pusat *cluster*(*centroids*) awal, hal ini mengurangi ketidakstabilan yang terjadi pada *K-Means* sehingga dapat memberikan hasil klusterisasi yang baik dan stabil[7]. Evaluasi terhadap algoritma *K-Means++* melalui perhitungan akurasi menggunakan metode *silhouette coefficient*. Kualitas pengelompokan dapat dinilai menggunakan rata-rata siluet dimana memaksimalkan nilai indeks digunakan untuk mencari jumlah *cluster* yang ideal[8].

### 2. TINJAUAN PUSTAKA

Dalam bab kajian teoritis tentang teori-teori yang mendukung tema penelitian. Berikut adalah penjelasan tentang teori-teori yang mendukung penelitian.

## 2.1. Drop Out

Menurut Kamus Oxford, dropout adalah seseorang yang meninggalkan sekolah atau universitas sebelum menyelesaikan studinya. Sedangkan pengertian masa studi menurut peraturan Kementerian Riset, Teknologi, dan Pendidikan adalah masa perencanaan yang diperlukan seorang mahasiswa untuk menyelesaikan studinya. Batas waktu belajar adalah jumlah waktu maksimal yang dimiliki mahasiswa untuk menyelesaikan studinya. Mata kuliah sarjana terdiri dari minimal 144 atau 160 SKS yang dapat diselesaikan dalam waktu delapan semester atau kurang, atau sampai dengan 14 semester. Berikut ketentuan *drop out* untuk Pendidikan program sarjana (S1) [9]:

- Mahasiswa tidak memperoleh total SKS sebanyak 52 SKS pada akhir tahun kedua.
- Mahasiswa tidak boleh menyelesaikan total SKS sejumlah  $\geq 52$  SKS namun mencapai  $IPK < 2,00$  di akhir tahun kedua.
- Mahasiswa tidak boleh mengumpulkan jumlah total SKS yang dipersyaratkan (144 SKS) pada masa akhir studi maksimal (10 Semester).
- Jumlah total SKS yang dipersyaratkan (144 SKS) yang dikumpulkan oleh mahasiswa dibatas waktu maksimal masa studi, namun mendapatkan nilai E atau mendapatkan nilai D sejumlah  $> 10\%$  dari syarat minimal total SKS.
- Mahasiswa yang tidak membayar biaya UKT sesuai jadwal pada masing-masing semester dan yang bersangkutan tidak mengajukan *step-out* (SO).

## 2.2. Data Mining

Data mining adalah suatu proses untuk menemukan pengetahuan atau informasi yang berguna dari sejumlah besar data. Hal ini sering disebut dengan proses KDD (*Knowledge Discovery In Databases*). KDD atau *Knowledge Discovery In Databases from Data* merupakan proses terstruktur yaitu sebagai berikut [10]:

- Data Cleaning* adalah proses membersihkan data dari data *noise* dan *missing value*.
- Data Selection* adalah proses untuk memilih data dari *database* yang sesuai dengan tujuan analisis.

*Data transformation* adalah proses mengubah data bentuk data menjadi data yang sesuai untuk proses *mining*

## 2.3. Clustering

*Clustering* disebut sebagai pengelompokan. Metode ini mengidentifikasi kelompok dalam kasus berdasarkan kelompok dengan atribut yang sama. *Clustering* memisahkan sejumlah kelompok data berdasarkan karakteristiknya, dimana objeknya dapat berupa orang, peristiwa, dan lainnya yang didistribusikan ke dalam kelompok sehingga terdapat beberapa tingkatan yang saling berhubungan antar *cluster*. [11].

## 2.4. Metode K-Means++

*K-Means++ clustering* merupakan salah satu algoritma pengembangan dari *k-means clustering*, dimana salah satu metode *clustering* menggunakan pendekatan berbasis partisi. *K-means* tidak efisien waktu karena menentukan *centroid* secara acak, namun *K-Means++* digunakan untuk mengatasi kelemahan *K-Means* dari pemrosesan waktu yang lama sehingga efisien. Berikut langkah dalam menentukan nilai awal pada algoritma *K-Means++* :

- Menentukan salah satu titik pusat awal pada seluruh data dokumen secara acak.
- Tambahkan *center* baru  $c_i$  dari semua data yang belum terpilih sebagai titik pusat, menggunakan *weighted probability distribution* dimana data yang dipilih dengan probabilitas tertinggi dengan Persamaan 1. [12]:

$$K = \frac{D(x)^2}{\sum_{x \in X} D(x)^2} \quad (1)$$

Keterangan :

$D(x)^2 = \text{Euclidean distance}$

$\sum_{x \in X} D(x)^2 = \text{Jumlah jarak Euclidean distance}$

- Ulang Langkah ke 2, hingga sejumlah  $k$  *centroids* telah terpilih.
- Lalu di lanjutkan dengan algoritma *K-means*.

## 2.5. Silhouette coefficient

*Silhouette coefficient* merupakan metode yang dikembangkan oleh Rousseeuw pada tahun 1987 untuk mengukur kualitas sebuah pengelompokan, dengan fokus pada validasi dan interpretasi Kumpulan data. Selain itu, *silhouette coefficient* juga dapat digunakan untuk menunjukkan tingkat kepemilikan setiap objek di dalam suatu *cluster*. Rumus *silhouette coefficient* terdapat pada Persamaan 2.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

$s(i)$  = nilai *Silhouette Coefficient*.

$a(i)$  = rata-rata jarak suatu data  $i$ .

$b(i)$  = rata-rata jarak suatu data  $i$  dengan

semua objek .

Nilai rata-rata yang dimiliki oleh *Silhouette Coefficient* dari masing-masing data objek dalam suatu *cluster* menunjukkan seberapa layak data tersebut dimasukkan dalam *cluster*.

## 3. METODE PENELITIAN

### 3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari Unit Komputer dan Sistem Informasi (UKSI) Universitas Ibn Khaldun Bogor. Dataset penelitian ini melibatkan informasi dari 842 mahasiswa program Studi Teknik Informatika, mulai dari angkatan 2016 hingga 2022, dengan 84 mahasiswa memiliki status berpotensi *drop out*.

### 3.2. Tahapan Penelitian

Tahapan penelitian yang dilakukan dalam penelitian ini dibagi menjadi 2 bagian yaitu alur

penelitian yang dapat dilihat pada Gambar 1 dan alur *K-Means++* yang dapat dilihat pada Gambar 2.

**3.2.1. Alur Penelitian**



Gambar 1. Alur Penelitian

Berikut penjelasan dari penelitian ini :

**1. Pengumpulan Data**

Pada tahap ini, proses pengumpulan data dilakukan dengan mengambil data dari UKSI Universitas Ibn Khaldun Bogor yang terdiri dari data akademik mahasiswa S1 Program Studi Teknik Informatika dari Angkatan 2016-2022.

**2. Praproses Data**

Praproses data merupakan proses untuk mengelola data mentah sebelum dilakukan proses selanjutnya, yaitu diantaranya: *data selection*, *cleaning data* dan *transformation data*.

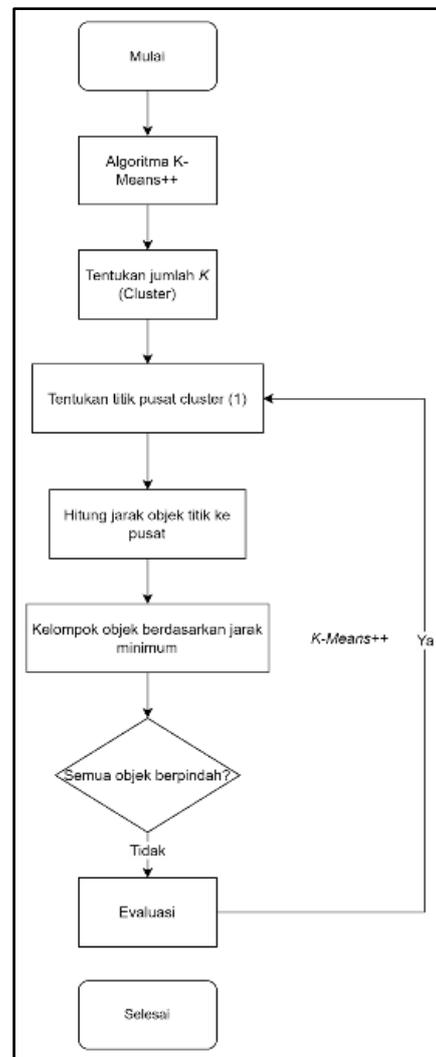
**3. Penerapan Metode *K-Means++***

Selanjutnya yaitu melakukan penerapan algoritma *K-Means++ clustering*. Adapun Alur (Gambar 2.) dan tahapan-tahapan dalam algoritma *K-Means++* yaitu sebagai berikut:

- a. Memilih nilai *K* untuk titik pusat dengan Persamaan 1. Setelah menentukan bayangan nilai *K* dan dihitung menggunakan Persamaan 1 akan mengeluarkan hasil yang akan digunakan untuk titik pusat.
- b. Kelompokkan berdasarkan jarak minimum terhadap titik pusat. Hasil dari nilai *K* yang sudah dihitung sebelumnya dikelompokkan berdasarkan jarak minimum terhadap titik pusat.
- c. Setelah mengelompokkan nilai titik pusat dengan jarak minimum, perbarui titik pusat dengan mencari rata-rata setiap *cluster*, Langkah ini mencari rata-rata dari setiap *cluster* untuk dijadikan nilai titik pusat.

- d. Mengulangi Langkah 2 dan 3 sampai semua objek tidak berpindah.

Menguji validasi sebuah *cluster* menggunakan *Silhouette Coefficient*.



Gambar 2. Alur *K-Means++*

**4. HASIL DAN PEMBAHASAN**

**4.1. Pengumpulan Data**

Data yang dikumpulkan merupakan data Mahasiswa Program Studi Teknik Informatika Universitas Ibn Khaldun Bogor yang didapat dari Unit Komputer dan Sistem Informasi dengan jumlah data 842 data Mahasiswa Universitas Ibn Khaldun Bogor. Data yang didapatkan adalah data Mahasiswa S1 Program Studi Teknik Informatika dari tahun 2016-2022. Data yang didapatkan merupakan data akademik mahasiswa yaitu, Nama Mahasiswa, IPS, IPK, Status, Penghasilan Orang Tua, Total SKS, Jumlah Nonaktif, Jumlah Aktif, dan Jalur Biaya. Rincian masing-masing atribut dapat dilihat pada Tabel 1.

Tabel 1. Data Awal Mahasiswa

| No | Nama                    | IPS | IPK  | Status        | Penghasilan Orang tua | SKS | Jumlah Nonaktif | Jumlah Aktif | Jalur   |
|----|-------------------------|-----|------|---------------|-----------------------|-----|-----------------|--------------|---------|
| 1  | JIMI HAFIADI            | 0   | 1,67 | BELUM MENIKAH | <Rp.500.000           | 57  | 9               | 3            | MANDIRI |
| 2  | AHMAD SYAJILI           | 0   | 1,33 | BELUM MENIKAH | <Rp.500.000           | 55  | 9               | 3            | MANDIRI |
| 3  | ZAINUL ULUM             | 0   | 1,53 | BELUM MENIKAH | <Rp.500.000           | 55  | 9               | 3            | MANDIRI |
| 4  | OYO SATRIO              | 0   | 1,35 | BELUM MENIKAH | <Rp.500.000           | 52  | 9               | 3            | MANDIRI |
| .. | ...                     | ..  | ..   | ..            | ..                    | ..  | ..              | ..           | MANDIRI |
| 84 | FATAHILLAH MUHAMMAD ALI | 0   | 0    | NULL          | NULL                  | 0   | 6               | 0            | MANDIRI |

#### 4.2. Praproses Data

Praproses data adalah proses mempersiapkan data sebelum melakukan proses selanjutnya yaitu *clustering*. Praproses data dalam penelitian ini terdiri dari *data selection*, *data cleaning*, dan *data transformation*.

##### 4.2.1. Data Selection

Menentukan variabel, variabel yang dipilih yaitu, Nama Mahasiswa, IPK, total SKS, jumlah aktif, jumlah non aktif, Penghasilan Orang tua, dan Jalur biaya agar bisa dilakukan tahap selanjutnya, variabel yang tidak terpakai maka akan dihapus. Variabel yang dihapus ialah variabel IPS, dan Status Mahasiswa. variabel yang dihapus dikarenakan banyaknya data kosong dan tidak update.

##### 4.2.2. Data Cleaning

Setelah dilakukan *data selection* maka selanjutnya dilakukan *data cleaning* atau penghapusan data yang tidak diperlukan selain variabel terdapat juga beberapa data yang tidak lengkap, dan seperti data kosong, Setelah melakukan *cleaning* data yang awalnya sebanyak 84 data Mahasiswa, setelah dilakukan *cleaning* terkumpul sejumlah 74 data Mahasiswa Program Studi Teknik Informatika.

##### 4.2.3. Data Transformation

Setelah dilakukan *data cleaning* maka selanjutnya dilakukan *data transformation* untuk mengubah data mentah menjadi data yang lebih tepat untuk memudahkan *clustering*.

Tabel 2. Data Mahasiswa Setelah Praproses Data

| NO | Nama         | IPK  | SKS | Jumlah Nonaktif | Jumlah Aktif | Penghasilan Orang Tua | Jalur |
|----|--------------|------|-----|-----------------|--------------|-----------------------|-------|
| 1  | Mahasiswa 1  | 1,67 | 57  | 9               | 3            | 1                     | 1     |
| 2  | Mahasiswa 2  | 1,33 | 55  | 9               | 3            | 3                     | 1     |
| 3  | Mahasiswa 3  | 1,53 | 55  | 9               | 3            | 4                     | 1     |
| 4  | Mahasiswa 4  | 1,35 | 52  | 9               | 3            | 1                     | 1     |
| 5  | Mahasiswa 5  | 1,76 | 72  | 8               | 4            | 1                     | 1     |
| .. | ...          | ...  | ... | ..              | ..           | ..                    | 1     |
| 74 | Mahasiswa 74 | 2,74 | 21  | 7               | 1            | 1                     | 1     |

#### 4.3. Implementasi Algoritma K-Means++

Implementasi *K-Means++* menggunakan sampel dari data yang diambil Angkatan 2016 – 2022 yaitu nilai Indeks Prestasi Kumulatif (IPK), jumlah SKS, Jumlah Nonaktif, Jumlah Aktif, Penghasilan

Orang Tua, dan Jalur Biaya. Pada penelitian ini data akan dikelompokkan menjadi 3 *cluster* sehingga butuh 3 *centroid*. Berikut *centroids* yang digunakan untuk melakukan *clustering* bisa dilihat pada Tabel 3.

Tabel 3. Centroids Terpilih K-Means++

| <i>centroid</i>                | IPK  | SKS | Jumlah Nonaktif | Jumlah Aktif | Penghasilan Orang Tua | Jalur |
|--------------------------------|------|-----|-----------------|--------------|-----------------------|-------|
| mahasiswa 12 <i>Centroid 1</i> | 1,22 | 89  | 7               | 5            | 1                     | 1     |
| mahasiswa 38 <i>Centroid 2</i> | 0    | 19  | 9               | 1            | 1                     | 1     |
| mahasiswa 24 <i>Centroid 3</i> | 3,42 | 136 | 0               | 8            | 1                     | 1     |

*Centroid* yang terpilih yaitu data Mahasiswa 12, Mahasiswa 38, dan Mahasiswa 24. Setelah memperoleh *centroid 1, 2, dan 3*. Setelah titik data (*centroid*) terpilih di metode *K-Means++*, maka dilanjutkan dengan Langkah algoritma *K-Means*, sebagai berikut:

#### 4.4. Algoritma K-Means

Adapun Langkah-langkah melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut :

Tahap pertama menentukan pusat cluster (*centroid*). Dalam penelitian ini dipilih *centroid* yang melalui algoritma *K-Means++* dan didapatkan *centroid*. Maka digunakanlah nilai titik data tersebut,

untuk nilai titik data (*centroid*) yang terpilih melalui algoritma K-Means++ dapat dilihat pada Tabel 3.

Jadi untuk nilai *centroid* terpilih yaitu:

- Cluster 1= 1.22, 89, 7, 5, 1, 1
- Cluster 2= 0, 19, 9, 1, 1, 1
- Cluster 3= 3.42, 136, 0, 8, 1, 1

Tahap kedua menghitung dengan Euclidian distance, setiap data dikelompokan berdasarkan

memilih angka terkecil, hasil dari *centroid* 1, 2, dan 3, lalu dimasukkan ke setiap *cluster* nya masing-masing. Pada contoh nilai dari ‘mahasiswa 1’ untuk nilai terkecilnya berada di kelas C1 maka untuk ‘mahasiswa 1’ ini berada di *cluster* 1. Untuk ‘mahasiswa 2’ karena nilai terkecilnya berada di kelas C1 maka untuk ‘mahasiswa 1’ ini termasuk ke dalam *cluster* 1, dan seterusnya sampai mahasiswa ke-74.

Tabel 4. Data Cluster pada iterasi ke-1

| Nama        | c1        | c2        | c3        | Jarak Terdekat | Cluster |
|-------------|-----------|-----------|-----------|----------------|---------|
| Mahasiswa 1 | 32,127908 | 38,207184 | 79,74373  | 32,12791       | 1       |
| Mahasiswa 2 | 34,176192 | 36,09389  | 81,684565 | 34,17619       | 1       |

Jumlah anggota *cluster* iterasi ke-1:

- Cluster 1= 23 Mahasiswa
- Cluster 2= 42 Mahasiswa
- Cluster 3 = 9 Mahasiswa

Tahap selanjutnya diulangi seperti diatas hingga nilai *centroid* dan keanggotaan masing-masing setiap *cluster* tidak berubah. Pada penelitian ini berhenti di iterasi ke-6, hasil sebagai berikut:

Tabel 5. Data Keanggotaan Pada Iterasi Ke-6

| Cluster 1    | Cluster 2    | Cluster 3    |
|--------------|--------------|--------------|
| 22 Mahasiswa | 39 Mahasiswa | 13 Mahasiswa |

Hasil dari iterasi ke-6:

- Cluster 1 = 22 Mahasiswa
- Cluster 2 = 39 Mahasiswa
- Cluster 3 = 13 Mahasiswa

Titik *centroid* setiap variabel tidak berubah maka perhitungan sudah selesai di iterasi ke-6. Hasil perhitungan dengan algoritma *K-Means* mahasiswa yang termasuk ke dalam *cluster* 1, 2, Dan 3 dapat dilihat pada Tabel 6.

Tabel 6. Anggota Cluster 1, 2, Dan 3

| Cluster 1   | Cluster 2    | Cluster 3    |
|-------------|--------------|--------------|
| mahasiswa 1 | mahasiswa 14 | mahasiswa 24 |
| mahasiswa 2 | mahasiswa 15 | mahasiswa 25 |
| mahasiswa 3 | mahasiswa 16 | mahasiswa 26 |
| mahasiswa 4 | mahasiswa 18 | mahasiswa 27 |
| mahasiswa 5 | mahasiswa 19 | mahasiswa 28 |
| mahasiswa 6 | mahasiswa 20 | mahasiswa 29 |
| mahasiswa 7 | mahasiswa 38 | mahasiswa 30 |
| mahasiswa 8 | mahasiswa 39 | mahasiswa 31 |

Tabel 7. Hasil Kelas Clustering

| Kelas  | Hasil Analisis  |
|--|---|
| Cluster 1 (S) termasuk kategori Berpotensi <i>drop out</i> sedang. | Memiliki nilai variabel IPK rendah (0 – 2,75), jumlah Total SKS dijalani sedang (50 - 100), jumlah Nonaktif lebih besar daripada jumlah Aktif tiap semester, mempunyai penghasilan Orang Tua rata-rata <Rp.500.000., dan jalur biaya Mandiri. |
| Cluster 2 (T) termasuk kategori Berpotensi                         | Memiliki nilai variabel IPK rendah (0, – 2,20), jumlah Total SKS dijalani rendah (0 – 50), jumlah Nonaktif lebih besar daripada jumlah Aktif tiap   |

| Cluster 1    | Cluster 2    | Cluster 3    |
|--------------|--------------|--------------|
| mahasiswa 9  | mahasiswa 40 | mahasiswa 32 |
| mahasiswa 10 | mahasiswa 41 | mahasiswa 33 |
| mahasiswa 11 | mahasiswa 42 | mahasiswa 34 |
| mahasiswa 12 | mahasiswa 43 | mahasiswa 35 |
| mahasiswa 13 | mahasiswa 44 | mahasiswa 36 |
| mahasiswa 17 | mahasiswa 45 |              |
| mahasiswa 21 | mahasiswa 46 |              |
| mahasiswa 22 | mahasiswa 47 |              |
| mahasiswa 23 | mahasiswa 48 |              |
| mahasiswa 37 | mahasiswa 49 |              |
| mahasiswa 50 | mahasiswa 51 |              |
| mahasiswa 54 | mahasiswa 52 |              |
| mahasiswa 55 | mahasiswa 53 |              |
| mahasiswa 56 | mahasiswa 57 |              |
|              | mahasiswa 59 |              |
|              | mahasiswa 60 |              |
|              | mahasiswa 61 |              |
|              | mahasiswa 62 |              |
|              | mahasiswa 63 |              |
|              | mahasiswa 64 |              |
|              | mahasiswa 65 |              |
|              | mahasiswa 66 |              |
|              | mahasiswa 67 |              |
|              | mahasiswa 68 |              |
|              | mahasiswa 69 |              |
|              | mahasiswa 70 |              |
|              | mahasiswa 71 |              |
|              | mahasiswa 72 |              |
|              | mahasiswa 73 |              |
|              | mahasiswa 74 |              |

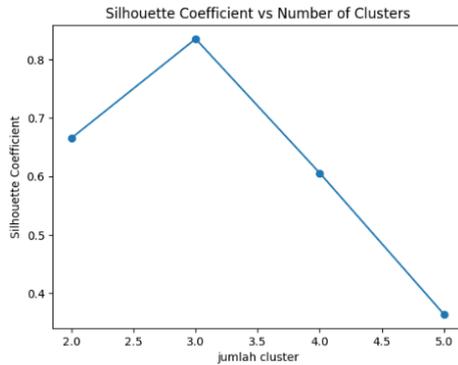
Setelah dianalisis untuk kasus mahasiswa *drop out* dari penelitian ini dibagi menjadi 3 kelas yaitu (T) Tinggi, (S) Sedang, dan (R) Rendah dari setiap mahasiswa, Bisa dilihat pada Tabel 7.

| Kelas  | Hasil Analisis  |
|--|---|
| <i>drop Out</i> tinggi.  | semesternya, mempunyai penghasilan Orang Tua rata-rata <Rp.500.000., dan jalur biaya Mandiri.   |
| Cluster 3 (R) termasuk kategori Berpotensi <i>drop Out</i> rendah. | Memiliki nilai variabel IPK tinggi (2,75 – 4,00) , jumlah Total SKS dijalani tinggi (100 >), jumlah Aktif lebih besardaripada jumlah Nonaktif tiap semester, mempunyai penghasilan Orang Tua rata-rata <Rp.500.000., dan jalur biaya Mandiri. |

**4.5. Silhouette Coefficient**

*Silhouette Coefficient* digunakan untuk mengukur validitas *clustering* dan mengindikasikan derajat kepemilikan setiap objek yang dalam *cluster*.

Pada Gambar 3 menunjukkan kualitas setiap *cluster*, sehingga dapat mengetahui *cluster* manakah yang memiliki hasil perhitungan yang kuat. Bisa dilihat pada tabel berikut adalah perhitungan indeks *Silhouette Coefficient*.



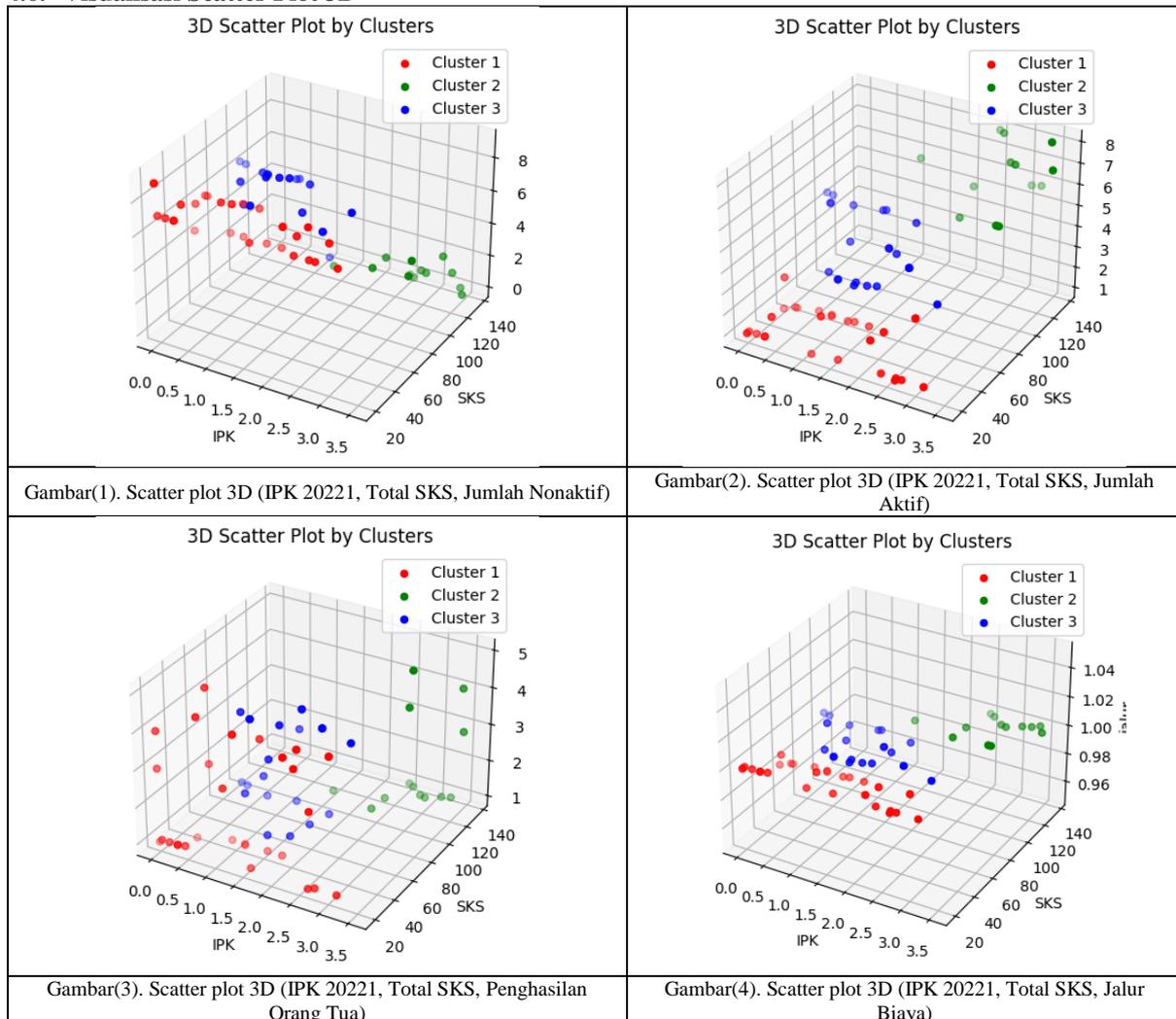
Gambar 3. *Silhouette Coefficient*

Tabel 8. Hasil Indeks *Silhouette Coefficient*

| Percobaan Cluster | Hasil Indeks <i>Silhouette Coefficient</i> |
|-------------------|--|
| 2                 | 0.6654                                     |
| 3                 | 0.8351                                     |
| 4                 | 0.6055                                     |
| 5                 | 0.3639                                     |

Pada Tabel 8 *Indeks Silhouette Coefficient* untuk percobaan *cluster* sebanyak 2 *cluster* memiliki indeks sebesar 0.6654. Begitu juga percobaan dengan 3 *cluster* yang memiliki rata-rata nilai *Indeks Silhouette Coefficient* sebesar 0.8351, untuk percobaan dengan 4 *cluster* memiliki rata-rata nilai *Indeks Silhouette Coefficient* sebesar 0.6055, untuk percobaan dengan 5 *cluster* memiliki rata-rata nilai *Indeks Silhouette Coefficient* sebesar 0.3639. Sehingga percobaan dengan 3 *cluster* yang paling terbesar berdasarkan nilai Indeksnya sebesar 0.8351

**4.6. Visualisasi Scatter Plot 3D**



Gambar(1). Scatter plot 3D (IPK 2021, Total SKS, Jumlah Nonaktif)

Gambar(2). Scatter plot 3D (IPK 2021, Total SKS, Jumlah Aktif)

Gambar(3). Scatter plot 3D (IPK 2021, Total SKS, Penghasilan Orang Tua)

Gambar(4). Scatter plot 3D (IPK 2021, Total SKS, Jalur Biaya)

Gambar 4. *Scatter Plot*

Pada Gambar 4 memiliki hasil 3 cluster, dimana cluster 1 berwarna merah, cluster 2 berwarna hijau, cluster 3 berwarna biru. Pada cluster 1 menunjukkan mahasiswa tersebut mahasiswa pada cluster 1 mempunyai IPK (0 – 2,20), menjalani total SKS (0 – 50), jumlah Nonaktif lebih besar daripada jumlah aktifnya, mempunyai penghasilan orang tua rata – rata <Rp.500.000, dan melalui jalur biaya Mandiri. Cluster 2 menunjukkan mahasiswa tersebut mahasiswa pada cluster 2 mempunyai IPK (2,75 – 4,00), menjalani total SKS (100>), jumlah aktif lebih besar daripada jumlah nonaktifnya, mempunyai penghasilan orang tua rata – rata <Rp.500.000, dan melalui jalur biaya Mandiri. Cluster 3 menunjukkan mahasiswa tersebut mahasiswa pada cluster 3 mempunyai IPK (0 – 2,75), menjalani total SKS (50 – 100), jumlah Nonaktif lebih besar daripada jumlah aktifnya, mempunyai penghasilan orang tua rata – rata <Rp.500.000, dan melalui jalur biaya Mandiri. Berdasarkan visualisasi Gambar 4.14 didapatkan bahwa penghasilan orang tua berada pada setiap cluster dengan kategori <Rp.500.000 yang dimana data tersebut tidak menjadi pembeda setiap cluster.

## 5. KESIMPULAN DAN SARAN

Setelah dilakukan *clustering* menggunakan algoritma *K-Means++* didapatkan hasil *Silhouette Coefficient* terbaik sebesar 0.8351 dengan jumlah 3 *cluster*. Hasil *cluster* dari 74 data didapatkan variabel yang memiliki rentang nilai yang berbeda antar *cluster* yaitu IPK, SKS, Jumlah Aktif, Penghasilan Orang Tua, dan Jalur Biaya. Kelima atribut ini menjadi penciri yang membedakan antar *cluster*. Dari tiap variabel Penghasilan Orang Tua tidak menjadikan ciri pembeda untuk setiap *cluster*, karena hampir setiap *cluster* memiliki penghasilan <Rp.500.000 yang mendominasi, sedangkan variabel lain menjadi salah satu penciri yang membedakan setiap *cluster*.

Peneliti mengusulkan untuk mengembangkan penelitian lebih lanjut mengenai “Penerapan *K-Means++* untuk pengelompokan Mahasiswa Berpotensi *drop out* (Studi Kasus: Universitas Ibn Khaldun Bogor)” sebagai berikut: Mengembangkan penelitian serupa dengan menggunakan teknik *clustering* lain dengan algoritma seperti *Hierarchical clustering*, *Fuzzy C-Means*, *K-Modes*, *K-Medoids*, dan lain-lain. Agar memperoleh hasil yang lebih beragam. Melakukan penelitian serupa dengan menambahkan variabel tambahan seperti tunggakan biaya, status mahasiswa, jumlah kehadiran, dan variabel lainnya yang berkaitan dengan mahasiswa potensial *drop out*. Saran dari hasil penelitian untuk Program Studi Teknik Informatika Universitas Ibn Khaldun Bogor sebaiknya dapat mengambil kebijakan terhadap mahasiswa yang berpotensi *drop out*. Misalnya menilai pembelajaran, memberikan motivasi di sebuah forum, dan memberikan solusi jika masih ada peluang untuk melanjutkan.

## DAFTAR PUSTAKA

- [1] S. Bahri, D. Marisa Midyanti, and P. Korespondensi, “Penerapan Metode K-Medoids Untuk Pengelompokan Mahasiswa Berpotensi Drop Out Application of K-Medoids Method for Dropout Potential Student Grouping,” vol. 10, no. 1, pp. 165–172, 2023, doi: 10.25126/jtiik.2023106643.
- [2] adi Firmansyah, acep irham Gufroni, and andy nur Rachman, “Data Mining Dengan Metode Clustering K-Mean Untuk Pengelompokan Mahasiswa Potensial Drop Out Pada Program Studi Teknik Informatika Universitas Siliwangi,” *Tek. Inform. ...*, pp. 1–6, 2017.
- [3] Yuli Mardi, “Data Mining: Klasifikasi Menggunakan Algoritma C4.5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ). Jurnal Edik Informatika,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- [4] Suparyanto dan Rosad, “Penerapan Algoritma C4.5 Pada Program Klasifikasi Mahasiswa Dropout,” *Suparyanto dan Rosad 2020*, vol. 5, no. 3, pp. 248–253, 2020.
- [5] I. Vhallah, S. Sumijan, and J. Santony, “Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Metode Clustering K-Means,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 2, pp. 572–577, 2018, doi: 10.29207/resti.v2i2.308.
- [6] U. F. Laili, C. Umatin, and M. U. Ridwanulloh, “ANALISIS POTENSIAL DROP OUT MAHASISWA DENGAN K-MEANS ++,” vol. 6356, pp. 145–153, 2023.
- [7] C. Ramdani and N. Safadila, “Analisis Data Akademis dengan Menerapkan Algoritme K-Means dan K-,” vol. 8798, pp. 155–160, 2022.
- [8] F. Nuraeni, D. Kurniadi, and G. Fauzian Dermawan, “Pemetaan Karakteristik Mahasiswa Penerima Kartu Indonesia Pintar Kuliah (KIP-K) menggunakan Algoritma K-Means++,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 437–443, 2023, doi: 10.32736/sisfokom.v11i3.1439.
- [9] Kemeristekdikti, “Permenristekdikti Nomor 44 Tahun 2015 tentang Standar Nasional Pendidikan Tinggi,” 2015. <https://lldikti13.kemdikbud.go.id/2016/01/26/standar-nasional-pendidikan-tinggi-tahun-2015/>
- [10] D. Sinaga, E. J. Solaiman, and F. J. Kaunang, “Penerapan Algoritma Decision Tree C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop out Di Universitas Advent Indonesia,” *TeIKa*, vol. 11, no. 2, pp. 167–173, 2021, doi: 10.36342/teika.v11i2.2613.
- [11] J. Nasir, “Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 690–703, 2021, doi: 10.24176/simet.v11i2.5482.

- [12] C. R. Sistiani, Widodo, and B. P. Padhi, “Kinerja Algoritma Kmeans++ pada Pengelompokan Dokumen Teks Pendek pada Abstrak di Jurusan Teknik Elektro Fakultas Teknik UNJ,” *PINTER J. Pendidik. Tek. Inform. dan Komput.*, vol. 2, no. 1, pp. 39–44, 2018, doi: 10.21009/pinter.2.1.6.