IMPLEMENTASI ALGORITMA *K-MEANS CLUSTERING* UNTUK PENGELOMPOKAN STATUS PENERIMA KIP KULIAH MAHASISWA UNIVERSITAS PAPUA

Eka Indriati, Nur'Ainun Suharyani Azisa, Ester Ivo Sihombing, Zharima Sukma Dewi Mokodompit
Teknik Informatika, Universitas Papua
Jln. Gunung Salju, Amban, Manokwari 98314
ekaindriati09@gmail.com

ABSTRAK

Penelitian ini berisi mengenai pengelompokan status penerima KIP Kuliah di Universitas Papua. KIP Kuliah merupakan program bantuan pendidikan yang diberikan oleh pemerintah Indonesia untuk meningkatkan akses pendidikan. Pada Universitas Papua proses seleksi calon penerima KIP Kuliah dilakukan secara manual dengan jumlah pendaftar yang sangat banyak sehingga penentuan KIP Kuliah kurang optimal. Penelitian ini bertujuan untuk menentukan calon penerima beasiswa KIP Kuliah di Universitas Papua menggunakan Algoritma *K-Means Clustering* dan bahasa pemrograman *python* pada *tools Jupyter Notebook*. Algoritma *K-Means Clustering* digunakan untuk mengelompokan setiap data kedalam *cluster* sehingga data yang mempunyai karakteristik yang sama akan dikelompokan dalam *cluster* yang sama dan sebaliknya jika data yang mempunyai karakteristik berbeda akan dikelompokkan kedalam *cluster* yang lain. Hasil penelitian ini terdapat 2 hasil *cluster* yang akan menjadi hasil akhir yaitu data yang diterima sebagai penerima KIP Kuliah sebanyak 687 data dan 547 data dikelompokkan sebagai data yang tidak diterima sebagai penerima KIP Kuliah.

Kata kunci: *Data mining, K-Means Clustering, KIP Kuliah*

1. PENDAHULUAN

Pemerintah Indonesia menyadari bahwa memiliki sumber daya manusia yang berkualitas sangat penting untuk kemajuan dan kesuksesan dalam berbagai aspek. Untuk mewujudkan hal tersebut, Pemerintah Indonesia berupaya meningkatkan akses pendidikan di semua tingkatan, termasuk di pendidikan tinggi tingkat universitas. Berdasarkan Undang-Undang Nomor 12 tahun 2012 tentang Pendidikan Tinggi, Pemerintah Indonesia berkewajiban meningkatkan akses dan kesempatan belajar di Perguruan Tinggi serta menyiapkan insan Indonesia yang cerdas dan kompetitif [1]. Maka dari itu pemerintah indonesia memiliki program yaitu PIP yang merupakan bantuan berupa uang tunai, perluasan akses, dan kesempatan belajar. Pemerintah memberikan bantuan kepada peserta didik dan mahasiswa yang berasal dari keluarga miskin/rentan miskin serta memiliki pencapaian akademik yang unggul untuk membiayai pendidikan[1][2]. PIP Pendidikan Tinggi untuk mahasiswa diberikan dalam bentuk Kartu Indonesia Pintar Kuliah atau KIP Kuliah yang diberikan pada kampus negri maupun swasta.

Universitas Papua merupakan salah satu universitas negeri yang dipilih oleh pemerintah untuk memberikan bantuan berupa KIP Kuliah kepada mahasiswanya. Setiap tahun, jumlah kuota penerima KIP Kuliah yang diberikan oleh pemerintah kepada Universitas Papua bervariasi, namun tidak dipungkiri bahwa kuota tersebut tidak selalu cukup untuk menampung seluruh jumlah mahasiswa yang mendaftar KIP Kuliah, terutama mengingat jumlah pendaftar yang terus meningkat setiap tahun. Hal ini menyebabkan para pengambil keputusan (pihak universitas) kesulitan dalam menentukan penerima KIP Kuliah secara optimal dan efesien. Pengambilan keputusan penerima K Kuliah di Universitas Papua juga masih dilakukan seacara manual sehingga membutuhkan waktu yang cukup lama. Pada tahun 2023 jumlah pendaftar KIP Kuliah di Universitas papua berkisar 1234 mahasiswa. Dengan jumah data mahasiswa pendaftar yang banyak tersebut pihak universitas harus memantau dengan cermat ketika melakukan pengambilan keputusan penerimaan KIP Kuliah pada setiap calon mahasiswa untuk memastikan pemberian program KIP Kuliah dilakukan secara tepat dan sesuai.

Untuk membantu pengambilan keputusan dalam proses seleksi calon penerima KIP Kuliah secara tepat dengan jumlah data yang cukup banyak, serta dapat memberikan efisiensi waktu maka dibutuhkan metode penerapan *Data mining* agar proses seleksi calon penerima KIP Kuliah Universitas Papua dapat dilakukan dengan cepat dan tepat sasaran[3]. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan mesin learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait berbagai *database* besar[3]. Pada *Data mining* terdapat beberapa metode, salah satunya yaitu metode algoritma *K-Means cluster*ing. Metode algoritma *K-Means cluster*ing akan mengelompokkan data yang memiliki karakteristik atau pola yang sama kedalam sebuah kelompok yang sama, dan data yang mempunyai karateristik atau pola yang berbeda kedalam kelompok lainya. Melalui *cluster*ing ini akan diketahui informasi dari sekumpulan data yang besar[3]. Kelebihan metode *K-Means cluster*ing

adalah sangat mudah diimplementasikan dan paling banyak digunakan dalam *cluster*ing. Algoritma K-Means relatif lebih efisien dalam mengolah data dengan jumlah yang besar karena memiliki tingkat ketelitian yang bagus. Disisi lain, Algoritma K-Means juga mempunyai kekurangan yaitu dalam menganalisa dan menentukan jumlah *cluster* dalam pengelompokkan sebuah data.

Dalam penggunaan algoritma *K-Means Cluster*ing sebagai metode pengambilan keputusan digunakan dua atribut yaitu Penghasilan orang tua dan jumlah tanggungan. Kedua atribut akan dikelompokkan berdasarkan kesamaan karakteristik data dan akan dibagi menjadi beberapa *cluster*. Pengelompokkan ini dilakukan dengan menggunakan bahasa pemrograman *python* dan menggunakan metode elbow untuk penentuan jumlah *cluster*. Hasil *cluster* ini nantinya yang akan menjadi penentuan dalam pengambilan keputusan penerima KIP Kuliah di Universitas Papua.

Penelitian ini bertujuan untuk mempermudah pengelompokkan data mahasiswa penerima KIP Kuliah di Universitas Papua dengan status diterima dan tidak diterima serta untuk mengevaluasi keakuratan metode *K-Means Cluster*ing terhadap pengambilan keputusan penerima KIP Kuliah.

2. TINJAUAN PUSTAKA.

2.1. *Data mining*

Data mining adalah proses mencari informasi yang tersembunyi dari dalam *database* yang besar serta memiliki manfaat bagi pihak yang berkepentingan. Terdapat beberapa metode dalam data mining yaitu asosiasi, regresi, *clustering*, klasifikasi dan lain-lain. Pada penelitian ini metode yang dipakai yaitu *clustering* menggunakan algoritma *K-Means* pada pemrograman pyhton. Didalam *data mining* terdapat langkah-langkah proses yang dimulai dengan menyeleksi data melalui data sumber ke data target, selanjutnya masuk pada tahap *processing* yaitu untuk memperbaiki suatu kualitas data, perubahan data, *data mining* dan tahapan interpretasi serta evaluasi yang menghasilkan suatu *output* seperti pengetahuan yang baru[4][3].

2.2. *Cluster*ing

*Cluster*ing adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripin yang minimum[5]. *Cluster* adalah sekelompok atau sekumpulan objek-objek data yang similar satu sama lain dalam *cluster* yang sama dan dissimilar terhadap objek-objek yang berbeda *cluster*. Objek-objek dikelompokkan dengan prinsip untuk memperbesar kesamaan antara objek-objek dalam satu kelompok (*cluster*) dan meningkatkan perbedaan di antara kelompok-kelompok yang berbeda. *Cluster*ing berujuan untuk mengidentifikasi pola alami atau struktur tersembunyi dalam data tanpa menggunakan kategori atau label sebelumnya. Proses pengelompokan melibatkan penggunaan algoritma yang menganalisis kesamaan atau perbedaan antara titik-titik data. [4][6].

2.3. Algoritma *K-Means*

Algoritma *K-Means* merupakan salah satu teknik untuk mengelompokkan data dalam kelompok atau sistem partisi yang disesuaikan dengan karakteristik masing-masing bagian data. Istilah pengelompokan dalam *K-Means* digunakan untuk menggambarkan suatu algoritma yang menetapkan setiap elemen terhadap *cluster* dengan centroid terdekat (*mean*) dimulai dengan memisahkan objek menjadi k *cluster* awal, lalu ditetapkan objek ke seluruh *cluster* centroid terdekat biasanya dihitung berdasarkan jarak Euclidean[7][8].

2.4. Metode Elbow

Metode dan analisis ini digunakan untuk pemilihan jumlah *cluster* atau kelompok yang optimal. Metode Elbow merupakan sebuah metode yang digunakan untuk memperoleh informasi jumlah *cluster* paling baik, dengan memperhatikan hasil perbandingan antara *cluster*-*cluste*r yang akan membentuk siku pada titik tertentu. Pada metode Elbow nilai *cluster* terbaik akan diambil dari nilai *Sum of Square Error* (SSE) yang mengalami penurunan yang signifikan dan berbentuk siku, semakin besar selisih penurunan SSE antar k dan berberntuk siku berdasarkan grafik maka hasil *cluster* tersebut yang paling optimal[9][10].

3. METODE PENELITIAN

Berikut tahapan penelitian yang dilakukan pada *cluster*ing data KIP Kuliah Universitas Papua menggunakan Algoritma *K-Means Cluster*ing :

3.1. Identifikasi Masalah

Pada tahap ini, dilakukan identifikasi terhadap berbagai permasalahan yang muncul dalam konteks penelitian. Permasalahan utama yang muncul adalah banyaknya jumlah data pendaftar KIP Kuliah dan pengerjaan pengambilan keputusan yang masih manual, yang berdampak pada ketidak tepatan dalam penyaluran KIP Kuliah karena kurangnya akurasi, serta menghabiskan waktu pengerjaan yang cukup lama. Untuk itu dibutuhkannya metode Algoritma *K-Means Cluster*ing sebagai penyelesaian dari masalah tersebut.

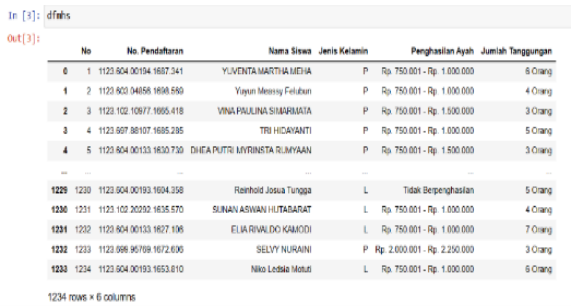
3.2. Studi Literatur

Dalam tahap Studi Literatur, dilakukan penelusuran dan pemahaman mendalam terhadap landasan teori yang diperoleh dari berbagai sumber, termasuk buku, jurnal ilmiah, serta referensi lainnya yang relevan terkait dengan permasalahan yang telah diformulasikan. Teori-teori ini menjadi acuan utama untuk menggali solusi yang dibutuhkan dalam penyelesaian masalah yang ada.

3.3. Pengumpulan Data

Metode pengumpulan data merupakan teknik atau metode yang digunakan untuk memperoleh data agar mendapatkan informasi yang dibutuhkan untuk mencapai tujuan penelitian [7]. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari *Website* KIP Kuliah. Data yang digunakan berupa data Mahasiswa Pendaftar KIP Kuliah di Universitas Papua tahun 2023, yang terdiri dari 1.234 pendaftar. Data yang telah diperoleh kemudian diolah terlebih dahulu untuk dapat digunakan[11].

Selanjutnya dilakukan *import* data ke *Jupyter Notebook* sebagai langkah awal dalam menganalisis data menggunakan bahasa pemrograman *python*.



In [3]: dfmhs

Out[3]:

	No	No. Pendaftaran	Nama Siswa	Jenis Kelamin	Penghasilan Ayah	Jumlah Tanggungan
0	1	1123.604.00194.1687.341	YUVENTA MARTHA MEHA	P	Rp. 750.001 - Rp. 1.000.000	6 Orang
1	2	1123.603.04658.1698.560	Yuyun Meassy Felubun	P	Rp. 750.001 - Rp. 1.000.000	4 Orang
2	3	1123.102.10977.1695.418	VINA PAULINA SIMARMATA	P	Rp. 750.001 - Rp. 1.500.000	3 Orang
3	4	1123.607.98107.1685.285	TRI HIDAYANTI	P	Rp. 750.001 - Rp. 1.000.000	5 Orang
4	5	1123.604.00133.1630.730	DHEA PUTRI MYRINSTA RUMYAAN	P	Rp. 750.001 - Rp. 1.500.000	3 Orang
...	...	...	...	...	...	...
1229	1230	1123.604.00193.1604.358	Reinhold Josua Tungga	L	Tidak Berpenghasilan	5 Orang
1230	1231	1123.102.20292.1635.570	SUNAN ASWAN HUTABARAT	L	Rp. 750.001 - Rp. 1.000.000	4 Orang
1231	1232	1123.604.00133.1627.106	ELIA RIVALDO KAMODI	L	Rp. 750.001 - Rp. 1.000.000	7 Orang
1232	1233	1123.699.95769.1672.606	SELVY NURAINI	P	Rp. 2.000.001 - Rp. 2.250.000	3 Orang
1233	1234	1123.604.00193.1653.810	Niko Ledsia Motuti	L	Rp. 750.001 - Rp. 1.000.000	6 Orang

1234 rows × 6 columns

Gambar 1. Data mahasiswa pendaftar KIP kuliah

Gambar 1 menunjukan daftar data calon penerima KIP Kuliah yang sudah di import kedalam *Jupyter Notebook* dengan jumlah data sebanyak 1234 pendaftar menggunakan pustaka *Pandas* yang sering digunakan dalam menganalisis data.

3.4. Pra Proses Data

Tahap pra proses *data mining* meliputi data *cleaning* atau pembersihan data. Dalam tahap ini apabila terdapat atribut kosong maupun tidak lengkap dan memerlukan data inisialisasi atau pengubahan data menjadi format yang sesuai untuk digunakan dalam proses *data mining* maka akan dilakukan pra-proses data. Adapun bentuk data inisialisasi yang telah di lakukan sebagai berikut :

Tabel 1. Inisialisasi data penghasilan

Penghasilan	Bobot	Keterangan
< Rp. 1.500.000	1	Sangat Rendah
Rp. 1.500.000 – Rp. 2.500.000	2	Rendah
Rp. 2.500.000 – Rp. 3.500.000	3	Sedang
Rp. 3.500.000 – Rp. 4.500.000	4	Tinggi
> Rp. 4.500.000	5	Sangat Tinggi

Tabel 2. Inisialisasi data jumlah tanggungan

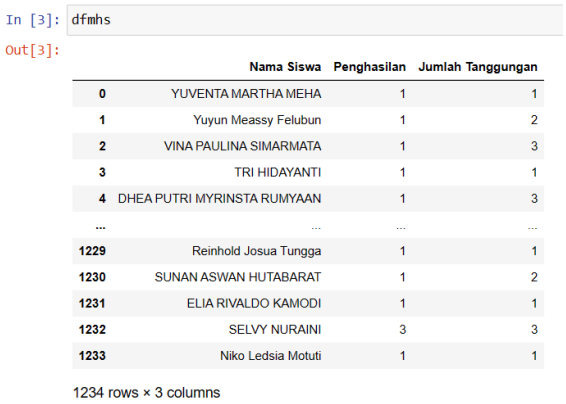
Jumlah Tanggungan	Bobot	Keterangan
> 5 Orang	1	Sangat Banyak
4 Orang	2	Banyak
3 Orang	3	Sedang
2 Orang	4	Sedikit
1 Orang	5	Sangat Sedikit

Tabel 1 dan 2 merupakan tabel hasil inisialisasi pada atribut penghasilan dan jumlah tanggungan. Inisialisasi dilakukan untuk mempermudah penerapan Algoritma *K-Means*.

4. HASIL DAN PEMBAHASAN

4.1. Data Set

Berdasarkan data yang diperoleh oleh penulis terdapat 1.234 data mahasiswa yang di telah mendaftar sebagai penerima KIP Kuliah di Universitas Papua. Data yang diperoleh tersebut akan digunakan pada penerapan Algoritma *K-Means Cluster*ing. Data yang digunakan sebelumnya telah melalui tahap pra proses data, dimana telah dilakukan pembersihan data pada atribut yang tidak digunakan serta menormalisasikan data agar siap digunakan. Data tersebut dapat dilihat pada gambar dibawah ini:



In [3]: dfmhs

Out[3]:

	Nama Siswa	Penghasilan	Jumlah Tanggungan
0	YUVENTA MARTHA MEHA	1	1
1	Yuyun Meassy Felubun	1	2
2	VINA PAULINA SIMARMATA	1	3
3	TRI HIDAYANTI	1	1
4	DHEA PUTRI MYRINSTA RUMYAAN	1	3
...	...	...	...
1229	Reinhold Josua Tungga	1	1
1230	SUNAN ASWAN HUTABARAT	1	2
1231	ELIA RIVALDO KAMODI	1	1
1232	SELVY NURAINI	3	3
1233	Niko Ledsia Motuti	1	1

1234 rows × 3 columns

Gambar 2. Data set

Gambar 2 menunjukan bahwa data yang telah di import sebanyak 1234 data kemudian diinisialisasikan agar dapat *clustering* pada tahap selanjutnya.

4.2. Penerapan Algoritma *K-Means*

Berikut langkah-langkah penerapan Algoritma *K-Means Cluster*ing :

1. Penentuan Jumlah Nilai *Cluster* (K)

   Penentuan jumlah nilai cluster merupakan langkah awal untuk memulai penerapan Algoritma K-Means Clustering. Penetuan nilai k disesuaikan dengan kebutuhan yang diinginkan. Pada penelitian ini nilai K=2.
2. Memilih Titik Pusat Awal

   Penentuan pusat centroid awal digunakan untuk pengelompokan atau *cluster* setiap data yang digunakan pada penelitian. Penentuan pusat centroid awal dilakukan secara acak sehingga bebas digunakan pada data keberapa. Dengan menggunakan centroid maka dapat di *cluster* data yang telah didapat menjadi 2 *cluster* [12]. Untuk

pusat centroid awal dapat dilihat pada tabel dibawah ini :

Proses Iterasi 1 :

Tabel 3. Data centroid awal

Centroid	Penghasilan	Jumlah Tanggungan
C0	0.07751092	0.08806405
C1	0.09734918	0.6750457

Tabel 3 merupakan hasil nilai perhitungan pusat centroid dari C0 dan C1 pada atribut penghasilan dan jumlah tanggungan.

3. Hitung Jarak Titik Pusat

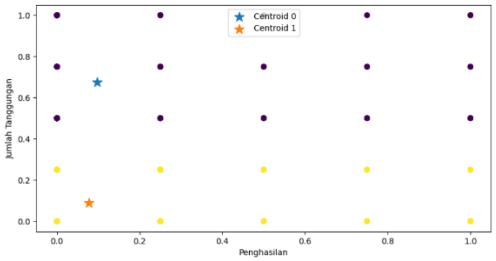
Perhitungan ini dilakukan dengan menggunakan *Sum of Squared Errors* (SSE) untuk mengukur sejauh mana setiap data dalam suatu *cluster* dari titik pusat *cluster*nya.

```
In [18]: inertias
Out[18]: [195.08230348460248,
          90.03769901460677,
          49.502910186233294,
          32.94163166628018,
          24.082089986160902,
          16.90145384216168,
          13.172305502441109,
          10.649285178660737,
          8.817863358094367]
```

Gambar 3. Jarak titik pusat

Gambar 3 merupakan hasil perhitungan jarak setiap data dari titik pusatnya dengan jumlah K= 2.

4. Hasil *Clustering*



Gambar 4. Hasil *clustering*

Dari gambar 4 tersebut menunjukan hasil pengelompokan menggunakan K-Means berdasarkan 2 cluster dengan warna ungu dan kuning.

5. Memilih Titik Pusat Baru K = 3

Proses Iterasi 2 :

Pemilihan titik pusat baru dilakukan untuk mengecek apakah nilai pada iterasi baru memiliki hasil yang sama dengan iterasi sebelumnya.

Tabel 4. Data centroid

Centroid	Penghasilan	Jumlah Tanggungan
C0	0,02146264	0,08823529
C1	0,02489837	0,67936992
C2	0,71460177	0,3539823

Tabel 4 merupakan hasil nilai perhitungan pusat centroid dari C0, C1, dan C2 pada atribut penghasilan dan jumlah tanggungan.

6. Hitung Jarak Titik Pusat

Pada proses iterasi kedua akhirnya dapat dilihat bahwa nilai inertias atau SSE sesuai dengan nilai pada proses iterasi pertama sehingga tidak diperlukan uji iterasi lagi. Berikut gambar hasil nilai inertias atau SSE pada proses iterasi 2 :

```
In [19]: inertias
Out[19]: [195.08230348460248,
          90.03769901460677,
          49.502910186233294,
          32.94163166628018,
          24.082089986160902,
          16.90145384216168,
          13.172305502441109,
          10.462504166094227,
          8.703617500610719]
```

Gambar 5. Jarak titik pusat

Gambar 5 merupakan hasil perhitungan jarak setiap data dari titik pusatnya dengan jumlah K=3.

Berdasarkan hasil uji iterasi 1 dan iterasi 2 maka dapat ditentukan bahwa pengelompokan data mahasiswa dibagi menjadi 2 *cluster*.

```
In [14]: dfmhs
Out[14]:
```

	Nama Siswa	Penghasilan	Jumlah Tanggungan	Cluster
0	YUVENTA MARTHA MEHA	1	1	1
1	Yuyun Meassy Felubun	1	2	1
2	VINA PAULINA SIMARMATA	1	3	0
3	TRI HIDAYANTI	1	1	1
4	DHEA PUTRI MYRINSTA RUMYAAN	1	3	0
...	...	...	...	...
1229	Reinhold Josua Tungga	1	1	1
1230	SUNAN ASWAN HUTABARAT	1	2	1
1231	ELIA RIVALDO KAMODI	1	1	1
1232	SELVY NURAINI	3	3	0
1233	Niko Ledsia Motuti	1	1	1

1234 rows × 4 columns

Gambar 6. Pengelompokan data mahasiswa 2 *cluster*

Gambar 6 menunjukan hasil pengelompokan data mahasiswa kedalam 2 cluster.

7. Statistika

```
Statistik Cluster 0:
       Penghasilan  Jumlah Tanggungan  Cluster
count   547.000000         547.000000    547.0
mean      1.389397           3.700183      0.0
std       0.955089           0.771315      0.0
min       1.000000           3.000000      0.0
25%       1.000000           3.000000      0.0
50%       1.000000           4.000000      0.0
75%       1.000000           4.000000      0.0
max       5.000000           5.000000      0.0

Statistik Cluster 1:
       Penghasilan  Jumlah Tanggungan  Cluster
count   687.000000         687.000000    687.0
mean      1.310044           1.352256      1.0
std       0.819728           0.478021      0.0
min       1.000000           1.000000      1.0
25%       1.000000           1.000000      1.0
50%       1.000000           1.000000      1.0
75%       1.000000           2.000000      1.0
max       5.000000           2.000000      1.0
```

Gambar 7. Hasil statistika K=2

Gambar 7 menunjukan hasil statistika cluster 0 dan cluster 1 dari atribut penghasilan dan jumlah tanggungan.

Pengelompokkan status mahasiswa akan dibagi menjadi 2 kategori sesuai dengan hasil *cluster*ing menggunakan algoritma *K-Means* yaitu Diterima dan Tidak Diterima. Pengelompokkan status mahasiswa dilihat berdasarkan hasil Statistika *cluster* 2.

Pada Statistika *Cluster* 0 dapat dilihat bahwa nilai rata-rata (*mean*) dari atribut Pengahasilan = 1.389397 dan nilai rata-rata dari atribut Jumlah Tanggungan = 3.700183. Sedangkan pada Statistika *Cluster* 1 dapat dilihat bahwa nilai rata-rata dari atribut Penghasilan = 1.310044 dan nilai rata-rata dari atribut Jumlah Tanggungan = 1.352256.

Berdasarkan hasil statistika diatas dapat disimpulkan bahwa nilai rata-rata dari Penghasilan dan Jumlah Tanggungan *Cluster* 0 lebih tinggi dibandingkan nilai rata-rata dari Penghasilan dan Jumlah Tanggungan *Cluster* 1. Hal tersebut dapat dilihat berdasarkan tabel inisialisasi 1 dan 2 bahwa semakin besar nilai bobot pada setiap atribut maka semakin kecil kemungkinan status mahasiswa dinyatakan sebagai penerima KIP Kuliah. Sehingga *Cluster* 0 masuk pada pengelompokkan status mahasiswa Diterima dan *Cluster* 1 masuk pada pengelompokkan status mahasiswa Tidak Diterima.

8. Hasil Pengelompokan Mahasiswa

Berikut gambar pengelompokkan penerima KIP Kuliah berdasarkan status Diterima dan Tidak Diterima :

```
In [23]: dfmhs
```

Out[23]:

	Nama Siswa	Penghasilan	Jumlah Tanggungan	Cluster
0	YUVENTA MARTHA MEHA	1	1	Diterima
1	Yuyun Meassy Felubun	1	2	Diterima
2	VINA PAULINA SIMARMATA	1	3	Tidak Diterima
3	TRI HIDAYANTI	1	1	Diterima
4	DHEA PUTRI MYRINSTA RUMYAAN	1	3	Tidak Diterima
...	...	...	...	...
1229	Reinhold Josua Tungga	1	1	Diterima
1230	SUNAN ASWAN HUTABARAT	1	2	Diterima
1231	ELIA RIVALDO KAMODI	1	1	Diterima
1232	SELVY NURAINI	3	3	Tidak Diterima
1233	Niko Ledsia Motuti	1	1	Diterima

1234 rows × 4 columns

Gambar 8. Hasil pengelompokan status mahasiswa

Dari gambar 8 dapat dilihat bahwa satiap nama mahasiswa telah berhasil dikelompokan kedalam cluster dengan status diterima dan tidak diterima.

Hasil clustering dengan 2 cluster mempunyai total data pembagian kedalam cluster sebagai berikut :

```
Total data pembagian kedalam cluster :
Diterima          687
Tidak Diterima    547
Name: Cluster, dtype: int64
```

Gambar 9. Total data pembagian cluster

Dari gambar 9 dapat dilihat bahwa jumlah cluster yang terbentuk dibagi menjadi 2. Cluster dengan kategori diterima sebanyak 687 data dan cluster dengan kategori tidak diterima sebanyak 547 data.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan oleh peneliti dapat disimpulkan bahwa dengan menerapkan metode algoritma *K-Means Cluster*ing dan penggunaan bahasa pemrograman *python* pada *tools Jupyter Notebook* dapat dilakukan pengelompokkan status mahasiswa penerima KIP menjadi 2 *cluster* yaitu diterima (*cluster* 1) dan tidak diterima (*cluster* 0) dari jumlah keseluruhan data sebanyak 1234 data. Dari 1234

data yang digunakan merupakan data pendaftar beasiswa KIP Kuliah tahun 2023. Pengolahan data yang dilakukan menghasilkan kelompok penerima KIP Kuliah yang dipertimbangkan menggunakan 2 atribut dalam penerapan algoritma *K-Means* yaitu penghasilan dan jumlah tanggungan. Dari data tersebut terdapat 687 data mahasiswa dengan status diterima dan 547 data mahasiswa dengan status tidak diterima. Saran unuk penelitian berikutnya sebaiknya menggunakan atribut yang lebih banyak agar hasil akurasi yang didapat bisa lebih baik dari penelitian sebelumnya. Penelitian ini dapat dikembangkan menggunakan algoritma lain kemudian membandingkannya dengan hasil penelitian menggunakan *K-Means Cluster*ing pada penelitian ini agar dapat mengetahui perbedaan serta mendapatkan evaluasi yang lebih baik.

DAFTAR PUSTAKA

[1] P. Tinggi dan K. I. P. K. Merdeka, "KARTU INDONESIA PINTAR KULIAH," 2023.

[2] J. N. Utamajaya, "Sistem Pendukung Keputusan Penerima Beasiswa Program Indonesia Pintar Menggunakan Metode Algoritma K-Means Clustering," vol. 9, no. 2, hal. 167–175, 2022, doi: 10.30865/jurikom.v9i2.3971.

[3] H. Kurniawan dan S. Defit, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," vol. 1, no. 2, hal. 80–89, 2020.

[4] C. Penerima, B. Kip, dan S. Usanto, "Penerapan Data Mining Dengan Mengimplementasikan Algoritma K- Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa," vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3411.

[5] D. T. Yuliana *et al.*, "Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah dengan Menggunakan Metode K-Means Clustering," vol. 5, no. 1, hal. 127–141, 2022, doi: 10.30762/f.

[6] S. Kasus, D. I. Universitas, dan M. Rawas, "DATA MINING PENERAPAN METODE K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN PENERIMA BANTUAN KARTU INDONESIA PINTAR," hal. 1095–1100, 2023.

[7] T. Elektro, U. A. Dahlan, S. Informasi, U. A. Dahlan, M. Informatika, dan U. A. Dahlan, "Penerapan Algoritma K - Means pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan efektif dan akurat , salah satu metode untuk pengelompokan data yaitu metode algoritma K - Penelitian terdahulu mengenai metode k-means yang dilakukan oleh Darlinda dan yayasan menggunakan metode k-means , dimana jumlah klaster dari hasil perhitungan," vol. 8, no. 2, hal. 352–366, 2022.

[8] M. S. Sompa dan R. Ishak, "Clustering Tingkat Ekonomi Mahasiswa Calon Penerima Kartu Indonesia Pintar (KIP) Kuliah Metode K-Means," vol. 1, no. 2, hal. 65–71, 2022.

[9] A. T. Rahman dan U. S. Maret, "Coal Trade Data Clusterung Using K-Means (Case Study PT . Global Bangkit Utama)," vol. 6, no. 1, hal. 24–31, 2017.

[10] A. Ilmiah, H. Riset, A. F. Febrianti, A. H. Cabral, dan G. Anuraga, "K-MEANS CLUSTERING DENGAN METODE ELBOW UNTUK PENGELOMPOKAN KABUPATEN DAN KOTA DI JAWA TIMUR," hal. 863–870, 2018.

[11] S. M. Dewi, A. P. Windarto, I. S. Damanik, dan H. Satria, "Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah," hal. 620–625, 2019.

[12] C. Astria, A. P. Windarto, A. Wanto, dan E. Irawan, "Metode K-Means Pada Pengelompokan Wilayah Pendistribusian Listrik," hal. 306–312, 2019.