

Analisis Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan (SPP) dengan Menggunakan Algoritma *K-Means* (Studi Kasus Politeknik TEDC Bandung)

Elfi Rosdianti¹, Castaka Agus Sugianto²

^{1,2} Program Studi Teknik Informatika- Politeknik TEDC Bandung

Jl. Politeknik-Pesantren KM2 Cibabat Cimahi Utara – Cimahi Jawa Barat - Indonesia

elfirosdianti01@gmail.com, castaka@poltektedc.ac.id

Abstrak— Sumbangan Pembinaan Pendidikan (SPP) adalah pembayaran yang dikenakan kepada mahasiswa setiap semester selama masa studi. Pada Politeknik TEDC Bandung biaya SPP untuk setiap mahasiswa dari tiap program studi adalah berbeda-beda. Namun pada penerapannya, masih terdapat beberapa mahasiswa yang mengalami keterlambatan dalam pembayarannya, dimana terdapat beberapa atribut yang menjadi alasan keterlambatan mahasiswa dalam melakukan pembayaran SPP. Oleh karena itu, pihak kampus perlu menerapkan kebijakan yang tepat untuk mengatasi masalah ini. Penulis tertarik untuk melakukan pengelompokan data menggunakan algoritma *K-Means* pada data mahasiswa yang mengalami keterlambatan pembayaran SPP. Berdasarkan hasil pengujian dengan metode *elbow* pada algoritma *K-Means*, ditemukan bahwa jumlah *cluster* terbaik adalah empat. *Cluster 0* terdiri dari 97 data yang didominasi oleh uang saku per bulan, *cluster 1* terdiri dari 3 data yang didominasi oleh program studi, *cluster 2* terdiri dari 46 data yang didominasi oleh keterlambatan pembayaran SPP, dan *cluster 3* terdiri dari 78 data yang didominasi oleh pendapatan orang tua. Pengujian kinerja dengan algoritma *K-Means* menunjukkan nilai *performance* berdasarkan nilai rata-rata jarak *centroid* adalah 14.604.447.397,425 dan nilai *Davies Bouldin Index* adalah 0,484. Hasil penelitian ini diharapkan dapat membantu pihak kampus dalam pengambilan keputusan terkait kebijakan pembayaran SPP bagi mahasiswa.

Kata Kunci— Sumbangan Pembinaan Pendidikan (SPP), Politeknik TEDC Bandung, Algoritma *K-Means*, Metode *Elbow*.

Abstract— *Educational Development Contributions (SPP) are fees charged to students every semester during their studies. At Politeknik TEDC Bandung, the SPP fees for each student vary by study program. However, in practice, some students still experience delays in their payments, where several attributes are the reasons for students' delays in paying the SPP. Therefore, the campus needs to implement appropriate policies to address this issue. The author is interested in clustering data using the K-Means algorithm on data from students who are late in paying the SPP. Based on tests using the elbow method on the K-Means algorithm, it was found that the best number of clusters is four. Cluster 0 consists of 97 data points dominated by monthly pocket money, cluster 1 consists of 3 data points dominated by study programs, cluster 2*

consists of 46 data points dominated by delays in SPP payments, and cluster 3 consists of 78 data points dominated by parents' income. Performance testing with the K-Means algorithm shows a performance value based on the average within-cluster distance to the centroid of 14,604,447,397.425 and a Davies Bouldin Index of 0.484. The results of this research are expected to help the campus in making decisions regarding SPP payment policies for students.

Keywords— *Educational Development Contribution (SPP), TEDC Bandung Polytechnic, K-Means Algorithm, Elbow Method.*

I. PENDAHULUAN

Pembayaran biaya kuliah merupakan sumber dana yang penting untuk mendukung operasional perguruan tinggi. Mengingat pentingnya pembayaran yang dilakukan oleh mahasiswa, terutama di perguruan tinggi swasta di mana biaya operasional sebagian besar ditanggung oleh kampus itu sendiri, maka pengelolaan pembayaran harus dilakukan dengan baik.

Politeknik TEDC Bandung, sebagai penyedia fasilitas pendidikan, memiliki kebijakan terkait pembayaran Sumbangan Pembinaan Pendidikan (SPP). SPP adalah biaya yang harus dibayar oleh mahasiswa setiap semester selama masa studi mereka. Besarnya SPP bervariasi antara mahasiswa program diploma/sarjana dan pascasarjana, serta antara mahasiswa lokal dan internasional [1]. Pada awal setiap semester, mahasiswa di Politeknik TEDC harus melakukan registrasi bagi mahasiswa baru dan herregistrasi bagi mahasiswa yang sudah terdaftar sebelumnya. Mahasiswa dapat membayar biaya semester secara penuh, atau dengan mencicil, di mana 50% dari jumlah SPP per semester dibayar saat herregistrasi, kemudian 75% dibayar saat UTS, dan dilunasi saat UAS. Kampus berharap pembayaran dilakukan tepat waktu karena dana SPP digunakan untuk operasional kampus.

Berdasarkan hasil observasi dan kuesioner, ditemukan bahwa masalah yang muncul di Politeknik TEDC adalah kecenderungan mahasiswa untuk menunda pembayaran SPP pada awal semester baru. Walaupun jumlah keterlambatan tidak signifikan, fenomena ini terus meningkat setiap semester.

Keterlambatan pembayaran SPP di Politeknik TEDC Bandung telah dicatat dalam data historis keuangan, yang menciptakan tantangan bagi manajemen institusi.

Untuk mengatasi permasalahan ini, perlu dilakukan pendekatan analisis yang lebih mendalam terhadap data keterlambatan pembayaran. Sebagai solusi, direkomendasikan penerapan teknik data *mining* dengan fokus pada *cluster* data. Dengan demikian, Politeknik TEDC dapat mengidentifikasi pola keterlambatan pembayaran SPP dan mengambil tindakan preventif yang lebih efektif. Langkah ini diharapkan dapat mengurangi angka keterlambatan pembayaran di masa mendatang, memastikan kelancaran administrasi keuangan institusi.

Data mining adalah proses yang memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi serta mengidentifikasi informasi berharga dan pengetahuan dari berbagai database besar [2]. Terdapat banyak teknik data mining yang telah dikembangkan dan diterapkan. Data mining berfungsi dengan menemukan pola-pola pengetahuan melalui berbagai metode. Beberapa peran penting dalam data mining adalah estimasi, prediksi, clustering, dan asosiasi. Setiap teknik data mining memiliki aturan dan metodologi yang unik yang menentukan jenis masalah yang dapat diselesaikan [3].

Dalam penelitian ini menggunakan metode *clustering* yang bertujuan untuk membagi atau mengelompokkan objek ke dalam *cluster* tertentu dengan tingkat kemiripan yang tinggi. Objek akan dikelompokkan ke dalam satu atau lebih *cluster* sehingga objek-objek yang berada di dalam satu *cluster* akan mempunyai kesamaan yang tinggi antara satu dengan yang lainnya. Kesamaan objek biasanya diperoleh dari nilai-nilai atribut yang menjelaskan objek data [4].

Algoritma *K-Means* merupakan teknik *clustering* yang paling sering digunakan dalam penelitian [5]. Algoritma *K-Means* adalah metode *clustering* yang paling sering digunakan dalam penelitian. Algoritma *K-Means* adalah salah satu teknik *clustering* non-hierarkis yang bertujuan untuk mempartisi atau mengelompokkan data ke dalam *cluster* dengan karakteristik yang sama, sedangkan data dengan karakteristik berbeda akan dikelompokkan ke dalam *cluster* lainnya [6].

Algoritma *K-Means* memiliki kelebihan yaitu skalabilitas yang baik untuk menangani dataset besar, implementasi yang mudah, dan kemampuan deteksi anomaly terhadap data, yang dapat membantu mengidentifikasi potensi risiko pembayaran yang tinggi. Fleksibilitas algoritma ini terhadap perubahan *cluster* memungkinkan adaptasi terhadap perubahan pola pembayaran dari waktu ke waktu. Visualisasi hasil clustering juga memudahkan interpretasi dan pengambilan keputusan, sementara pemahaman yang lebih mendalam terhadap setiap *cluster* memungkinkan pengembangan solusi atau strategi yang lebih spesifik untuk setiap kelompok siswa [7].

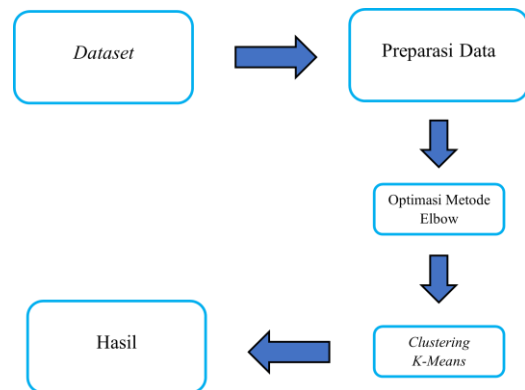
Menurut penelitian, algoritma *K-Means* memiliki keunggulan dibandingkan algoritma *K-Medoids* [8]. Dari segi efisiensi waktu, algoritma *K-Means* dapat melakukan *clustering* lebih cepat daripada algoritma SFCM [9]. Dalam penelitian yang dilakukan oleh Defiyanti, ditemukan bahwa tingkat akurasi *clustering* algoritma *K-Means* memiliki nilai

purity measure sebesar 80,56%, sedangkan untuk *dataset* asli nilai *purity* adalah 75% [10]. Hal ini menunjukkan bahwa *dataset* yang dikoordinasikan secara keseluruhan lebih baik dibandingkan *dataset* yang hanya sebagian dikodifikasi. Proses *clustering* dengan menggunakan *K-Means* juga lebih cepat dan sederhana. Sementara itu, algoritma *K-Medoids* menggunakan *medoid* (*median*) dan tidak menghitung nilai rata-rata dari objek dalam *cluster*, sehingga mengurangi ketidaksamaan antara setiap objek dengan titik acuan yang sesuai (*medoid*) [10].

Berdasarkan latar belakang di atas maka penulis mengajukan tugas akhir yang berjudul “Analisis Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan (SPP) dengan Menggunakan Algoritma *K-Means* (Studi Kasus Politeknik TEDC Bandung).

II. METODE PENELITIAN

Proses pengolahan data menggunakan algoritma *K-Means* pada penelitian ini ditunjukkan pada gambar 1.



Gbr. 1 Metode Penelitian

Berdasarkan gambar 1, terdapat beberapa tahapan yaitu diantaranya sebagai berikut.

A. Dataset

Penelitian ini menggunakan *dataset* yang mencakup data mahasiswa Politeknik TEDC yang sedang menjalani kuliah, termasuk mereka yang sudah membayar SPP dan yang belum. Data tersebut diperoleh dari hasil observasi dan penyebaran kuisioner di Politeknik TEDC. Data yang diperoleh berjumlah 200 data, dengan *field* yang terdiri atas nomor, nama lengkap, jenis kelamin, alamat, program studi, semester, angkatan, status, pekerjaan orang tua, pendapatan ayah, pendapatan ibu, jumlah saudara, tempat tinggal, hobi, pemenuhan konsumsi per hari, uang saku per bulan, keterangan pembayaran spp, dan spp yang belum di bayarkan. Detailnya dapat dilihat pada tabel 1 di bawah ini.

TABEL I
DATA KUISIONER MAHASISWA

No.	Nama Lengkap	Jenis Kelamin	...	SPP yang Belum Dibayar		
				1	...	8
1.	Taya Vania	Perempuan	...	-		
2.	Ishmah Nailufar	Perempuan	...	-		
3.	Akbar Maulana	Laki - Laki	...	-		
4.	Taufik Ardiansyah	Laki - Laki	...	-		
5.	Resa Hardiyanti	Perempuan	...	-		
6.	Amelia Firstiana	Perempuan	...	6 dan 7		
7.	Fadly Ikram Sulaiman	Laki-Laki	...	6		
8.	Farhan Maulana Yusuf	Laki-Laki	...	7		
9.	Gina Fajriani	Perempuan	...	5		
⋮	⋮	⋮	⋮	⋮		
200	Fitria Azzahra	Perempuan	...	-		

B. Preparasi Data

Pada tahap ini, data yang terkumpul akan dibersihkan dengan mengeliminasi informasi yang tidak relevan, sehingga data siap untuk diolah. Proses ini penting karena algoritma *K-Means* hanya dapat memproses data numerik. Terdapat 200 *record* dengan 18 variabel, yaitu nomor, nama lengkap, jenis kelamin, alamat, program studi, semester, angkatan, status, pekerjaan orang tua, pendapatan ayah, pendapatan ibu, jumlah saudara, tempat tinggal, hobi, pemenuhan konsumsi, uang saku, keterangan pembayaran SPP, dan SPP yang belum dibayar.

Penulis hanya menggunakan 12 variabel pada penelitian ini yaitu terdiri atas jenis kelamin, program studi, semester, pendapatan ayah, pendapatan ibu, jumlah saudara, tempat tinggal, pemenuhan konsumsi per hari, uang saku per bulan, keterangan pembayaran SPP, dan SPP yang belum dibayar. Data yang telah dipreparasi detailnya dapat dilihat pada tabel 2.

TABEL II
HASIL DATA PREPARASI

Jenis Kelamin	Program Studi	Semester	Status	Pendapatan Ayah	Pendapatan Ibu	Jumlah Saudara	...	SPP yang Belum Dibayar		
								1	...	8
1	6	5	0	2	4	1	...	0	...	0
1	6	5	0	2	2	1	...	0	...	0
1	6	5	0	2	4	1	...	0	...	0
0	0	7	0	4	4	1	...	0	...	0
0	0	7	0	4	2	1	...	0	...	0
1	0	5	0	4	2	1	...	0	...	0
1	5	7	0	2	4	2	...	0	...	0

Jenis Kelamin	Program Studi	Semester	Status	Pendapatan Ayah	Pendapatan Ibu	Jumlah Saudara	...	SPP yang Belum Dibayar		
								1	...	8
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	2	5	0	2	3	2	...	0	...	0

C. Optimasi Metode Elbow

Pada tahap ini, data yang telah diperoleh akan diolah untuk menentukan jumlah *cluster* optimal pada data tersebut. Proses ini sangat penting dalam analisis *clustering* karena menentukan berapa banyak *cluster* yang terbaik untuk membagi data secara efektif.

Langkah pertama dalam proses ini adalah melakukan uji coba dengan berbagai parameter awal. Parameter ini bisa berupa jumlah *cluster* yang berbeda, misalnya mulai dari 2 hingga nilai maksimal yang masuk akal berdasarkan ukuran *dataset*. Setiap kali parameter ini diuji, algoritma *clustering* akan menghasilkan nilai tertentu yang dikenal sebagai nilai *avg. within distance centroid*. Nilai ini mengukur seberapa dekat data dalam satu *cluster* dengan pusat (*centroid*) dari *cluster* tersebut. Semakin kecil nilai ini, semakin homogen atau serupa data dalam *cluster* tersebut.

Setelah semua parameter diuji, langkah selanjutnya adalah mengonversi hasil dari nilai *avg. within distance centroid* tersebut ke dalam bentuk grafik. Grafik ini biasanya adalah plot dari jumlah *cluster* (sumbu x) terhadap nilai *avg. within distance centroid* (sumbu y). Dengan memplot data ini, kita mencari suatu titik dalam grafik yang membentuk sudut atau siku (*elbow*).

Titik *elbow* ini menunjukkan jumlah *cluster* yang optimal. Sebab pada titik tersebut, penurunan nilai *avg. within distance centroid* mulai melambat secara signifikan. Artinya, menambahkan lebih banyak *cluster* setelah titik *elbow* tidak akan memberikan pengurangan yang signifikan pada nilai *avg. within distance centroid*. Ini mengindikasikan bahwa titik *elbow* adalah jumlah *cluster* yang efisien untuk menggambarkan struktur data yang ada. Dengan demikian, melalui proses ini, kita dapat menentukan jumlah *cluster* optimal yang tidak hanya meningkatkan keakuratan *clustering*, tetapi juga efisien dalam hal komputasi dan interpretasi hasil.

D. Clustering K-Means

Penelitian ini menggunakan aplikasi *RapidMiner* karena kemampuannya dalam melakukan analisis data *mining*, *text mining*, dan prediksi [11]. Dalam aplikasi *RapidMiner*, penulis akan mengelompokkan data mahasiswa terkait pembayaran SPP menggunakan algoritma *K-Means*. Berikut adalah algoritma metode *K-Means* [12]:

- 1) Masukkan data yang akan diklaster.
- 2) Tentukan jumlah *cluster* menggunakan metode *elbow*.
- 3) Pilih data secara acak sebanyak jumlah *cluster* sebagai pusat *cluster* (*centroid*).

4) Hitung jarak antara data dengan pusat *cluster* menggunakan persamaan:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Keterangan :

D (i,j) = Jarak data ke i ke pusat *cluster* j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

5) Hitung ulang pusat *cluster* dengan keanggotaan *cluster* yang baru.

6) Jika pusat *cluster* tidak berubah, proses *clustering* selesai. Jika belum, ulangi langkah 4 hingga pusat *cluster* stabil.

E. Hasil

Penelitian ini menghasilkan *cluster* data mahasiswa yang berkaitan dengan pembayaran SPP di Politeknik TEDC. Selain itu, penelitian ini juga mencakup pengujian algoritma *K-Means* untuk *clustering* data mahasiswa terkait pembayaran SPP, yang bertujuan menemukan nilai k optimal dan menentukan jumlah *cluster* terbaik dalam pengelompokan data tersebut.

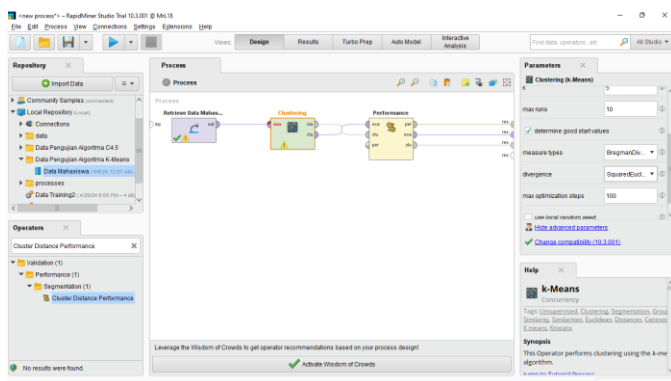
III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Hasil penelitian merupakan sub bagian yang membahas tiga tahapan yaitu penentuan jumlah *cluster* terbaik, pengujian menggunakan algoritma *k-means*, dan hasil data *cluster k-means*.

1) Penentuan Jumlah *Cluster* Terbaik

Pada tahap ini, metode *elbow* digunakan untuk menguji dan menentukan hasil *cluster* yang paling optimal dengan memasukkan data yang telah dipersiapkan sebelumnya ke dalam *software RapidMiner* dan menyeleksi data yang akan digunakan. Data yang akan diproses dengan *RapidMiner* dapat dilihat pada gambar 2 berikut.

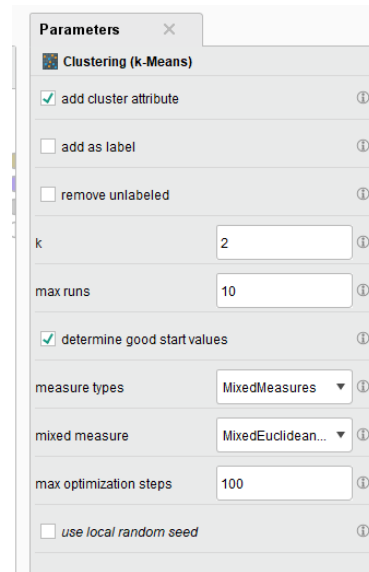


Gbr. 2 Pengolahan Data pada *RapidMiner*

Berdasarkan gambar 2, data yang akan diuji akan dihubungkan dengan opsi *K-Means (Clustering)* lalu dihubungkan ke *Cluster Distance Performance* dan *Cluster Distance Performance* disambungkan ke *result*.

Langkah selanjutnya adalah menyesuaikan parameter *clustering* dengan jumlah pengujian yang direncanakan. Pada tahap ini, penulis akan melakukan 10 pengujian, sehingga

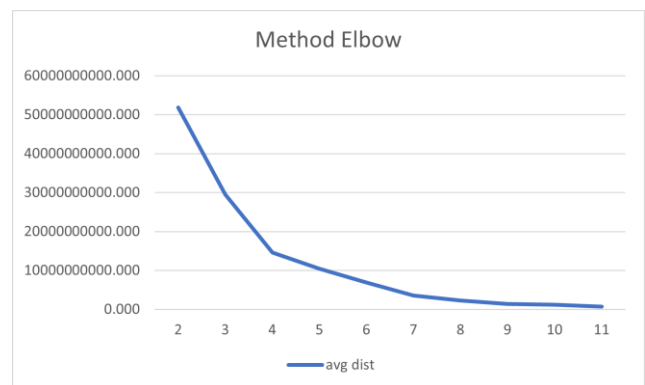
nilai k akan diubah dari 2 hingga 11 untuk mendapatkan nilai rata-rata jarak dalam kluster sebagai dasar untuk pengujian metode *elbow*. Jenis pengukuran diubah menjadi *MixedMeasures* untuk menyesuaikan dengan jenis atribut yang digunakan. Detailnya dapat dilihat pada gambar 3 berikut.



Gbr. 3 Mengatur Parameter *Clustering*

2) Pengujian Algoritma *K-Means* Menggunakan Metode *Elbow*

Hasil pengujian dengan parameter k=2 hingga k=11 menunjukkan bahwa nilai rata-rata jarak dalam kluster akan dikonversi menjadi diagram untuk mengidentifikasi kurva dengan sudut siku paling jelas yang dihasilkan oleh metode *elbow*, sehingga dapat menentukan nilai k yang tepat untuk *clustering* data. Pengujian menggunakan metode *elbow* menemukan bahwa nilai k optimal adalah 4, dengan rata-rata jarak dalam kluster sebesar 14.604.447.397,425. Detail hasilnya dapat dilihat pada gambar 4 berikut.



Gbr. 4 Hasil Pengujian Metode *Elbow*

Data hasil pengujian menggunakan metode *elbow* dari 10 parameter *cluster* yang diuji dapat dilihat pada tabel 3 berikut.

TABEL III
HASIL PENGUJIAN METODE *ELBOW*

No.	Jumlah Cluster	Nilai Avg. Within Centroid Distance	Davies-Bouldin Index (DBI)
1.	2	51.817.134.942,890	0.539
2.	3	29.471.098.943,136	0.470
3.	4	14.604.447.397,425	0.484
4.	5	10.459.845.767,451	0.484
5.	6	6.944.404.308,515	0.379
6.	7	3.528.378.667,064	0.314
7.	8	2.222.602.804,933	0.317
8.	9	1.380.390.681,041	0.325
9.	10	1.172.582.247,052	0.307
10.	11	738.874.687,052	0.215

3) Hasil Data Cluster *K-Means*

Dari hasil proses penentuan cluster menggunakan metode *elbow*, ditemukan bahwa jumlah cluster optimal adalah empat. Data dalam penelitian ini terbagi menjadi empat kelompok: cluster 0 dengan 41 data, cluster 1 dengan 70 data, cluster 2 dengan 86 data, dan cluster 3 dengan 3 data, totalnya menjadi 220 data. Nilai rata-rata jarak dalam klaster adalah 14.604.447.397,425 dan nilai *Davies Bouldin Index* adalah 0,484. Pengelompokan data ini ditunjukkan pada Gambar 5 dan Gambar 6 berikut.

Cluster Model

```
Cluster 0: 41 items
Cluster 1: 70 items
Cluster 2: 86 items
Cluster 3: 3 items
Total number of items: 200
```

Gbr. 5 Hasil Pengujian Data Berdasarkan Cluster

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 14604447397.425
Avg. within centroid distance_cluster_0: 17311124347.477
Avg. within centroid distance_cluster_1: 12455836757.732
Avg. within centroid distance_cluster_2: 7820443505.657
Avg. within centroid distance_cluster_3: 22222222236.889
Davies Bouldin: 0.484
```

Gbr. 6 Hasil Performance Vector

B. Pembahasan

Pada bagian ini akan dibahas menjadi tiga tahapan yaitu analisis hasil algoritma *K-Means*, pengelompokan data mahasiswa yang mengalami keterlambatan pembayaran SPP dengan algoritma *K-Means*, dan upaya untuk mengurangi keterlambatan pembayaran SPP.

1) Analisis Hasil Algoritma *K-Means*

Berdasarkan pengujian yang dilakukan oleh penulis dengan menggunakan metode *elbow* pada algoritma *K-Means*, ditemukan bahwa 4 cluster adalah yang paling optimal. Dari 220 data mahasiswa, terbentuk 4 kelompok data yaitu cluster 0 yang didominasi oleh uang saku per bulan, cluster 1 yang didominasi oleh program studi, cluster 2 yang didominasi oleh keterlambatan pembayaran SPP, dan cluster 3 yang didominasi oleh pendapatan orang tua.

2) Pengelompokan Data Mahasiswa Berdasarkan Cluster

Berdasarkan hasil pengujian dengan metode *elbow*, didapatkan hasil terbaik yaitu dengan k=4. Kategorisasi didasarkan pada uang saku per bulan, program studi, keterlambatan pembayaran SPP, dan pendapatan orang tua (ayah dan ibu). Oleh karena itu, dataset dibagi menjadi empat kelompok cluster berdasarkan dominasi data masing-masing. Berikut adalah penjelasan pengelompokan data mahasiswa berdasarkan masing-masing cluster, di mana setiap cluster memiliki atribut yang mendominasi.

Pengelompokan data mahasiswa berdasarkan uang saku per bulan menunjukkan bahwa cluster 0 didominasi oleh 33 data dengan uang saku sebesar Rp. 700.000 – Rp. 980.000. Cluster 1 terdiri dari 40 data dengan uang saku Rp. 2.000.000 – Rp. 2.800.000. Cluster 2 terdiri dari 49 data dengan uang saku Rp. 1.100.000 – Rp. 1.380.000. Cluster 3 terdiri dari 2 data dengan uang saku Rp. 480.000 – Rp. 760.000. Maka, rata-rata uang saku per bulan mahasiswa adalah Rp. 809.600. Detail lebih lanjut dapat dilihat pada tabel 4 berikut.

TABEL IV
PENGELOMPOKKAN DATA BERDASARKAN UANG SAKU PER BULAN

Pengelompokan Data		Jumlah
Cluster	Range Uang Saku	
Cluster_0	Rp. 700.000 - Rp. 980.000	33
	Rp. 980.000 - Rp. 1.260.000	8
Cluster_1	Rp. 2.000.000 - Rp. 2.800.000	40
	Rp. 2.840.000 - Rp. 3.120.000	30
Cluster_2	Rp. 1.100.000 - Rp. 1.380.000	49
	Rp. 1.380.000 - Rp. 1.660.000	37
Cluster_3	Rp. 200.000 - Rp. 480.000	1
	Rp. 480.000 - Rp. 760.000	2

Pengelompokan data mahasiswa berdasarkan program studi didominasi oleh cluster_1 dengan 23 data yang memiliki prodi D4 – Mekanik Industri dan Desain. Lalu untuk cluster_0 dengan 8 data yang memiliki prodi D4 – Teknik Informatika dan D4 – Komputerisasi Akuntansi. Untuk cluster_2 dengan 15 data yang memiliki prodi D4 – Mekanik Industri dan Desain. Dan Untuk cluster_3 dengan 3 data yang memiliki prodi D3 – Teknik Komputer. Maka rata-rata program studi yang mengalami keterlambatan pembayaran SPP adalah prodi D4 - Mekanik Industri dan Desain. Untuk lebih jelasnya dapat dilihat pada tabel 5 berikut.

TABEL V
PENGELOMPOKKAN DATA BERDASARKAN PROGRAM STUDI

Pengelompokan Data		Jumlah
Cluster	Program Studi	
Cluster_0	D4 - Teknik Informatika	8
	D4 - Teknik Otomasi	2
Cluster_0	D4 - Komputerisasi Akuntansi	8
	D4 - Konstruksi Bangunan	5
	D4 - Mekanik Industri dan Desain	3
	D3 - Teknik Mesin	3
	D3 - Teknik Komputer	5
	D3 - Teknik Kimia	5
	D3 - Rekam Medik dan Informasi Kesehatan	1
	D3 - Akuntansi	0
	D3 - Mesin Otomotif	0
	D3 - Teknik Elektronika	1
Cluster_1	D4 - Teknik Informatika	5
	D4 - Teknik Otomasi	5
	D4 - Komputerisasi Akuntansi	1
	D4 - Konstruksi Bangunan	3
	D4 - Mekanik Industri dan Desain	23
	D3 - Teknik Mesin	6
	D3 - Teknik Komputer	2
	D3 - Teknik Kimia	7
	D3 - Rekam Medik dan Informasi Kesehatan	3
	D3 - Akuntansi	3
	D3 - Mesin Otomotif	5
	D3 - Teknik Elektronika	7
Cluster_2	D4 - Teknik Informatika	10
	D4 - Teknik Otomasi	4
	D4 - Komputerisasi Akuntansi	4
	D4 - Konstruksi Bangunan	4
	D4 - Mekanik Industri dan Desain	15
	D3 - Teknik Mesin	5
	D3 - Teknik Komputer	7
	D3 - Teknik Kimia	4
	D3 - Rekam Medik dan Informasi Kesehatan	10
	D3 - Akuntansi	7
D3 - Mesin Otomotif	10	
Cluster_3	D4 - Teknik Informatika	0
	D4 - Teknik Otomasi	0
	D4 - Komputerisasi Akuntansi	0
	D4 - Konstruksi Bangunan	0
	D4 - Mekanik Industri dan Desain	0
	D3 - Teknik Mesin	0
	D3 - Teknik Komputer	3
	D3 - Teknik Kimia	0

Pengelompokan Data		Jumlah
Cluster	Program Studi	
	D3 - Rekam Medik dan Informasi Kesehatan	0
	D3 - Akuntansi	0
	D3 - Mesin Otomotif	0
	D3 - Teknik Elektronika	0

Pengelompokan data mahasiswa berdasarkan program studi didominasi oleh *cluster_2* dengan 49 data yang sudah membayar. Lalu untuk *cluster_0* dengan 24 data yang sudah membayar. Untuk *cluster_1* dengan 48 data yang sudah membayar. Dan Untuk *cluster_3* dengan 3 data yang sudah membayar. Maka rata-rata mahasiswa telah membayar SPP. Detailnya dapat dilihat pada tabel 6 berikut.

TABEL VI
PENGELOMPOKKAN DATA BERDASARKAN KETERLAMBATAN PEMBAYARAN SPP

Pengelompokan Data		Jumlah
Cluster	Keterangan	
Cluster_0	Sudah	24
	Belum	17
Cluster_1	Sudah	48
	Belum	22
Cluster_2	Sudah	49
	Belum	37
Cluster_3	Sudah	3
	Belum	0

Pengelompokan data mahasiswa berdasarkan pendapatan ayah menunjukkan bahwa *cluster 3* didominasi oleh 3 data dengan pendapatan Rp 1.000.000 - Rp 3.000.000. *Cluster 0* terdiri dari 21 data dengan pendapatan yang sama. *Cluster 1* mencakup 40 data dengan pendapatan Rp 3.000.000 - Rp 5.000.000, sementara *cluster 2* terdiri dari 24 data dengan pendapatan di atas Rp 5.000.000. Dengan demikian, rata-rata pendapatan ayah dari setiap mahasiswa adalah Rp 1.000.000 - Rp 3.000.000. Detail lebih lanjut dapat dilihat pada tabel 7 berikut.

TABEL VII
PENGELOMPOKKAN DATA BERDASARKAN PENDAPATAN AYAH

Pengelompokan Data		Jumlah
Cluster	Range Pendapatan	
Cluster_0	Dibawah Rp 500.000	0
	Rp 500.000 - Rp 1.000.000	10
	Rp 1.000.000 - Rp 3.000.000	21
	Rp 3.000.000 - Rp 5.000.000	6
	Diatas Rp 5.000.000	4
Cluster_1	Dibawah Rp 500.000	10
	Rp 500.000 - Rp 1.000.000	4
	Rp 1.000.000 - Rp 3.000.000	12
	Rp 3.000.000 - Rp 5.000.000	40
	Diatas Rp 5.000.000	4
Cluster_2	Dibawah Rp 500.000	15
	Rp 500.000 - Rp 1.000.000	10
	Rp 1.000.000 - Rp 3.000.000	18
	Rp 3.000.000 - Rp 5.000.000	19

Pengelompokan Data		Jumlah
Cluster	Range Pendapatan	
	Diatas Rp 5.000.000	24
Cluster_3	Dibawah Rp 500.000	0
	Rp 500.000 - Rp 1.000.000	0
	Rp 1.000.000 - Rp 3.000.000	3
	Rp 3.000.000 - Rp 5.000.000	0
	Diatas Rp 5.000.000	0

Pengelompokan data mahasiswa berdasarkan pendapatan ibu menunjukkan bahwa *cluster* 3 didominasi oleh 2 data dengan pendapatan di atas Rp 3.000.000. *Cluster* 0 terdiri dari 23 data dengan pendapatan yang sama. *Cluster* 1 mencakup 41 data dengan pendapatan di atas Rp 3.000.000, sementara *cluster* 2 terdiri dari 51 data dengan pendapatan Rp 1.000.000 - Rp 2.000.000. Dengan demikian, rata-rata pendapatan ibu dari setiap mahasiswa adalah Rp 1.000.000 - Rp 2.000.000. Untuk detail lebih lanjut dapat dilihat pada tabel 8 berikut.

TABEL VIII
PENGELOMPOKAN DATA BERDASARKAN PENDAPATAN IBU

Pengelompokan Data		Jumlah
Cluster	Range Pendapatan	
Cluster_0	Dibawah Rp 500.000	4
	Rp 500.000 - Rp 1.000.000	2
	Rp 1.000.000 - Rp 2.000.000	9
	Rp 2.000.000 - Rp 3.000.000	3
	Diatas Rp 3.000.000	23
Cluster_1	Dibawah Rp 500.000	18
	Rp 500.000 - Rp 1.000.000	4
	Rp 1.000.000 - Rp 2.000.000	3
	Rp 2.000.000 - Rp 3.000.000	4
	Diatas Rp 3.000.000	41
Cluster_2	Dibawah Rp 500.000	14
	Rp 500.000 - Rp 1.000.000	3
	Rp 1.000.000 - Rp 2.000.000	51
	Rp 2.000.000 - Rp 3.000.000	3
	Diatas Rp 3.000.000	15
Cluster_3	Dibawah Rp 500.000	0
	Rp 500.000 - Rp 1.000.000	0
	Rp 1.000.000 - Rp 2.000.000	1
	Rp 2.000.000 - Rp 3.000.000	0
	Diatas Rp 3.000.000	2

3) Upaya Mengurangi Keterlambatan Pembayaran SPP

Berdasarkan hasil pengelompokan *cluster*, maka diperoleh 4 pengelompokan data berdasarkan dominasi *cluster* yaitu *cluster_0* didominasi oleh uang saku per bulan, *cluster_1* didominasi oleh program studi, *cluster_2* didominasi oleh keterlambatan pembayaran SPP, dan *cluster_3* didominasi oleh pendapatan orang tua atau pendapatan ayah dan ibu. Oleh karena itu, diperlukan upaya atau solusi untuk mengurangi atau mengantisipasi hal tersebut.

Menurut penulis beberapa upaya yang perlu dilakukan untuk mengurangi keterlambatan pembayaran SPP oleh mahasiswa adalah sebagai berikut:

- 1) Menerapkan sistem pembayaran cicilan untuk SPP yang memungkinkan mahasiswa membayar dalam beberapa kali angsuran selama satu semester atau tahun akademik.
- 2) Menerapkan skema subsidi atau diskon SPP bagi mahasiswa dari keluarga berpendapatan rendah. Tujuan langkah ini adalah untuk mendukung mahasiswa dari latar belakang ekonomi kurang mampu agar tetap dapat mengakses pendidikan tinggi tanpa beban biaya yang tinggi. Dengan adanya subsidi atau diskon SPP, diharapkan beban keuangan mahasiswa dan keluarganya dapat berkurang, sehingga mereka dapat fokus pada proses belajar-mengajar.
- 3) Mengadakan penyuluhan atau *workshop* tentang pentingnya pengelolaan keuangan pribadi termaksud strategi untuk mengatur keuangan dan memprioritaskan pembayaran SPP tepat waktu serta konsekuensinya bagi mahasiswa yang terlambat melakukan pembayaran SPP.
- 4) Tiap prodi atau program studi dapat mengadakan bimbingan tiap awal semester atau memaksimalkan peran wali masiswa dalam konseling terhadap setiap mahasiswanya setiap awal ataupun akhir semester sehingga dapat membantu memantau perkembangan akademik siswa dan membantu menyelesaikan serta memberi solusi bagi mahasiswa yang bermasalah.

IV. PENUTUP

A. Kesimpulan

Berdasarkan pengujian dan pengelompokan data mahasiswa untuk mengetahui keterlambatan pembayaran SPP menggunakan metode *elbow* pada algoritma *k-means*, dapat diambil kesimpulan sebagai berikut:

- 1) Pengujian dengan metode *elbow* pada algoritma *k-means* menghasilkan empat *cluster* terbaik. Pengelompokan dibagi menjadi empat kelompok: *cluster* 0 terdiri dari 33 data dengan uang saku Rp. 700.000 - Rp. 980.000, 8 data dari program studi D4 - Teknik Informatika dan D4 - Komputerisasi Akuntansi, 17 data dengan keterlambatan pembayaran, 21 data dengan pendapatan ayah Rp. 1.000.000 - Rp. 3.000.000, dan 23 data dengan pendapatan ibu di atas Rp. 3.000.000. *Cluster* 1 terdiri dari 40 data dengan uang saku Rp. 2.840.000 - Rp. 3.120.000, 23 data dari program studi D4 - Mekanik Industri dan Desain, 22 data dengan keterlambatan pembayaran, 40 data dengan pendapatan ayah Rp. 3.000.000 - Rp. 5.000.000, dan 41 data dengan pendapatan ibu di atas Rp. 3.000.000. *Cluster* 2 terdiri dari 49 data dengan uang saku Rp. 1.100.000 - Rp. 1.380.000, 15 data dari program studi D4 - Mekanik Industri dan Desain, 37 data dengan keterlambatan pembayaran, 24 data dengan pendapatan ayah di atas Rp. 5.000.000, dan 51 data dengan pendapatan ibu Rp. 1.000.000 - Rp. 2.000.000. *Cluster* 3 terdiri dari 2 data dengan uang saku Rp. 480.000 - Rp. 760.000, 3 data dari program studi D3 - Teknik Komputer, 0 data dengan keterlambatan pembayaran, 3 data dengan pendapatan ayah Rp. 1.000.000 - Rp. 3.000.000, dan 2 data dengan pendapatan ibu di atas Rp. 3.000.000.

2) Berdasarkan pengujian yang telah dilakukan oleh penulis menggunakan metode *elbow* untuk memperoleh nilai *k* atau *cluster* yang terbaik, dimana pengujian tersebut dilakukan sebanyak 10 kali percobaan. Dari 10 percobaan tersebut, setelah dikonversi kedalam bentuk grafik untuk mengetahui *k* atau *cluster* mana yang memiliki bentuk yang menyerupai siku sempurna, maka diperoleh hasil $k=4$. Lalu pengujian untuk mengetahui *performance* dengan algoritma *k-means* menghasilkan nilai *performance* berdasarkan nilai *avg. within distance centroid* adalah 14.604.447.397,425 dan nilai *Davies Bouldin Index* adalah 0.484 sehingga data tersebut termasuk kedalam hasil pengujian yang baik.

B. Saran

Adapun beberapa saran untuk pengembangan atau penelitian selanjutnya adalah sebagai berikut:

- 1) Untuk penelitian berikutnya cobalah untuk menerapkan algoritma yang lain misalnya *Fuzzy C-Means*. Hal ini agar dapat memperoleh hasil yang lebih baik kemudian membandingkan dengan penelitian yang sekarang.
- 2) Pada penelitian berikutnya diharapkan memperoleh data mahasiswa yang lebih banyak lagi khususnya bagi mahasiswa yang telah lulus ataupun alumni agar semakin memperluas jangkauan pengujian guna memperoleh data *clustering* atau pengelompokan yang lebih beragam.
- 3) Untuk penelitian berikutnya lebih menambahkan atribut yang lebih banyak lagi dalam datanya, agar data yang diolah dapat menjadi lebih beragam.

REFERENSI

- [1] Astriyani, E., Sari, M. M., & Herman, H. (2020). Perancangan Sistem Informasi Pembayaran Spp Berbasis Web Menggunakan Notifikasi SMS Gateway (Studi Kasus: SMP Puspita Tangerang). *Journal CERITA*, 6(1), 106–116. <https://doi.org/10.33050/cerita.v6i1.893>.
- [2] Kurniawan, H., Defit, S., & Sumijan. (2020). Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal. *Journal of Applied Computer Science and Technology*, 1(2), 80–89. <https://doi.org/10.52158/jacost.v1i2.102>.
- [3] Uska, M., Wirasasmita, R., Usuluddin, & Arianti, B. (2020). Evaluation of Rapidminer-Aplication in Data Mining Learning using PeRSIVA Model. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 164–171. <https://doi.org/10.29408/edumatic.v4i2.2688>.
- [4] Suprawoto, T. (2016). Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran. *JIKO (Jurnal Informatika Dan Komputer)*, 1(1), 12–18. <https://doi.org/10.26798/jiko.2016.v1i1.9>.
- [5] Sonang, S., Purba, A. T., & Pardede, F. O. I. (2019). Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia Dengan Metode K-Means. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 2(2), 166–173. <https://doi.org/10.37600/tekinkom.v2i2.115>.
- [6] Darmi, Y. D., & Setiawan, A. (2017). Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk. *Jurnal Media Infotama*, 12(2), 148–157. <https://doi.org/10.37676/jmi.v12i2.418>.
- [7] Widiarina, & Wahono, R. S. (2019). Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial. *Journal of Intelligent Systems*, 1(1), 33–35.
- [8] Sugianto, C. A., & Bokings, T. P. O. R. (2021). K-Means Algorithm For Clustering Poverty Data in Bangka Belitung Island Province. *Journal of Computer Networks, Architecture, and High-Performance Computing*, 3(1), 58–67. <https://doi.org/10.47709/cnahpc.v3i1.934>.
- [9] Widayani, W., & Harliana, H. (2020). Perbandingan Algoritma K-Means dan SFCM Pada Pengelompokan Rumah Tangga Miskin. *Jurnal Sains Dan Informatika*, 6(1), 1–9. <https://doi.org/10.34128/jsi.v6i1.200>.
- [10] Defiyanti, S., Jajuli, M., & Rohmawati, N. (2017). K-Medoid Algorithm in Clustering Student Scholarship Applicants. *Scientific Journal of Informatics*, 4(1), 27–33. <https://doi.org/10.15294/sji.v4i1.8212>.
- [11] Rizmayanti, A. I., Hidayati, N., Nugraha, F. S., & Gata, W. (2021). Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decision Tree (Studi Kasus Smk Multicom Depok). *Swabumi*, 9(1), 9–18. <https://doi.org/10.31294/swabumi.v9i1.8363>.
- [12] Ramadhan, A., Efendi, Z., & Mustakim. (2017). Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI)*, hal. 219-226.