

IMPLEMENTASI METODE K-MEANS CLUSTERING TUNGGAKAN UANG KULIAH PADA STMIK ROYAL KISARAN

Risnawati¹, Nurul Rahmadani²

STMIK Royal, Kisaran

e-mail: rhisnawati716@gmail.com

Abstract: Tuition fees are tuition fees that need to be paid by students each semester independently, more precisely called the SPP (Educational Development Contribution). STMIK Royal is a private university that operates in the field of computers, which has approximately 2535 active students, who have tuition arrears of Rp. 399,720,500 for semester 1 of the information systems study program. This makes it difficult for the administration to make reports on arrears of tuition fees based on class and semester. In this study, the method used was K-Means Clustering taken from 17 semester 1 classes by calculating tuition arrears from term 1, term 2, term 3, and term 4 to group the data with the provisions of 3 clusters, namely the highest arrears, medium arrears, and lowest arrears. The results of the research and grouping the data using the K-Means Clustering method of the 3 clusters determined that there were 3 classes with the highest arrears, 9 classes of medium arrears, and 2 classes of the lowest arrears.

Keywords: Tuition arrears, STMIK Royal, K-Means Clustering, Cluster

Abstrak: Uang kuliah adalah biaya kuliah yang perlu dibayarkan oleh mahasiswa setiap semesternya secara mandiri, lebih tepatnya di sebut dengan SPP (Sumbangan Pembinaan Pendidikan). STMIK Royal adalah perguruan tinggi swasta yang bergerak dibidang komputer yang memiliki jumlah mahasiswa lebih kurang 2535 mahasiswa yang aktif, yang memiliki tunggakan kuliah sebesar Rp. 399.720.500 untuk semester 1 program studi sistem informasi. Hal tersebut mengakibatkan bagian administrasi kesulitan untuk membuat laporan tunggakan uang kuliah berdasarkan kelas dan semesternya. Pada penelitian ini metode yang digunakan adalah K-Means Clustering diambil dari 17 kelas semester 1 dengan menghitung tunggakan uang kuliah dari termin 1, termin 2, termin 3, dan termin 4 untuk dilakukan pengelompokan data dengan ketentuan 3 cluster yaitu tunggakan tertinggi, tunggakan sedang, dan tunggakan terendah. Hasil dari penelitian dan pengelompokan data dengan metode K-Means Clustering dari 3 cluster yang ditentukan terdapat 3 kelas dengan tunggakan tertinggi, 9 kelas tunggakan sedang, dan 2 kelas tunggakan terendah.

Kata kunci: Tunggakan Uang Kuliah, STMIK Royal, K-Means Clustering, Klaster

PENDAHULUAN

Uang kuliah adalah biaya kuliah yang perlu dibayarkan oleh mahasiswa setiap semesternya secara mandiri, lebih tepatnya di sebut dengan SPP (Sumbangan Pembinaan Pendidikan). Terhambatnya pembayaran uang kuliah sering dialami oleh perguruan tinggi swasta dimanapun berada. Berbeda dengan Perguruan Tinggi Negeri (PTN) yang banyak mendapat subsidi maupun bantuan dari pemerintah, dikelola oleh

Kementerian Riset Teknologi, dan Pendidikan Tinggi Republik Indonesia, sedangkan perguruan tinggi swasta adalah perguruan tinggi yang dikelola oleh yayasan atau lembaga non pemerintah. Semua perguruan tinggi swasta di Indonesia dikoordinasikan oleh badan yang disebut Lembaga Layanan Pendidikan Tinggi (LLDIKTI) (Purba et al., 2019).

STMIK Royal Kisaran adalah perguruan tinggi swasta satu-satunya yang terakreditasi B dari Institusi dan

Program Studi di Kabupaten Asahan Sumatera Utara yang bergerak di bidang komputer dengan jenjang prodi S-1 dan D-3 yang memiliki mahasiswa lebih kurang 2.535 mahasiswa aktif. Sebagai mahasiswa yang aktif, maka mahasiswa harus membayar uang kuliah sebagai tanda kewajiban melaksanakan perkuliahan. namun dari banyaknya mahasiswa yang aktif juga tidak menutup kemungkinan bahwa tunggakan uang kuliah mahasiswa pada STMIK Royal Kisaran mencapai Rp.399.720.500 terhitung di akhir semester ganjil sampai termin ke 4, hal tersebut akan mengakibatkan kegiatan operasional menjadi terhambat, terlebih pada bagian administrasi keuangan yang merasa kesulitan untuk menentukan kelas mana saja yang penunggakan uang kuliahnya paling banyak untuk menjadi data laporan uang kuliah.

Penunggakan uang kuliah pada STMIK Royal Kisaran bisa terjadi karena kesulitan ekonomi dimasa pandemi, keteledoran pembayaran uang kuliah, ketidaktahuan metode pembayaran, dan lain sebagainya sehingga menyulitkan pihak administrasi keuangan untuk membuat laporan tunggakan uang kuliah berdasarkan kelas. Salah satu cara untuk mengatasi hal tersebut adalah dilakukannya pengelompokan kelas untuk mengetahui secara tepat kelas mana saja yang paling banyak mengalami penunggakan dengan ketentuan cluster tunggakan tertinggi, tunggakan sedang, dan tunggakan terendah. Cluster juga merupakan proses pengelompokan satu set objek ke dalam dengan karakteristik yang serupa.(Ananda, 2018). Cluster juga bagian dari data mining dengan arti lain data mining adalah proses menemukan pola yang menarik dari sejumlah besar data.(Metisen & Sari, 2015) Penelitian ini menggunakan 17 sampel kelas prodi Sistem Informasi semester 1. Data yang digunakan dalam perhitungan algoritma k-means adalah jumlah kelas yang memiliki penunggakan, termin per semester, dan jumlah tunggakan. Salah satu cara yaitu dengan memanfaatkan

teknologi data mining untuk pengelompokan kelas dengan teknik Clustering K-Means.

METODE

Data Mining (DM) adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar pengambilan keputusan (Metisen & Sari, 2015). Data Mining adalah Analisa terhadap data untuk menentukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut(Risnawati & Rohminatin, 2020). Data mining dari arti lain adalah suatu proses yang menemukan hubungan berarti seperti pola dan kecenderungan dalam memeriksa sekumpulan besar data yang tersimpan yang menggunakan teknik pengenalan pola, statistik, dan matematika (Rohman & Rochcham, 2019). Sedangkan dari arti lain data mining adalah proses menemukan pola yang menarik dari sejumlah besar data (Wijayanti et al., 2017).

K-Means merupakan salah satu metode pengelompokan data non-hierarki (sekatan) yang berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data kedalam kelompok sehingga data berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain.adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok (Rahmadani & Kurniawan, 2020). *Algoritma K-Means* merupakan salah satu algoritma dalam fungsi *clustering* atau pengelompokan. *Clustering* mengacu pada pengelompokan data, observasi atau kasus berdasar

kemiripan objek yang diteliti (Risnawati & Rohminatn, 2020).

Clustering merupakan bagian dari ilmu *Data Mining* yang bersifat tanpa arahan (*unsupervised*). *Clustering* adalah proses pembagian data ke dalam kelas atau cluster berdasarkan tingkat kesamaannya. Dalam *clustering*, data yang memiliki kesamaan dimasukkan ke dalam *cluster* yang sama, sedangkan yang tidak memiliki kesamaan dimasukkan dalam *cluster* yang berbeda. *Clustering* akan melakukan pengelompokan data-data ke dalam sejumlah kelompok *cluster* berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada (Ananda, 2018).

K-Means Clustering merupakan proses pengelompokan elemen-elemen dataset ke dalam cluster k. Proses dari pengelompokan dilakukan dengan menghitung jarak euclidean antara elemen dan centroid dimana centroid adalah elemen kerapatan seragam (Muningsih & Kiswati, 2015). Secara umum, pengelompokan data clustering menggunakan metode *K-Means* dilakukan dengan algoritma dasar sebagai berikut:

1. Menentukan jumlah *cluster* yang akan dibentuk
2. Alokasikan data ke dalam cluster secara random, untuk pemilihan bilangan random pusat awal cluster harus dengan urutan data.
3. Menghitung pusat kelompok dari masing-masing data kelompok. Dari masing-masing centroid akan diambil rata-rata nilainya. Jika rata-rata dinyatakan sebagai sebuah kelompok, maka *i* adalah fitur, *p* adalah dimensi data yang persamaannya adalah untuk menghitung centroid fitur *i* yang digunakan persamaan 1.

$$C_i = \frac{1}{M} \sum_{j=1}^M X_j \dots\dots\dots (1)$$

Persamaan 1 dilakukan sebanyak *p* dimensi *i=1* sampai dengan *i=p* (Rahmadani & Kurniawan, 2020)

4. Alokasikan masing-masing data ke centroid yang terdekat dengan mengukur jarak ruang yang dapat di cari persamaannya.

$$d = \sqrt{(x_1 + x_2)^2 + (y_1 + xy_2)^2} \dots\dots\dots (2)$$

Alokasikan data ke masing-masing metode *K-means* berdasarkan perbandingan jarak antara centroid ke setiap kelompok.

$$a_{i1} \begin{cases} 1 & d = \min\{D(x_1, C_1)\} \\ 0 & \text{lainnya} \end{cases} \dots\dots\dots (3)$$

a_{i1} adalah nilai anggota titik *x_i* ke pusat kelompok *c₁*, *d* adalah jarak terpendek dari data *x_i* ke kelompok setelah dilakukan perbandingan, dan *c₁* adalah pusat kelompok ke 1.

$$j = \sum_{i=1}^n \sum_{i=1}^k a_{ic} D(X_i, C_1) \dots\dots\dots (4)$$

n adalah jumlah data, *k* adalah jumlah kelompok, *a_{i1}* adalah nilai anggota titik data *x_i* ke kelompok *c_i* yang diikuti. *a* mempunyai nilai 0 atau 1, data merupakan anggota kelompok.

5. Lakukan perhitungan kembali jika masih terjadi perubahan data sampai selesai.

Karakteristik dari algoritma *K-Means* adalah penentuan titik pusat awal *cluster* karena *K-Means* dapat membangkitkan titik pusat cluster awal secara random.

HASIL DAN PEMBAHASAN

Proses yang dilakukan untuk membantu administrasi keuangan dalam pembuatan laporan tunggakan uang kuliah pada STMIK Royal Kisaran menggunakan metode *K-Means Clustering* dengan ketentuan 3 cluster yaitu Tunggakan tertinggi, tunggakan sedang, dan tunggakan terendah. Cluster yang dilakukan berdasarkan jumlah kelas yang memiliki tunggakan uang kuliah

sehingga menyulitkan bagian administrasi keuangan untuk membuat laporan data keuangan.

Adapun data yang diolah ada sebanyak 17 kelas yang memiliki tunggakan uang kuliah antara lain:

Tabel Daftar Tunggakan Uang Kuliah

NO	KELAS	JUMLAH TUNGGAKAN (Rp)			
		Termin 1	Termin 2	Termin 3	Termin 4
1	SI-IA	900000	2250000	10757000	21107000
2	SI-IB	450000	900000	13000000	26912500
3	SI-IC	450000	900000	11950000	24100000
4	SI-ID	0	1000	11576000	24176000
5	SI-IE	450000	900000	9025000	19475000
6	SI-IF	450000	900000	10500000	21300000
7	SI-IG	450000	900000	14500000	28900000
8	SI-IH	0	0	12525000	25575000
9	SI-II	450000	900000	9200000	20450000
10	SI-IJ	0	0	8800000	21362500
11	SI-IK	0	450000	12325000	26275000
12	SI-IL	0	0	5775000	14775000
13	SI-IM	0	450000	5575000	10975000
14	SI-IN	0	450000	10725000	22387500
15	SI-IO	500000	1500000	17425000	33425000
16	SI-IP	0	0	12600000	27500000
17	SI-IQ	750000	1750000	16025000	32025000

Tabel di atas adalah data yang akan dihitung menggunakan K-Means Clustering dengan jumlah cluster 4, jumlah data 17, dan jumlah atribut 3. Perhitungan manual di ambil dari jumlah tunggakan per termin sebanyak 4 termin yaitu termin 1, termin 2, termin 3, dan termin 4.

Penentuan pusat cluster awal diambil dari data 1 sebagai pusat cluster 1 yaitu Rp.500.000, Rp.1.500.000, Rp.17.425.000, Rp.33.425.000

Data 2 sebagai pusat cluster 2 yaitu Rp.900.000, Rp. 2.250.000, Rp. 10.757.000, Rp. 21.107.000.

Data 3 sebagai pusat 3 yaitu Rp. 0, Rp. 450.000, Rp. 5.575.000, Rp. 10.975.000.

Adapun perhitungan jarak data 1 diambil dari masing-masing cluster:

$$d = \sqrt{(500.000 - 500.000)^2 + (1.500.000 - 1.500.000)^2 + (17.425.000 - 17.425.000)^2 + (33.425.000 - 33.425.000)^2} = 14.032.742$$

$$d = \sqrt{(900.000 - 900.000)^2 + (2.250.000 - 2.250.000)^2 + (10.757.000 -$$

$$10.757.000)^2 + (21.107.000 - 21.107.000)^2} = 0,00$$

$$d = \sqrt{(0 - 0)^2 + (450.000 -$$

$$450.000)^2 + (5.575.000 -$$

$$5.575.000)^2 + (10.975.000 - 10.975.000)^2}$$

$$= 11.556.840$$

untuk melihat hasil persamaan dan perhitungan pusat cluster dari 17 data dapat dilihat pada tabel di bawah ini:

Tabel Hasil Perhitungan Jarak Pusat Cluster menggunakan 17 Data

C1	C2	C3	Jarak Terpendek
14.032.74		11.556.84	
2	0	0	0
		17.593.73	
7.896.568	6.384.346	6	6.384.346
10.830.22		14.605.17	
4	3.522.258	9	3.522.258
11.056.75		14.507.92	
4	3.994.687	9	3.994.687
17.158.81			
7	3.457.202	8.279.946	3.457.202
13.976.18		11.457.14	
5	1.458.869	8	1.458.869
		20.034.12	
5.421.600	8.761.615	7	5.421.600
		16.176.06	
9.387.891	5.381.575	3	5.381.575
15.374.12		10.164.70	
6	2.209.275	6	2.209.275
14.912.89		10.885.92	
8	3.125.321	1	3.125.321
		16.722.81	
8.859.176	5.763.406	4	5.763.406
22.046.42			
8	8.413.504	3.831.775	3.831.775
25.412.15	11.556.84		
3	0	0	0
12.964.13		12.520.68	
9	2.385.520	9	2.385.520
	14.032.74	25.412.15	
0	2	3	0
		17.961.86	
7.802.964	7.080.932	9	7.080.932
	12.133.71	23.549.04	
2.011.219	5	5	2.011.219

Setelah dilakukannya perhitungan jarak pusat cluster untuk mencari nilai terdekat, maka akan dilakukan pengelompokan grup yang akan di beri tanda 1 yang berarti bahwa data tersebut berada dalam grup kelompok data.

Tabel Pengelompokan Data 1

No.	C1	C2	C3
1		1	
2		1	
3		1	
4		1	
5		1	
6		1	
7	1		
8		1	
9		1	
10		1	
11		1	
12			1
13			1
14		1	
15	1		
16		1	
17	1		

Setelah hasil pengelompokan data 1 berhasil, maka masuk untuk melakukan perhitungan berdasarkan anggota setiap cluster sesuai dengan perhitungan sebagai berikut:

Cluster baru pertama:

$$\frac{450.000+500.000+75.0000}{3} = 566.667$$

$$\frac{900.000+1.500.000+1.750.000}{3} = 1.383.333$$

$$\frac{14.500.000 + 17.425.000 + 16.025.000}{3}$$

$$= 15.983.333$$

$$\frac{28.900.000 + 33.425.000 + 32.025.000}{3}$$

$$= 31.450.000$$

Cluster baru kedua:

$$\frac{900.000 + 450.000 + 450.000 + 450.000 + 0 + 450.000 + 0 + 0 + 0}{9}$$

$$= 262.500$$

$$\frac{2.250.000 + 900.000 + 900.000 + 900.000 + 0 + 900.000 + 0 + 450.000 + 0}{9}$$

$$= 700.000$$

$$\frac{10.757.000 + 11.950.000 + 9.025.000 + 10.500.000 + 12.525.000 + 9.200.000 + 8.800.000 + 10.725.000 + 12.600.000}{9}$$

$$= 11.081.917$$

$$\frac{21.107.000 + 24.100.000 + 18.475.000 + 21.300.000 + 25.575.000 + 20.450.000 + 21.362.500 + 22.387.500 + 27.500.000}{9}$$

$$= 23.301.708$$

Cluster baru ketiga:

$$\frac{0 + 0}{2} = 0$$

$$2$$

$$\frac{0+450.000}{2} = 225.000$$

$$\frac{5.775.000+5.575.000}{2} = 5.675.000$$

$$\frac{14.775.000+10.975.000}{2} = 12.875.000$$

Adapun penentuan pusat awal cluster baru sebagai berikut:

Cluster baru ke 1:

566.667, 1.383.333, 15.983.33, 31.450.000

Cluster baru ke 2:

262.500, 0, 11.081.917, 23.301.708

Cluster baru ke 3:

0, 225.000, 5.675.000, 12.875.000

Lakukan perhitungan berulang sampai kelompok data memiliki nilai yang sama dengan nilai kelompok data sebelumnya dari hasil clustering. Dari hasil perhitungan iterasi yang di dapat berhenti pada iterasi ke 3, terlihat pada tabel di bawah ini:

Tabel Hasil perhitungan Clustering Iterasi ke 3

NO	KELAS	JUMLAH TUNGGAKAN (Rp)				C1	C2	C3	Jarak Terpendek
		Termin1	Termin2	Termin3	Termin4				
1	SI-1A	800000	2250000	10757000	21107000	11625593,81	2815778,625	8924876,473	2815778,625
2	SI-1B	450000	800000	13000000	26912500	5453109,778	4101326,479	15854499,54	4101326,479
3	SI-1C	450000	900000	11850000	24100000	8398660,808	1222643,454	12885432,67	1222643,454
4	SI-1D	0	1000	11576000	24176000	8633254,464	1217640,69	12750865,77	1217640,69
5	SI-1E	450000	800000	9025000	18475000	14731471,41	5256618,52	6575760,412	5256618,52
6	SI-1F	450000	800000	16500000	21300000	11547149,72	2108978,533	8742657,492	2108978,533
7	SI-1G	450000	800000	14500000	28900000	2091655,06	6587205,038	18312273,89	2091655,06
8	SI-1H	0	0	12525000	25575000	6979284,825	2779525,727	14431324,44	2779525,727
9	SI-1I	450000	800000	9200000	20450000	12892895,99	3451890,079	8394506,106	3451890,079
10	SI-1J	0	0	8800000	21862500	12473678,54	3072961,339	8947314,864	3072961,339
11	SI-1K	0	450000	12325000	26750000	6490882,521	3238798,121	14961053,61	3238798,121
12	SI-1L	0	0	5775000	14775000	19608682,18	10066952,78	1915887,523	1915887,523
13	SI-1M	0	450000	5575000	10975000	22994591,76	13504738,55	1915887,523	1915887,523
14	SI-1N	0	450000	10725000	22387500	16544286,78	1035082,833	10772222,67	1035082,833
15	SI-1O	500000	1500000	17425000	33425000	2440894,308	11979812,57	23711613,72	2440894,308
16	SI-1P	0	0	12600000	27500000	5411484,701	4517261,193	16183228,2	4517261,193
17	SI-1Q	750000	1750000	16025000	32025000	707401,3477	10089753,34	21834219,13	707401,3477

Setelah dilakukannya perhitungan berulang dan berhenti pada iterasi ke-4, maka dapat dilihat tabel di bawah ini:

Tabel Pengelompokan Data Iterasi ke 3

No.	C1	C2	C3
1		1	
2		1	
3		1	
4		1	
5		1	
6		1	
7	1		
8		1	
9		1	
10		1	
11		1	
12			1
13			1
14		1	
15	1		
16		1	
17	1		

Dari tabel di atas, maka dari pengelompokan data tunggakan uang kuliah yang sudah dihitung dari 4 termin terdapat 3 kelas tunggakan tertinggi, 9 kelas tunggakan sedang, dan 2 kelas tunggakan terendah. Dapat dilihat pada tabel 6 mengenai kelas yang sudah di cluster.

Tabel Kelas yang sudah di Cluster

C1	C2	C3
SI-1O	SI-1A	SI-1L
SI-1Q	SI-1B	SI-1M
SI 1G	SI-1C	
	SI-1D	
	SI-1E	
	SI-1F	
	SI-1H	
	SI-1I	
	SI-1J	
	SI-1K	
	SI-1N	
	SI-1P	

SIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang sudah dilakukan, maka dapat disimpulkan sebagai berikut:

1. Metode K-Means Clustering yang diterapkan oleh data mining untuk menentukan tunggakan uang kuliah di STMIK Royal Kisaran berhasil diterapkan sehingga mempermudah bagian keuangan untuk membuat laporan keuangan yang menunggak
2. Pengelompokan data yang sudah dilakukan terdapat 3 jenis cluster, yaitu tunggakan tertinggi, tunggakan sedang, dan tunggakan terendah. Dari 3 cluster yang ditentukan terdapat 3 kelas tunggakan tertinggi, 9 kelas tunggakan sedang, dan 2 kelas tunggakan terendah.

DAFTAR PUSTAKA

- Ananda, L. R. (2018). Clustering Untuk Menentukan Calon Mahasiswa Berprestasi. *Jiti*, 1(2), 16–19.
- Metisen, B. M., & Sari, H. L. (2015). Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2), 110–118.
- Muningsih, E., & Kiswati, S. (2015). Penerapan Metode K-Means untuk Clustering Produk Online Shop dalam Penentuan Stok Barang. *Jurnal Bianglala Informatika*, 3(1), 10–17.
- Purba, L. P., Hartama, D., Irawan, E., & Wanto, A. (2019). Memprediksi Faktor Tunggakan Uang Kuliah Menggunakan Metode Naive Bayes. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 298. <https://doi.org/10.30645/senaris.v1i0.35>
- Rahmadani, N., & Kurniawan, E. (2020). Implementasi Metode K-Means Clustering Tunggakan Rekening Listrik pada PT . PLN (Persero) Gardu Induk Kisaran. *J-SISKO TECH*, 3(1), 103–117.
- Risnawati, R., & Rohminatin, R. (2020). K-MEANS CLUSTERING HWI PRODUCTS (Case Study: HWI

- Kisaran Distributor). *International Conference on ...*, 4509, 27–36. <https://jurnal.stmikroyal.ac.id/index.php/ICoSSIT/article/view/824>
- Rohman, A., & Rochcham, M. (2019). Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 5(1), 23–29. <https://doi.org/10.37760/neoteknika.v5i1.1379>
- Wijayanti, S., Azahari, & Andrea, R. (2017). K-Means cluster analysis for students graduation (case study: STMIK widya cipta dharma). *ACM International Conference Proceeding Series, Part F1296*, 20–23. <https://doi.org/10.1145/3108421.3108430>