

Analisis Pengelompokan Pola Pembayaran UKT Mahasiswa Menggunakan Algoritma K-Means Clustering

Desmulyati^{1*}, Imam Budiawan², Feri Andrianto³, Reafael Andrian Canavaro⁴, Muhammad Haikal Nugroho⁵, Sofiyon Aris Saputra⁶

¹Informatika, ²⁻⁶Sistem Informasi, Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat
E-mail: desmulyati.dmy@bsi.ac.id

* Corresponding Author

 <https://doi.org/10.31004/jerkin.v4i3.4967>

ARTICLE INFO

Article history

Received: 23 Nov 2025

Revised: 05 Dec 2025

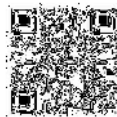
Accepted: 30 Dec 2025

Kata Kunci:

UKT, K-Means Clustering, Pola Pembayaran, Machine Learning, Pendidikan Tinggi.

Keywords:

UKT, K-Means Clustering, Payment Pattern, Machine Learning, Higher Education.



ABSTRACT

Uang Kuliah Tunggal (UKT) berperan penting dalam pembiayaan pendidikan tinggi, namun keterlambatan dan tunggakan pembayaran sering sulit dianalisis secara manual. Penelitian ini bertujuan mengelompokkan pola pembayaran UKT mahasiswa menggunakan algoritma K-Means berdasarkan pendapatan per kapita, besaran UKT, keterlambatan, kategori keterlambatan, dan total tunggakan. Data yang digunakan sebanyak 300 mahasiswa yang telah dibersihkan dan distandarasi. Penentuan jumlah cluster menggunakan metode Elbow dan Silhouette Score, dengan hasil terbaik pada $k = 3$ ($SSE = 524,06$; Silhouette Score = $0,5609$). Tiga cluster tersebut meliputi mahasiswa berpendapatan tinggi dengan pembayaran tertib, mahasiswa berpendapatan rendah dengan keterlambatan ringan, dan mahasiswa berisiko tinggi dengan keterlambatan serta tunggakan besar. Hasil ini membantu perguruan tinggi memetakan risiko pembayaran UKT dan menyusun kebijakan penagihan serta keringanan yang lebih tepat sasaran.

Single Tuition Fee (UKT) plays a crucial role in financing higher education, but late and arrear payments are often difficult to analyze manually. This study aims to classify student UKT payment patterns using the K-Means algorithm based on per capita income, UKT amount, lateness, lateness category, and total arrears. The data used were 300 cleaned and standardized students. The number of clusters was determined using the Elbow and Silhouette Score methods, with the best results at $k = 3$ ($SSE = 524.06$; Silhouette Score = 0.5609). The three clusters include high-income students with regular payments, low-income students with minor delays, and high-risk students with large delays and arrears. These results help universities map UKT payment risks and develop more targeted collection and relief policies.



This is an open access article under the CC-BY-SA license.

How to Cite: Desmulyati, et al (2025). Analisis Pengelompokan Pola Pembayaran UKT Mahasiswa Menggunakan Algoritma K-Means Clustering, 4(3) 16592-16597. <https://doi.org/10.31004/jerkin.v4i3.4967>

PENDAHULUAN

UKT (Uang Kuliah Tunggal) diterapkan untuk menyederhanakan skema pembiayaan pendidikan tinggi sekaligus menjaga keberlanjutan operasional perguruan tinggi. Namun, dalam praktiknya, tidak semua mahasiswa mampu membayar UKT secara tepat waktu. Kondisi ekonomi keluarga yang beragam, besaran UKT yang berbeda, serta kurangnya pengawasan berbasis data menyebabkan munculnya beberapa permasalahan, seperti: keterlambatan pembayaran berulang, tunggakan yang menumpuk, dan sulitnya memetakan mahasiswa mana yang berisiko tinggi menunggak..

Selama ini, pemantauan pembayaran UKT umumnya hanya dilakukan dengan melihat status “lunas” atau “belum lunas” pada sistem keuangan, tanpa analisis lebih dalam mengenai pola keterlambatan, besaran tunggakan, dan keterkaitannya dengan kondisi ekonomi mahasiswa. Padahal, dengan pemetaan pola pembayaran yang lebih detail, perguruan tinggi dapat:

1. Mengidentifikasi kelompok mahasiswa yang selalu membayar tepat waktu,
2. Mengidentifikasi mahasiswa yang mulai menunjukkan gejala keterlambatan, dan
3. Mengantisipasi mahasiswa yang berpotensi mengalami tunggakan jangka panjang.

Perkembangan machine learning, khususnya unsupervised learning, memberikan peluang untuk melakukan analisis pola pembayaran secara lebih objektif. Salah satu algoritma yang banyak digunakan adalah K-Means Clustering, yang mampu mengelompokkan objek berdasarkan kemiripan karakteristiknya [1]–[5]. Dalam konteks UKT, K-Means dapat digunakan untuk membentuk cluster mahasiswa berdasarkan kombinasi indikator: pendapatan per kapita, besaran UKT, jumlah hari terlambat, kategori keterlambatan, dan total tunggakan [2]–[4],[11]–[13].

Berbagai penelitian sebelumnya telah menerapkan K-Means untuk penentuan besaran UKT, pengelompokan kelompok UKT, maupun analisis perilaku mahasiswa di bidang pendidikan [1]–[4],[6]–[10]. Namun, masih relatif sedikit kajian yang secara khusus memfokuskan pada pola keterlambatan pembayaran UKT dan total tunggakan sebagai dasar pengelompokan risiko.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis pengelompokan pola pembayaran UKT mahasiswa menggunakan algoritma K-Means Clustering, sehingga perguruan tinggi dapat memperoleh gambaran yang lebih jelas mengenai sebaran risiko pembayaran dan menyusun strategi kebijakan keuangan yang lebih tepat sasaran.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma K-Means Clustering pada data pembayaran UKT mahasiswa. Tahapan penelitian meliputi: pengumpulan data, pra-proses data, pemodelan K-Means, evaluasi cluster, serta interpretasi hasil.

Pengumpulan Data

Dataset yang digunakan terdiri dari 300 data mahasiswa dengan atribut utama sebagai berikut:

1. Pendapatan perkapita (rupiah)
2. besaran_ukt (rupiah)
3. jumlah_hari_telat (hari)
4. kategori_keterlambatan (0 = tepat waktu, 1–3 = kategori telat)
5. total_tunggakan (rupiah)

Data bersumber dari simulasi/rekap pembayaran keuangan mahasiswa yang disesuaikan dengan pola nyata di perguruan tinggi. Seluruh data disimpan dalam file DATASET_UKT_300.csv yang kemudian diolah menggunakan Python.

No	ID Mahasiswa	Pendapatan perkapita	Besaran UKT	Jumlah Hari Telat	Kategori Keterlambatan	Total Tunggakan
1	MHS001	900000	500000	0	0	0
2	MHS002	1200000	750000	1	1	0
3	MHS003	850000	500000	2	1	0
4	MHS004	2500000	2000000	0	0	0
5	MHS005	5000000	4000000	3	1	0
6	MHS006	750000	500000	0	0	0
7	MHS007	1000000	750000	4	1	0
8	MHS008	4800000	3500000	1	1	0
9	MHS009	600000	500000	0	0	0
10	MHS010	5200000	4500000	2	1	0
11	MHS011	800000	750000	0	0	0
12	MHS012	950000	750000	1	1	0
13	MHS013	3000000	2500000	0	0	0
14	MHS014	4200000	3500000	2	1	0
15	MHS015	1100000	750000	0	0	0
16	MHS016	700000	500000	0	0	0
17	MHS017	2600000	2000000	60	3	2000000

18	MHS018	2400000	2000000	55	3	1800000
19	MHS019	2800000	2500000	58	3	1900000
20	MHS020	2300000	2000000	54	3	1600000
21	MHS021	900000	750000	1	1	0
22	MHS022	5000000	4250000	3	1	0
23	MHS023	4700000	4000000	2	1	0
24	MHS024	880000	700000	0	0	0
25	MHS025	920000	700000	1	1	0
26	MHS026	5100000	4200000	2	1	0
27	MHS027	4600000	4000000	3	1	0
28	MHS028	910000	700000	0	0	0
29	MHS029	2550000	2000000	57	3	1950000
30	MHS030	2650000	2500000	59	3	2100000

Pra-Proses Data

Tahap pra-proses dilakukan agar data siap di-cluster oleh K-Means:

1. Pembersihan Data (Cleaning)
 - a. Menghapus baris dengan nilai kosong (missing value).
 - b. Memastikan seluruh kolom numerik (pendapatan, UKT, hari telat, tunggakan) bertipe numerik.
2. Pemilihan Fitur
Fitur yang digunakan untuk proses clustering:
 - a. Pendapatan perkapita
 - b. `besaran_ukt`
 - c. `jumlah_hari_telat`
 - d. `kategori_keterlambatan`
 - e. `total_tunggakan`
3. Standarisasi Data
Karena skala antar fitur berbeda (misalnya rupiah vs hari), digunakan StandardScaler sehingga setiap fitur memiliki rata-rata 0 dan standar deviasi 1. Standarisasi ini penting agar K-Means tidak didominasi oleh fitur berskala besar [11]–[13],[15].

Pemodelan K-Means Clustering

Algoritma yang digunakan adalah K-Means Clustering, yaitu metode partitioning clustering yang membagi data ke dalam k cluster dengan meminimalkan jumlah kuadrat jarak setiap titik terhadap centroid cluster [1],[11],[12].

Tahapan pemodelan:

1. Menentukan rentang nilai k (jumlah cluster) dari $k = 2$ sampai $k = 10$.
2. Untuk setiap k dilakukan:
 - a. Inisialisasi centroid,
 - b. Proses iteratif pembaruan centroid,
 - c. Perhitungan Within-Cluster Sum of Squares (SSE),
 - d. Perhitungan Silhouette Score untuk mengukur kualitas pemisahan cluster.
3. Memilih k yang memberikan keseimbangan antara SSE yang relatif kecil dan Silhouette Score yang cukup tinggi.

Evaluasi Cluster

Cluster dievaluasi menggunakan dua ukuran utama:

1. SSE (Sum of Squared Errors / Inertia)
Mengukur seberapa kompak data dalam setiap cluster. Semakin kecil SSE, semakin kompak cluster yang terbentuk.
2. Silhouette Score
Mengukur seberapa baik setiap titik data berada dalam cluster yang “benar” dibandingkan dengan cluster terdekat lainnya, dengan nilai antara -1 s.d. 1. Nilai mendekati 1 menunjukkan pemisahan cluster yang baik.

Alat dan Perangkat Lunak

Penelitian ini menggunakan:

1. Python 3 (Google Colab)
2. Library: pandas, numpy, scikit-learn, dan matplotlib
3. File input: DATASET_UKT_300.csv

HASIL DAN PEMBAHASAN

Hasil Penentuan Jumlah Cluster

Berdasarkan perhitungan dengan rentang $k = 2$ sampai $k = 10$, diperoleh nilai SSE dan Silhouette Score untuk setiap k . Pola grafik Elbow menunjukkan penurunan SSE yang cukup tajam hingga $k = 3$, kemudian cenderung melandai. Sementara itu, Silhouette Score mencapai nilai yang relatif tinggi di sekitar $k = 3$.

Dari hasil perhitungan melalui kode Python, diperoleh:

1. Jumlah data yang di-cluster : 300 mahasiswa
2. k terpilih : 3
3. SSE untuk $k = 3$: 524,06
4. Silhouette Score untuk $k = 3$: 0,5609

Nilai Silhouette Score = 0,5609 menunjukkan bahwa pemisahan cluster tergolong cukup baik untuk data sosial ekonomi dan perilaku pembayaran yang cenderung kompleks [8],[9],[14].

Profil Cluster

Setelah model K-Means dengan $k = 3$ diterapkan, setiap mahasiswa memperoleh label cluster (0, 1, atau 2). Rata-rata fitur untuk setiap cluster dirangkum sebagai berikut:

1. Cluster 0 (101 mahasiswa):
 - a. Rata-rata pendapatan per kapita \approx Rp 4.745.049,50
 - b. Rata-rata besaran UKT \approx Rp 4.257.425,74
 - c. Rata-rata jumlah hari telat \approx 2,55 hari
 - d. Rata-rata kategori keterlambatan \approx 0,38
 - e. Rata-rata total tunggakan \approx Rp 0
2. Cluster 1 (174 mahasiswa):
 - a. Rata-rata pendapatan per kapita \approx Rp 882.494,25
 - b. Rata-rata besaran UKT \approx Rp 705.459,77
 - c. Rata-rata jumlah hari telat \approx 1,09 hari
 - d. Rata-rata kategori keterlambatan \approx 0,21
 - e. Rata-rata total tunggakan \approx Rp 0
3. Cluster 2 (25 mahasiswa):
 - a. Rata-rata pendapatan per kapita \approx Rp 2.608.120,00
 - b. Rata-rata besaran UKT \approx Rp 2.440.000,00
 - c. Rata-rata jumlah hari telat \approx 56,80 hari
 - d. Rata-rata kategori keterlambatan \approx 3,00
 - e. Rata-rata total tunggakan \approx Rp 1.970.000,00

Interpretasi Cluster

Berdasarkan profil di atas, cluster dapat diinterpretasikan sebagai berikut:

1. Cluster 1 – Kelompok pendapatan rendah, namun relatif tertib (low risk)
 - a. Pendapatan per kapita paling rendah di antara ketiga cluster.
 - b. Besaran UKT juga paling rendah.
 - c. Keterlambatan rata-rata hanya sekitar 1 hari, tanpa tunggakan.
 - d. Dapat diinterpretasikan sebagai mahasiswa dari keluarga ekonomi lemah hingga menengah, namun masih mampu menjaga ketertiban pembayaran.
2. Cluster 0 – Kelompok pendapatan tinggi, UKT besar, tetapi tetap relatif tertib (medium risk)
 - a. Pendapatan per kapita dan besaran UKT paling tinggi.
 - b. Keterlambatan rata-rata sedikit lebih tinggi (2–3 hari), namun tanpa tunggakan.
 - c. Dapat diinterpretasikan sebagai mahasiswa dari keluarga yang relatif mampu, dengan potensi keterlambatan administratif, namun risiko tunggakan tetap rendah.

3. Cluster 2 – Kelompok keterlambatan ekstrem dan tunggakan tinggi (high risk)
 - a. Pendapatan per kapita kategori menengah dan besaran UKT sedang.
 - b. Keterlambatan rata-rata sangat tinggi (lebih dari 50 hari).
 - c. Total tunggakan rata-rata hampir Rp 2.000.000 per mahasiswa.
 - d. Cluster ini dapat diidentifikasi sebagai kelompok mahasiswa berisiko tinggi yang membutuhkan perhatian khusus (misalnya pendampingan atau skema keringanan dan cicilan UKT).

Implikasi bagi Pengelolaan Keuangan UKT

Hasil pengelompokan menunjukkan bahwa:

1. Mayoritas mahasiswa berada pada cluster yang relatif tertib (Cluster 0 dan 1), sehingga risiko sistemik terhadap keuangan kampus relatif terkendali.
2. Sejumlah kecil mahasiswa terkonsentrasi di Cluster 2 dengan pola keterlambatan ekstrem dan tunggakan tinggi. Justru kelompok kecil inilah yang berpotensi memberikan dampak besar terhadap piutang UKT jika tidak dikelola dengan baik.

Dengan memanfaatkan hasil clustering, kampus dapat:

1. Menyusun strategi penagihan diferensial, misalnya:
 - a. Peningkat otomatis dan edukasi keuangan untuk cluster tertib,
 - b. Pendekatan personal, konseling, atau skema cicilan khusus untuk cluster berisiko tinggi.
2. Merancang kebijakan keringanan atau beasiswa internal yang lebih tepat sasaran dengan mempertimbangkan kombinasi pendapatan per kapita dan pola keterlambatan.
3. Menggunakan cluster sebagai dasar analitik lanjutan, misalnya menggabungkan dengan algoritma lain (PSO, SVM, dsb.) seperti yang disarankan pada penelitian [6],[9],[14],[24].

SIMPULAN

Penelitian ini menganalisis pola pembayaran UKT mahasiswa menggunakan algoritma K-Means Clustering dengan memanfaatkan 300 data mahasiswa dan lima atribut utama: pendapatan per kapita, besaran UKT, jumlah hari keterlambatan, kategori keterlambatan, dan total tunggakan.

Berdasarkan hasil evaluasi menggunakan metode Elbow dan Silhouette, diperoleh bahwa:

1. Jumlah cluster terbaik adalah $k = 3$ dengan
 - a. $SSE = 524,06$, dan
 - b. Silhouette Score = $0,5609$.

Tiga cluster yang terbentuk dapat disimpulkan sebagai:

1. Cluster 1: Mahasiswa berpendapatan relatif rendah dengan UKT kecil, tetapi cenderung membayar tepat waktu dan tanpa tunggakan.
2. Cluster 0: Mahasiswa berpendapatan tinggi dengan UKT besar, memiliki sedikit keterlambatan, namun tidak memiliki tunggakan signifikan.
3. Cluster 2: Mahasiswa dengan pendapatan menengah dan UKT sedang, namun memiliki rata-rata keterlambatan sangat tinggi dan tunggakan besar, sehingga dikategorikan sebagai cluster berisiko tinggi.

Secara keseluruhan, penelitian ini menunjukkan bahwa algoritma K-Means efektif digunakan untuk:

1. Memetakan mahasiswa berdasarkan risiko pembayaran UKT,
2. Memberikan wawasan bagi perguruan tinggi dalam merancang kebijakan penagihan dan keringanan yang lebih tepat sasaran,
3. Menjadi dasar pengembangan sistem pendukung keputusan di bidang pengelolaan keuangan mahasiswa.

Saran pengembangan ke depan:

1. Menambah variabel lain seperti status pekerjaan orang tua, bantuan sosial, dan riwayat akademik untuk memperkaya analisis cluster.
2. Mencoba algoritma clustering lain (misalnya Hierarchical, DBSCAN, atau K-Means yang dioptimasi PSO) untuk dibandingkan dengan K-Means standar [9],[14],[24].
3. Mengintegrasikan hasil clustering ke dalam dashboard web agar dapat digunakan langsung oleh bagian keuangan kampus dalam pengambilan keputusan.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada pihak yang sudah berkontribusi dalam pelaksanaan penelitian dan penyusunan artikel ini.

REFERENSI

- Yustanti, W., Nurhidayat, A. I. & Java, M. I., 2025. Optimizing Tuition Fee Determination with K-Means Cluster Relabeling Based on Centroid Mapping of Principal Component Pattern. *Journal of Information Systems Engineering and Business Intelligence*, 11(3), pp. 445–458.
- 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 (JACOST)*, 1(2), pp. 80–89.
- Sophia, A. & Jasmir, J., 2024. Penerapan Data Mining dalam Pengelompokan Uang Kuliah Tunggal (UKT) Menggunakan Metode K-Means pada Universitas Jambi. *Jurnal Manajemen Sistem Informasi*, 9(1), pp. 24–38.
- Fahriya, K. & Yustanti, W., 2021. Optimalisasi Jumlah Kluster Uang Kuliah Tunggal pada Data Sosial Ekonomi Mahasiswa. *JEISBI (Journal of Emerging Information Systems and Business Intelligence)*, 2(2), pp. 73–77.
- Rosdianti, E. & Sugianto, C. A., 2023. Analisis Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan (SPP) dengan Menggunakan Algoritma K-Means (Studi Kasus Politeknik TEDC Bandung). *Journal Informatics and Electronics Engineering*, 3(1), pp. 25–32.
- Talib, N. I. M., Abd Majid, N. A. & Sahran, S., 2023. Identification of Student Behavioral Patterns in Higher Education Using K-Means Clustering and Support Vector Machine. *Applied Sciences*, 13(5), art. 3267.
- Goh, Y. L., Goh, Y. H., Yip, C.-C., Ting, C. H., Bin, R. L. L. & Chen, K. P., 2020. Prediction of Students' Academic Performance by K-Means Clustering. *Science Proceedings Series*, 2(1), pp. 1–6.
- Liu, R., 2022. Data Analysis of Educational Evaluation Using K-Means Clustering Method. *Computational Intelligence and Neuroscience*, 2022, Article ID 3762431, 10 pages.
- Syahputra, I., Fonna, N., Fitriani, R. & Tahlil, T., 2025. Application of K-Means Algorithm to Cluster Students' Reading Patterns in the Digital Age. *Jurnal Inovtek Polbeng Seri Informatika*, 10(1), pp. 320–331.
- Siregar, H. L., Hidayanthi, R. & Sakti, A. L. D., 2024. Implementation of K-Means Clustering on Student Learning Achievements Based on Social Economic and Social Related. *Research and Development in Education (RaDEn)*, 4(2), pp. 1447–1459.