

Penerapan Data Mining dalam Pengelompokan Uang Kuliah Tunggal (UKT) Menggunakan Metode K-Means Pada Universitas Jambi

Aya Sophia¹, Jasmir²

*Pascasarjana, Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi
Jl. Jend. Sudirman Thekok-Jambi Telp: 0741-35096 Fax : 35093
E-mail: sophia.aya@gmail.com¹, ijay_jasmir@yahoo.com²*

Abstract

UKT is a single tuition fee borned by each undergraduate student at a state university in Indonesia, to be paid every semester. At Jambi University, 8 UKT groups apply to students based on the level of the student's economic condition. Determining the UKT group for new students manually is less effective because it really depends on the assessment of the assessor, the criteria for economic conditions are quite a lot, the potential for subjectivity, and the amount of data is quite large. This study grouped new students into 8 UKT groups by applying data mining using the k-Means method. The k-Means method performs clustering based on the similarity of data on student economic conditions. K-Means analysis was performed using Rapidminer and SPSS tools. The results show that there are 7 variables/attributes of economic condition parameters that can be used in k-means analysis, including: total parent/guardian income including additional income, parent/guardian employment, electricity bills, parent status, land and building tax bills, housing conditions, and the number of dependents who are still at school. The results show that between the interpretation of the k-means analysis and the real data, there was a similarity in determining the UKT group above 50% in both Rapidminer and SPSS. Thus it can be concluded that the k-Means method can be applied to support decision making in determining UKT groups for students.

Keywords: data mining, k-Means, clustering, single tuition fee, UKT

Abstrak

Uang Kuliah Tunggal (UKT) adalah biaya yang dikenakan kepada setiap mahasiswa pada perguruan tinggi negeri di Indonesia, untuk dibayarkan pada setiap semester. Pada Universitas Jambi berlaku 8 kelompok UKT yang dikenakan kepada mahasiswa berdasarkan tingkatan kondisi ekonomi mahasiswa. Penetapan kelompok UKT bagi mahasiswa baru secara manual kurang efektif karena sangat tergantung dari penilaian penilai, kriteria kondisi ekonomi yang cukup banyak, potensi subjektivitas, dan jumlah data yang cukup banyak. Penelitian ini melakukan pengelompokan mahasiswa baru ke dalam 8 kelompok UKT dengan menerapkan *data mining* menggunakan metode *k-Means*. Metode *k-Means* melakukan pengelompokan (*clustering*) berdasarkan kemiripan data kondisi ekonomi mahasiswa. Analisis *k-Means* dilakukan menggunakan *tools* Rapidminer dan SPSS. Hasil penelitian menunjukkan terdapat 7 variabel/atribut parameter kondisi ekonomi yang dapat digunakan dalam analisis *k-Means*, meliputi: penghasilan total orang tua/wali termasuk pendapatan tambahan, pekerjaan orang tua/wali, tagihan listrik, status orang tua, tagihan pajak bumi dan bangunan, kondisi rumah, dan jumlah tanggungan yang masih sekolah. Hasil penelitian menunjukkan antara hasil interpretasi analisis *k-Means* dan data riil memiliki tingkat kesamaan penentuan kelompok UKT di atas 50% baik pada Rapidminer maupun SPSS. Dengan demikian dapat disimpulkan bahwa metode *k-Means* dapat diterapkan dalam mendukung pengambilan keputusan penetapan kelompok UKT bagi mahasiswa.

Kata kunci: data mining, k-Means, clustering, uang kuliah tunggal, UKT

© 2024 Jurnal MANAJEMEN SISTEM INFORMASI.

1. Pendahuluan

Pesatnya perkembangan teknologi informasi dewasa ini telah memberikan banyak manfaat dalam berbagai bidang. Tak dapat dipungkiri perkembangan teknologi informasi juga membawa pada semakin besarnya jumlah data yang dikumpulkan dan dikelola di berbagai industri dan organisasi. Ketersediaan data dalam jumlah besar memiliki potensi untuk memberikan pengetahuan yang berharga yang dapat mendukung pengambilan keputusan. Teknologi mengenai pencarian pengetahuan dari sejumlah besar data dikenal dengan istilah "*data mining*" [1,10]. Dengan menggunakan data relevan yang diperoleh dari dalam dan luar organisasi, *data mining* membangun model untuk menemukan pola di antara atribut yang disajikan dalam kumpulan data [10]. Secara umum, *data mining* berusaha mengidentifikasi empat jenis pola utama, yaitu asosiasi, prediksi, klaster, dan hubungan berurutan.

Universitas Jambi yang merupakan salah satu perguruan tinggi negeri di Provinsi Jambi - Indonesia, setiap tahunnya melaksanakan penerimaan mahasiswa baru. Terhadap setiap mahasiswa yang baru diterima, Universitas Jambi menetapkan besaran Uang Kuliah Tunggal (UKT) untuk dibayarkan oleh mahasiswa pada setiap semester. Penetapan UKT dilakukan dengan mempertimbangkan kemampuan ekonomi mahasiswa. Pada Universitas Jambi, berlaku sebanyak 8 kelompok UKT bagi program diploma dan program sarjana, yaitu dari UKT I yang merupakan UKT terendah, hingga UKT VIII yang merupakan UKT tertinggi. Dalam proses daftar ulang mahasiswa baru melalui sistem informasi registrasi *online* Universitas Jambi, mahasiswa melakukan input data-data untuk persyaratan penentuan UKT, seperti jumlah tanggungan orang tua yang masih sekolah, pekerjaan orang tua/wali, penghasilan per bulan orang tua/wali, status orang tua, serta mengunggah dokumen pendukung seperti kartu keluarga, bukti rekening listrik, bukti pembayaran pajak bumi dan bangunan (PBB), slip gaji orang tua, foto kondisi rumah, dan lain-lain. Berdasarkan data-data tersebut, pihak universitas menilai kemampuan ekonomi setiap mahasiswa baru dan menentukan besaran UKT yang akan dikenakan mengacu pada 8 kelompok UKT. Dalam pelaksanaannya, penentuan kelompok UKT bagi mahasiswa baru sangat tergantung dari penilaian setiap penilai. Penilaian yang dilakukan secara manual dengan kriteria kondisi ekonomi mahasiswa yang cukup banyak berpotensi adanya subjektivitas. Disamping itu, jumlah data yang cukup banyak menjadi kerumitan tersendiri bagi universitas dalam melakukan proses penilaian dan penentuan kelompok UKT secara manual. Berangkat dari kondisi tersebut, maka dibutuhkan suatu mekanisme yang lebih efektif bagi Universitas Jambi dalam menentukan kelompok UKT bagi mahasiswa baru.

K-Means merupakan salah satu metode dalam *data mining* untuk melakukan *clustering* atau mengidentifikasi pola klaster. *Clustering* mempartisi kumpulan hal-hal (misalnya objek, peristiwa, disajikan dalam kumpulan data terstruktur) ke dalam beberapa segmen (pengelompokan alami) yang anggotanya memiliki karakteristik yang sama [10]. Algoritma *k-Means Clustering* (di mana *k* adalah jumlah klaster yang telah ditentukan sebelumnya) merupakan salah satu algoritma *clustering* yang paling umum digunakan dan paling banyak direferensikan [10].

Rumusan masalah dari penelitian ini adalah bagaimana mengelompokkan mahasiswa baru Universitas Jambi ke dalam 8 kelompok UKT berdasarkan kemiripan data kondisi ekonomi mahasiswa dengan menerapkan *data mining* menggunakan metode *k-Means*. Tujuan penelitian adalah (1) merancang variabel parameter kondisi ekonomi mahasiswa agar dapat diproses dengan *data mining* menggunakan metode *k-Means*, dan (2) mengelompokkan mahasiswa baru menggunakan metode *k-Means* untuk penentuan kelompok UKT bagi mahasiswa baru Universitas Jambi.

2. Tinjauan Pustaka

Data mining adalah proses yang menggunakan teknik statistik, matematika, dan kecerdasan buatan untuk mengekstraksi dan mengidentifikasi informasi yang berguna dan pengetahuan selanjutnya (atau pola) dari kumpulan data yang besar [10]. *Data mining* juga dikenal sebagai *Knowledge Discovery in Databases* (KDD). Durham [10] meringkas proses KDD terdiri dari pemilihan data, pra-pemrosesan data, transformasi data, *data mining*, dan interpretasi/evaluasi [10].

2.1 Penelitian Sebelumnya

Penelitian mengenai penentuan kelompok UKT pada Universitas Jambi sebelumnya dilakukan oleh Mauladi, dkk [4] yang menggunakan kombinasi metode Naïve Bayes Classifier dan Fuzzy-TOPSIS. Penelitian menyimpulkan bahwa kombinasi metode tersebut dapat diimplementasikan untuk menentukan kelas UKT, dan rekomendasi dapat dipertimbangkan bagi pembuat keputusan. Penelitian mengenai

penentuan kelompok UKT lainnya dilakukan oleh Sugiyono [12] pada IAIN Samarinda menggunakan metode *k-Means Clustering*. Penelitian menyimpulkan bahwa penambahan algoritma *k-Means Clustering* pada Sistem Informasi UKT mendukung Manajemen Perguruan Tinggi dalam menetapkan kelompok tarif UKT.

Terdapat penelitian yang menerapkan metode *k-Means* menggunakan Rapidminer dan SPSS. Dalam penelitian Silveira, dkk [12] yang menggunakan metode *k-Means* dalam melakukan analisis pelanggan, hasil validasi menunjukkan SPSS memberikan nilai akurasi yang lebih tinggi daripada Rapidminer, dengan nilai akurasi pada kedua *tools* di atas 80%. Sementara itu dalam penelitian Zulfadhilah, dkk [16] yang menggunakan metode *k-Means* untuk *clustering* berdasarkan jumlah pengunjung *website*, hasil *clustering* Rapidminer dan SPSS menunjukkan data klaster yang sama.

2.2 Clustering

Menurut Sharda, dkk [10], *clustering* mempartisi kumpulan hal (misalnya objek, peristiwa, disajikan dalam kumpulan data terstruktur) ke dalam segmen (atau pengelompokan alami) yang disebut klaster, yang anggotanya memiliki karakteristik yang sama. Tidak seperti pada klasifikasi, pada *clustering* label kelas tidak diketahui. Tujuan *clustering* adalah untuk membuat kelompok sehingga anggota dalam setiap kelompok memiliki kesamaan maksimum dan anggota lintas kelompok memiliki kesamaan minimum. Kebanyakan metode analisis klaster melibatkan penggunaan ukuran jarak untuk menghitung kedekatan antara pasangan titik data. Ukuran jarak yang populer di antaranya jarak Euclidian (jarak biasa antara dua titik yang akan diukur dengan penggaris) dan jarak Manhattan (disebut juga jarak *rectilinear*, atau jarak *taxicab*, di antara dua titik). [10]

2.2 K-Means

Menurut Sharda, dkk [10], algoritma *k-Means* (dimana *k* adalah jumlah klaster yang telah ditentukan) adalah algoritma *clustering* yang paling direferensikan. Pada algoritma *k-Means*, objek dikelompokkan ke dalam sejumlah klaster berdasarkan atribut atau fitur [2]. Pengelompokan objek dilakukan dengan meminimalkan jumlah kuadrat jarak, yaitu jarak Euclidean antara data dan pusat klaster (*centroid*) yang sesuai [2].

Rumus jarak Euclidean antara dua titik pada bidang dengan koordinat (x,y) dan (a,b) adalah [2]:

$$\text{Euclidean dist}((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2} \quad (1)$$

Cara kerja algoritma *k-Means Clustering* diilustrasikan dalam lima langkah berikut [2]:

Langkah 1: Mulai dengan pemilihan nilai *k* dimana *k* = jumlah klaster

Langkah 2: Pembuatan matriks jarak di antara *centroid* dan tiap sampel data

Langkah 3: Tempatkan setiap sampel dalam klaster dengan *centroid* terdekat (jarak minimal)

Langkah 4: Perbarui *centroid* baru untuk setiap klaster

Langkah 5: Ulangi sampai tidak ada perubahan lebih lanjut

2.2 Software Tools Data Mining

2.2.1 Rapidminer

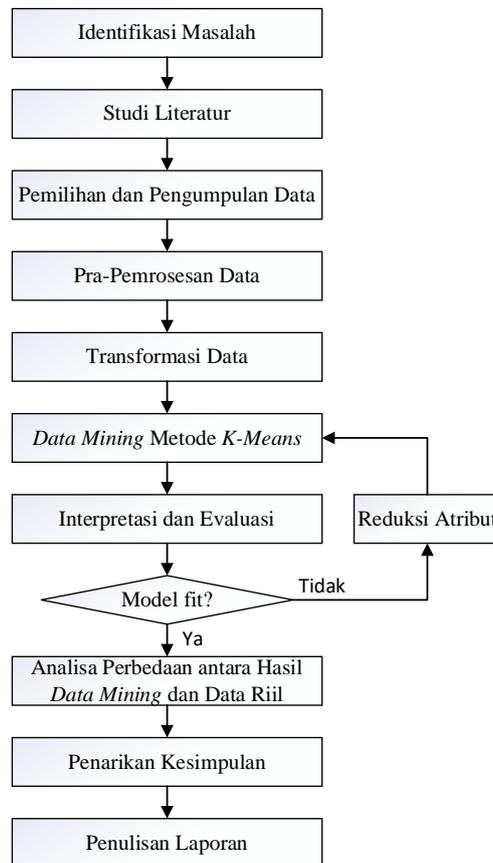
Rapidminer termasuk salah satu *tools data mining* yang *open-source* gratis dan berkemampuan tinggi. Antarmuka pengguna yang ditingkatkan secara grafis, penggunaan sejumlah besar algoritma, dan penggabungan berbagai fitur visualisasi data membedakannya dari *tools data mining* gratis lainnya. Rapidminer juga dapat berjalan di sistem operasi Windows dan Mac. Dengan perubahan terbaru dalam penawarannya, Rapidminer telah membuat versi *tool* analisis yang di-*scale-down* secara gratis (edisi komunitas) sambil membuat produk komersial lengkap. [10]

2.2.2 SPSS

SPSS pada awalnya sebagai *software* statistik yang diperuntukkan bagi pengolahan data statistik untuk ilmu sosial (SPSS saat itu adalah singkatan dari *Statistical Package for the Social Sciences*). Dalam perkembangannya saat ini SPSS tidak hanya menangani permasalahan statistik saja, namun sudah meluas ke *data mining* dan *predictive analytic*. Kepanjangan SPSS pun berubah menjadi *Statistical Product for Service Solutions*. Pada akhir 2009 SPSS secara resmi diakuisisi oleh IBM, dan namanya berubah menjadi IBM SPSS. [9]

3. Metodologi

Alur penelitian sebagaimana pada gambar berikut:



Gambar 1. Alur Penelitian

Dalam melakukan *data mining* dengan metode *k-Means*, penelitian ini menggunakan *tools* Rapidminer (Rapidminer Studio versi 9.10.011) dan SPSS (IBM SPSS Statistics versi 27). Hasil interpretasi *data mining* dievaluasi dengan melihat tingkat kesamaan pengelompokan UKT antara hasil *clustering* dan data riil. Jika tingkat kesamaan belum mencapai 50% (model dianggap belum fit), maka dilakukan kembali proses *data mining* dengan reduksi atribut.

Data penelitian adalah data mahasiswa baru program sarjana (S1) Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Jambi tahun 2022 yang bersumber dari *database* Sistem Registrasi *Online* Universitas Jambi.

4. Hasil dan Pembahasan

4.1. Pemilihan dan Pengumpulan Data

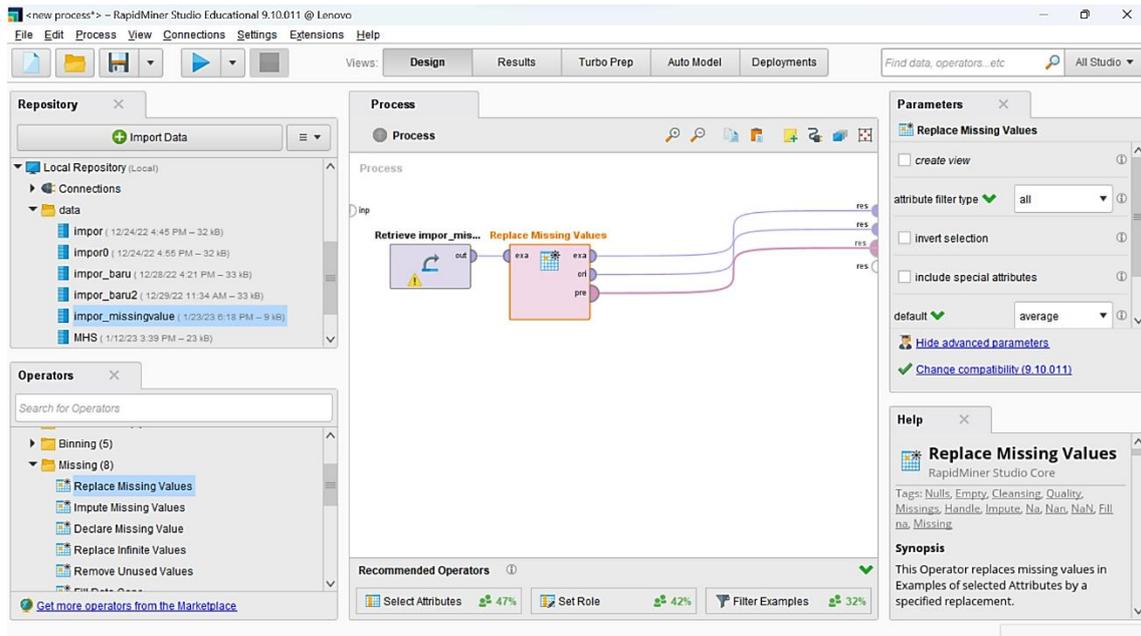
Data penelitian sebanyak 219 *record* data mahasiswa baru. Atribut data sebagian tercatat pada tabel *database* sistem registrasi *online* dan sebagian tercatat dalam *file* dokumen pendukung. Untuk proses *data mining*, dipilih atribut sebagai berikut:

- dari tabel *database*: nomor tes, nomor KIP (nomor pendaftar program Kartu Indonesia Pintar Kuliah), status ayah, status ibu, nama wali, pekerjaan ayah, pekerjaan ibu, pekerjaan wali, penghasilan ayah, penghasilan ibu, penghasilan wali, daya listrik, dan jumlah tanggungan (yang masih sekolah),
- dari *file* dokumen pendukung: tagihan PBB, tagihan listrik, kepemilikan rumah orang tua, kondisi rumah, pendapatan tambahan, dan kendaraan yang dimiliki

Untuk evaluasi, dipilih 3 atribut dari tabel *database* yaitu UKT, estimasi penghasilan, dan catatan pertimbangan.

4.2. Pra-pemrosesan Data

Dalam tahap ini dilakukan *data cleaning* penanganan data inkonsisten dan penanganan *missing value*. Perbaikan data inkonsisten dilakukan terhadap 83 nilai atribut pada 65 *record* data, dengan mengacu pada *file* dokumen pendukung. Penanganan *missing value* dilakukan dengan metode diisi dengan nilai rata-rata atribut. Proses pengisian *missing value* dilakukan menggunakan Rapidminer (Gambar 1). Pengisian *missing value* dilakukan terhadap 54 nilai atribut pada 43 *record* data.



Gambar 2. Desain Proses Replace Missing Values pada Rapidminer

4.3. Transformasi Data

Transformasi data dilakukan dengan mengkategorikan data dan memberikan skor sesuai dengan tingkatan kondisi ekonomi. Semakin tinggi tingkat kesejahteraan atau kondisi ekonomi, semakin tinggi skor yang diberikan. Selain itu, dilakukan juga penggabungan beberapa atribut menjadi atribut gabungan. Transformasi data menghasilkan 11 atribut sebagai berikut:

1. Penghasilan Total (PT): gabungan dari penghasilan ayah, penghasilan ibu, penghasilan wali, dan pendapatan tambahan.

Tabel 1. Transformasi Data Penghasilan Total

Penghasilan Total (PT)	Skor
Rp0 – Rp1.000.000	1
> Rp1.000.000 – Rp2.000.000	2
> Rp2.000.000 – Rp3.000.000	3
> Rp3.000.000 – Rp4.000.000	4
> Rp4.000.000 – Rp5.000.000	5
> Rp5.000.000 – Rp6.000.000	6
> Rp6.000.000 – Rp7.000.000	7
> Rp7.000.000 – Rp8.000.000	8
> Rp8.000.000	9

2. Jumlah Tanggungan (JT)

Tabel 2. *Transformasi Data Jumlah Tanggungan (Masih Bersekolah)*

Jumlah Tanggungan (JT)	Skor
1 orang	3
2-3 orang	2
≥ 4 orang	1

3. Pendaftar KIP Kuliah (PKIPK)

Tabel 3. *Transformasi Data Pendaftar Program KIP-Kuliah*

Pendaftar KIP Kuliah (PKIPK)	Skor
Pendaftar program KIP-Kuliah	1
Bukan pendaftar program KIP-Kuliah	2

4. Status Orang Tua (SO): rata-rata skor dari status ayah dan status ibu.

Tabel 4. *Transformasi Data Status Ayah dan Status Ibu*

Status Ayah, Status Ibu	Skor
Sudah meninggal	1
Cerai	2
Masih hidup	3

5. Pekerjaan (PEK): rata-rata skor dari pekerjaan ayah, pekerjaan ibu, dan pekerjaan wali, dengan formulasi:

- Jika dibiayai orang tua: $PEK = \frac{1}{2}$ (skor pekerjaan ayah + skor pekerjaan ibu)
- Jika dibiayai wali dan salah satu dari ayah/ibu: $PEK = \frac{1}{2}$ (skor pekerjaan wali + salah satu skor pekerjaan ayah/ibu yang membiayai)
- Jika hanya dibiayai wali: $PEK = \frac{1}{2}$ (skor pekerjaan wali + 1)

Tabel 5. *Transformasi Data Pekerjaan Ayah, Pekerjaan Ibu, Pekerjaan Wali*

Pekerjaan Ayah, Pekerjaan Ibu, Pekerjaan Wali	Skor
Sudah Meninggal	1
Tidak Bekerja	
Nelayan	2
Petani	
Peternak	
Pedagang Kecil	
Buruh	
Pensiunan	
Lainnya	3
Pedagang Besar	
Wiraswasta	
Wiraswasta	4
PNS/TNI/Polri	
Karyawan Swasta	
Pegawai BUMN	
Pegawai BUMD	

6. Kepemilikan Rumah (KPR)

Tabel 6. Transformasi Data Kepemilikan Rumah Orang Tua

Kepemilikan Rumah (KPR)	Skor
sewa rumah, menumpang tinggal	1
kredit berjalan, milik perusahaan/dinas	2
milik sendiri	3

7. Kondisi Rumah (KOR)

Tabel 7. Transformasi Data Kondisi Rumah

Kondisi Rumah (KOR)	Skor
sangat sederhana	1
sederhana	2
menengah	3
mewah	4
sangat mewah	5

8. Tagihan Listrik (TL)

Tabel 8. Transformasi Data Tagihan Listrik

Tagihan Listrik (TL)	Skor
Rp0 – Rp100.000	1
> Rp100.000 – Rp250.000	2
> Rp250.000 – Rp500.000	3
> Rp500.000	4

9. Daya Listrik (DL)

Tabel 9. Transformasi Data Daya Listrik

Daya Listrik (DL)	Skor
$\leq 450^*$	1
900	2
1300	3
≥ 2200	4

10. Tagihan PBB (PBB)

Transformasi data tagihan PBB mengacu pada nilai PBB untuk tingkatan Nilai Jual Objek Pajak (NJOP) Rp100 juta, Rp250 juta, dan Rp 500 juta. Berdasarkan peraturan yang berlaku [6, 13]: (a) apabila NJOP kurang dari satu miliar rupiah, Nilai Jual Kena Pajak (NJKP) sebesar 20% dari NJOP [6], (b) NJOP Tidak Kena Pajak (NJOPTKP) paling sedikit Rp10 juta [13], dan (c) tarif PBB-P2 maksimum 0,5% [13]. Perhitungan $PBB=0,5\%*20\%*(NJOP-Rp10 \text{ juta})$, maka untuk NJOP Rp100 juta, Rp250 juta, dan Rp500 juta, nilai PBB adalah Rp90 ribu, Rp240 ribu, dan Rp490 ribu.

Tabel 10. Transformasi Data Tagihan PBB

Tagihan PBB (PBB)	Skor
Kepemilikan rumah: sewa rumah, menumpang tinggal, milik perusahaan/dinas	1
Rp0 – Rp90.000	
> Rp90.000 – Rp240.000	2
> Rp240.000 – Rp490.000	3
> Rp490.000	4

11. Kendaraan (K)

Tabel 11. Transformasi Data Kendaraan yang Dimiliki

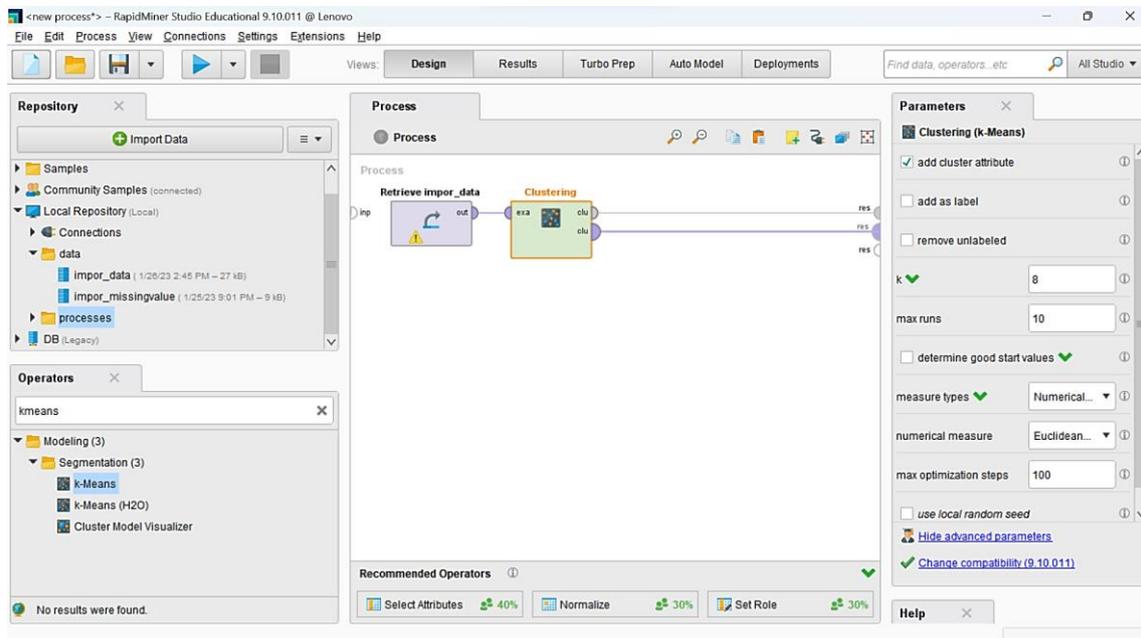
Kendaraan (K)	Skor
tidak ada	1
sepeda motor	2
mobil	3

4.4. Data Mining K-Means Clustering dengan 11 Atribut, Interpretasi, dan Evaluasi

Pada tahap ini dilakukan *data mining* metode *k-Means* menggunakan 11 atribut hasil transformasi data (PT, JT, PKIPK, SO, PEK, KPR, KOR, TL, DL, PBB, dan K).

4.4.1. K-Means Clustering dengan 11 Atribut pada Rapidminer

Tampilan desain proses *k-Means Clustering* pada Rapidminer sebagaimana Gambar 2.



Gambar 3. Desain Proses K-Means Clustering pada Rapidminer

Setting parameter yang digunakan yaitu:

- “*add cluster attribut*”, agar informasi kluster tampil pada data hasil *clustering*
- $k = 8$, yaitu jumlah kluster yang ditentukan
- *measure types* = “*NumericalMeasures*”
- *numerical measure* = “*EuclideanDistance*”

Hasil proses *data mining* berupa jumlah anggota kluster dan *centroid* akhir sebagaimana Tabel 12.

Tabel 12. Jumlah Anggota Klaster dan Centroid Akhir dengan 11 Atribut pada Rapidminer

Klaster	Jumlah Anggota	Centroid										
		PT	JT	PKIPK	SO	PEK	KPR	KOR	TL	DL	PBB	K
0	20	5,600	2,200	2	2,950	2,675	2,600	2,550	2,150	2,850	1,050	2,150
1	23	7,783	2,087	2	3	3,174	2,826	2,870	2,087	2,435	1,130	2,609
2	23	3,609	2,043	2	2,826	2,130	2,652	2,826	3,348	2,870	1,522	2,261
3	47	1,660	2,021	1,340	2,809	1,702	3	1,277	1,170	1,979	1,064	1,851
4	20	8,800	2,050	2	3	3,375	2,950	3,350	3,400	3,200	1,300	2,750
5	31	2,097	2,194	1,516	2,710	1,661	1	1,677	1,645	2,290	1	1,871
6	34	2,529	2,265	1,824	2,971	1,618	2,971	1,676	2,353	2,735	1,059	2,029
7	21	3,714	1,857	1,619	2,810	2,262	2,524	1,952	1,333	1,762	1	2,190

Interpretasi pemetaan 8 klaster kepada 8 kelompok UKT dilakukan berdasarkan urutan nilai atribut penghasilan total (PT) dari yang terendah hingga tertinggi pada 8 *centroid* akhir. Dengan demikian, kelompok UKT I hingga VIII berturut-turut adalah klaster 3, 5, 6, 2, 7, 0, 1, 4.

Evaluasi tingkat kesamaan pengelompokan UKT antara hasil interpretasi *data mining* dan data riil disajikan pada Tabel 13, yaitu sebesar 38,81%.

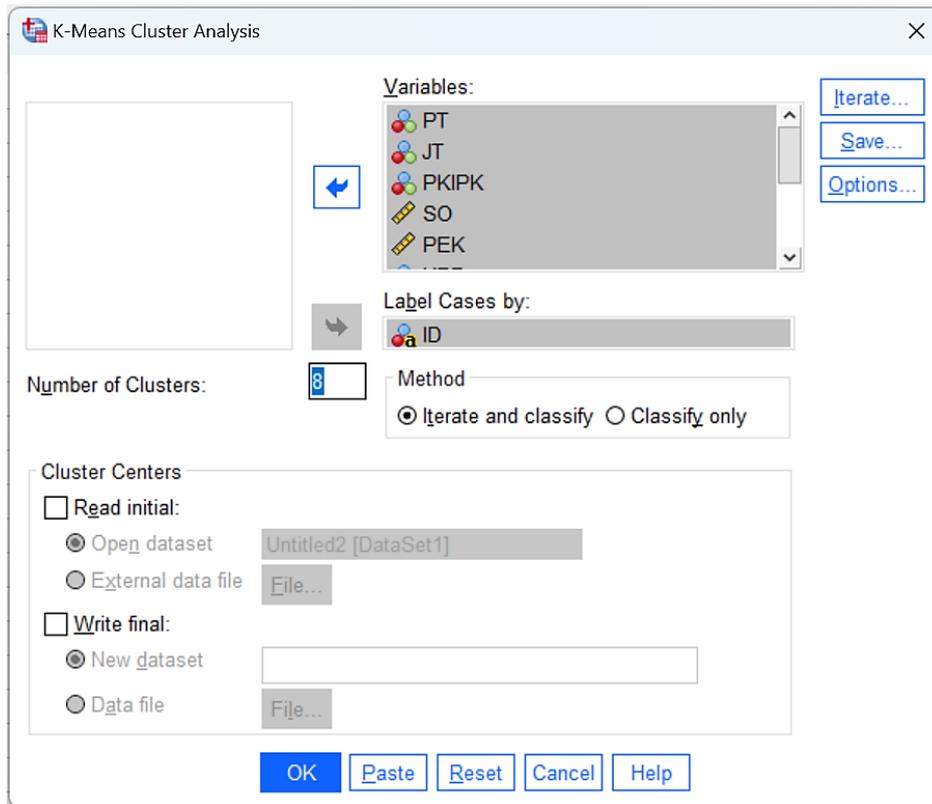
Tabel 13. Evaluasi Tingkat Kesamaan Pengelompokan UKT antara Hasil Clustering 11 Atribut pada Rapidminer dan Data Riil

Kelompok UKT K-Means	Kelompok UKT Data Riil									Total
	KIPK	I	II	III	IV	V	VI	VII	VIII	
I	22	12	12	1						47
II	8	11	12							31
III	3	4	23	3	1					34
IV			7	5	7	4				23
V	8		4	7			1		1	21
VI				1	3	6	4	1	5	20
VII					1	2	8		12	23
VIII							3		17	20
Total	41	27	58	17	12	12	16	1	35	219
Total Sama	85	38,81%								

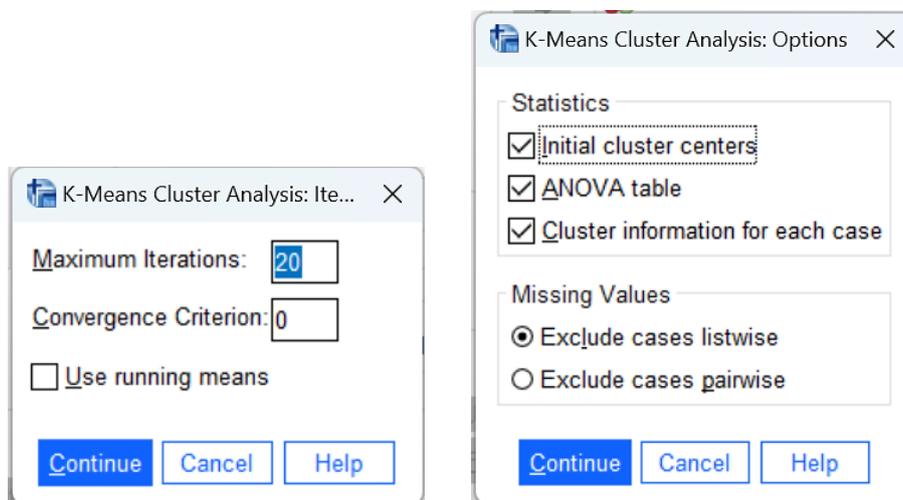
Pada data riil, kolom KIPK menunjukkan mahasiswa penerima beasiswa program KIP-Kuliah (d/h Bidikmisi). Berdasarkan aturan UKT bahwa penerima Bidikmisi bersama penerima UKT kelompok I dan II termasuk dalam paling sedikit 20% dari seluruh mahasiswa baru di perguruan tinggi negeri [3]. Oleh karena itu penerima KIPK dikomparasi dengan kelompok UKT I dan II.

4.4.2. K-Means Clustering dengan 11 Atribut pada SPSS

Tampilan *setting K-Means Cluster Analysis, Iterate, dan Options* pada SPSS sebagaimana Gambar 3 dan Gambar 4.



Gambar 4. Setting K-Means Cluster Analysis pada SPSS



Gambar 5. Setting Iterate dan Options K-Means Cluster Analysis pada SPSS

Output data mining berupa histori iterasi menunjukkan bahwa jumlah iterasi proses *k-Means Clustering* sebanyak 12 kali.

Tabel ANOVA menunjukkan atribut mana yang paling berkontribusi pada solusi kluster. Variabel dengan nilai F besar memberikan pemisahan terbesar antar kluster [2]. Output Tabel ANOVA pada Tabel 14 menunjukkan bahwa PT (penghasilan total) memiliki nilai F terbesar, yang artinya PT paling berkontribusi pada solusi kluster.

Tabel 14. Tabel ANOVA dengan 11 Atribut pada SPSS

Atribut	Kluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
PT	177,885	7	0,617	211	288,371	0,000

Atribut	Klaster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
JT	0,274	7	0,280	211	0,978	0,448
PKIPK	2,326	7	0,129	211	17,990	0,000
SO	0,354	7	0,119	211	2,965	0,006
PEK	12,923	7	0,290	211	44,628	0,000
KPR	14,515	7	0,161	211	89,958	0,000
KOR	14,777	7	0,349	211	42,394	0,000
TL	19,440	7	0,338	211	57,562	0,000
DL	5,938	7	0,496	211	11,969	0,000
PBB	2,914	7	0,129	211	22,546	0,000
K	2,813	7	0,159	211	17,648	0,000

Hasil proses *data mining* berupa jumlah anggota klaster dan *centroid* akhir sebagaimana Tabel 15.

Tabel 15. Jumlah Anggota Klaster dan Centroid Akhir dengan 11 Atribut pada SPSS

Klaster	Jumlah Anggota	Centroid										
		PT	JT	PKIPK	SO	PEK	KPR	KOR	TL	DL	PBB	K
1	54	1,833	2,019	1,333	2,815	1,722	2,981	1,333	1,167	1,963	1,056	1,870
2	4	3	2,250	2	2,5	1,375	3	2,750	3,250	2,750	3,250	2
3	16	4,688	2,125	2	2,938	2,563	2,500	3	3,500	3,063	1,188	2,313
4	20	8,500	2,050	2	3	3,225	2,950	3,350	3,400	3,400	1,350	2,800
5	25	7,880	2,040	2	3	3,360	2,840	2,880	2,120	2,240	1,080	2,520
6	28	4,679	2	1,893	2,875	2,375	2,571	2,286	1,571	2,357	1,036	2,250
7	33	2,152	2,152	1,515	2,712	1,697	1	1,697	1,697	2,303	1	1,879
8	39	2,615	2,256	1,872	2,923	1,705	2,949	1,769	2,487	2,692	1,051	2,051

Interpretasi pemetaan 8 klaster kepada 8 kelompok UKT dilakukan berdasarkan urutan nilai atribut penghasilan total (PT) dari yang terendah hingga tertinggi pada 8 *centroid* akhir. Dengan demikian, kelompok UKT I hingga VIII berturut-turut adalah klaster 1, 7, 8, 2, 6, 3, 5, 4.

Evaluasi tingkat kesamaan pengelompokan UKT antara hasil interpretasi *data mining* dan data riil disajikan pada Tabel 16, yaitu sebesar 37,44%.

Tabel 16. Evaluasi Tingkat Kesamaan Pengelompokan UKT antara Hasil Clustering 11 Atribut pada SPSS dan Data Riil

Kelompok UKT K-Means	Kelompok UKT Data Riil								Total	
	KIPK	I	II	III	IV	V	VI	VII		VIII
I	27	12	13	2						54
II	9	11	12	1						33
III	2	4	28	4	1					39
IV			3			1				4
V	3		2	9	4	3	4		3	28
VI				1	6	6	1	1	1	16
VII					1	2	5		17	25
VIII							6		14	20
Total	41	27	58	17	12	12	16	1	35	219
Total Sama	82	37,44%								

4.5. Data Mining K-Means Clustering dengan 7 Atribut, Interpretasi, dan Evaluasi

Untuk mendapatkan tingkat kesamaan pengelompokan UKT yang lebih tinggi antara hasil *data mining* dan data riil, maka dilakukan reduksi atribut. Dari beberapa percobaan *data mining* dengan reduksi atribut pada Rapidminer dan SPSS, diperoleh hasil terbaik adalah dengan 7 atribut, yaitu PT, PEK, TL, SO, PBB, KOR, dan JT.

4.5.1. K-Means Clustering dengan 7 Atribut pada Rapidminer

Hasil proses *data mining* berupa jumlah anggota klaster dan *centroid* akhir sebagaimana Tabel 17.

Tabel 17. Jumlah Anggota Klaster dan Centroid Akhir dengan 7 Atribut pada Rapidminer

Klaster	Jumlah Anggota	Centroid						
		PT	PEK	TL	SO	PBB	KOR	JT
0	27	6,333	2,963	2,593	2,963	1,111	2,778	2,074
1	18	8,833	3,972	2,778	3	1,056	3,389	2,222
2	14	3,786	2,143	3,643	2,857	1,857	2,929	2,071
3	69	1,681	1,551	1,406	2,768	1,014	1,449	2,174
4	13	8,615	2,308	2,462	3	1,385	2,846	2
5	24	2,625	2,104	1,208	2,833	1,083	1,333	2,083
6	29	2,690	1,741	2,862	2,931	1,069	1,966	2,034
7	25	4,200	2,300	1,520	2,880	1,040	2,240	1,960

Interpretasi pemetaan 8 klaster kepada 8 kelompok UKT dilakukan berdasarkan urutan nilai atribut penghasilan total (PT) dari yang terendah hingga tertinggi pada 8 *centroid* akhir. Dengan demikian, kelompok UKT I hingga VIII berturut-turut adalah klaster 3, 5, 6, 2, 7, 0, 4, 1.

Evaluasi tingkat kesamaan pengelompokan UKT antara hasil interpretasi *data mining* dan data riil disajikan pada Tabel 18, yaitu sebesar 50,23%.

Tabel 18. Evaluasi Tingkat Kesamaan Pengelompokan UKT antara Hasil Clustering 7 Atribut pada Rapidminer dan Data Riil

Kelompok UKT K-Means	Kelompok UKT Data Riil									Total
	KIPK	I	II	III	IV	V	VI	VII	VIII	
I	27	21	20	1						69
II	9	3	11	1						24
III	1	3	20	5						29
IV			3	1	6	4				14
V	4		4	9	3	2	2		1	25
VI					2	4	11	1	9	27
VII					1	2	3		7	13
VIII									18	18
Total	41	27	58	17	12	12	16	1	35	219
Total Sama	110	50,23%								

4.5.2. K-Means Clustering dengan 7 Atribut pada SPSS

Output data mining berupa histori iterasi menunjukkan bahwa jumlah iterasi proses *k-Means Clustering* sebanyak 8 kali.

Output Tabel ANOVA pada Tabel 19 menunjukkan bahwa PT (penghasilan total) memiliki nilai F terbesar, yang artinya PT paling berkontribusi pada solusi klaster.

Tabel 19. Tabel ANOVA dengan 7 Atribut pada SPSS

Atribut	Klaster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
PT	185,802	7	0,354	211	524,577	0,000
PEK	13,383	7	0,274	211	48,789	0,000
TL	18,152	7	0,380	211	47,707	0,000
SO	0,294	7	0,121	211	2,423	0,021
PBB	2,980	7	0,127	211	23,449	0,000
KOR	16,376	7	0,296	211	55,412	0,000
JT	0,269	7	0,280	211	0,962	0,460

Hasil proses *data mining* berupa jumlah anggota klaster dan *centroid* akhir sebagaimana Tabel 20.

Tabel 20. Jumlah Anggota Klaster dan Centroid Akhir dengan 7 Atribut pada SPSS

Klaster	Jumlah Anggota	Centroid						
		PT	PEK	TL	SO	PBB	KOR	JT
1	3	3	1,333	3	2,667	3,667	2,667	2,333
2	17	8,529	2,853	2	3	1,118	2,941	2,059
3	24	4,875	2,417	1,542	2,979	1,042	2,042	1,958
4	49	2,653	1,735	2,224	2,867	1,041	1,878	2,204
5	18	6,667	3,194	2,833	3	1,167	3	1,944
6	22	3,818	2,273	3,273	2,773	1,182	2,909	2,045
7	72	1,750	1,681	1,250	2,778	1,042	1,306	2,097
8	14	9	3,786	3,429	3	1,286	3,429	2,214

Interpretasi pemetaan 8 klaster kepada 8 kelompok UKT dilakukan berdasarkan urutan nilai atribut penghasilan total (PT) dari yang terendah hingga tertinggi pada 8 *centroid* akhir. Dengan demikian, kelompok UKT I hingga VIII berturut-turut adalah klaster 7, 4, 1, 6, 3, 5, 2, 8.

Evaluasi tingkat kesamaan pengelompokan UKT antara hasil interpretasi *data mining* dan data riil disajikan pada Tabel 21, yaitu sebesar 55,71%, yang merupakan hasil tertinggi dalam penelitian ini.

Tabel 21. Evaluasi Tingkat Kesamaan Pengelompokan UKT antara Hasil Clustering 7 Atribut pada SPSS dan Data Riil

Kelompok UKT K-Means	Kelompok UKT Data Riil									Total
	KIPK	I	II	III	IV	V	VI	VII	VIII	
I	31	21	19	1						72
II	6	6	32	4	1					49
III			2			1				3
IV			4	6	7	3	1		1	22
V	4		1	6	3	3	4	1	2	24
VI						3	8		7	18
VII					1	2	3		11	17
VIII									14	14
Total	41	27	58	17	12	12	16	1	35	219
Total Sama	122	55,71%								

4.6. Pembahasan Perbedaan antara Hasil Clustering dengan Data Riil

Berdasarkan evaluasi hasil *k-Means Clustering* dengan 11 atribut dan 7 atribut menggunakan Rapidminer dan SPSS, diperoleh hasil *clustering* dengan tingkat kesamaan tertinggi terhadap pengelompokan UKT data riil adalah pada *k-Means Clustering* dengan 7 atribut menggunakan SPSS, yaitu sebanyak 122 *record* data atau 55,71% dari total 219 *record* data. Adapun perbedaan pengelompokan UKT antara hasil *clustering* dengan 7 atribut pada SPSS dan data riil pada sebanyak 97 *record* (44,29%), secara umum dipengaruhi oleh:

- Pertimbangan penentuan kelompok UKT data riil cenderung memprioritaskan atribut tertentu (seperti penghasilan total / status orang tua / pekerjaan orang tua / jumlah tanggungan), sedangkan *k-Means* memperhitungkan semua atribut. *K-Means* menempatkan *record* pada suatu klaster karena seluruh nilai atribut *record* memberikan jarak *Euclidean* terdekat ke *centroid* klaster tersebut, sehingga seluruh anggota suatu klaster saling memiliki kemiripan nilai atribut dengan karakteristik yang tercermin pada *centroid* akhir klaster.
- Pada penentuan kelompok UKT data riil terdapat beberapa *human error* dalam catatan pertimbangan.
- Adanya komponen pendapatan tambahan dalam atribut penghasilan total (PT) yang menjadi parameter *k-Means* dalam penelitian ini.
- Pada penerima program KIP Kuliah, sebagian masih memenuhi syarat penghasilan total KIP Kuliah, sedangkan sebagian lainnya lolos sebagai penerima KIP Kuliah karena faktor selain penghasilan total.

5. Kesimpulan

Dari hasil penelitian pengelompokan sebanyak 219 *record* data mahasiswa baru Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Jambi tahun 2022 ke dalam 8 kelompok UKT dengan menerapkan *data mining* metode *k-Means*, dihasilkan beberapa kesimpulan dan saran sebagai berikut:

5.1 Simpulan

1. Hasil penelitian menunjukkan terdapat 7 variabel/atribut parameter kondisi ekonomi yang dapat digunakan dalam analisis *k-Means*, yaitu: penghasilan total orang tua/wali termasuk pendapatan tambahan, pekerjaan orang tua/wali, tagihan listrik, status orang tua, tagihan PBB, kondisi rumah, dan jumlah tanggungan yang masih sekolah.
2. Hasil penelitian menunjukkan antara hasil interpretasi analisis *k-Means* dengan 7 variabel/atribut parameter memiliki tingkat kesamaan dengan penentuan UKT data riil di atas 50% baik pada Rapidminer maupun SPSS. Tingkat kesamaan yang lebih tinggi dihasilkan pada SPSS, yaitu sebesar 55,71%, dengan jumlah anggota setiap klaster kelompok UKT meliputi UKT I sebanyak 72 orang, UKT II sebanyak 49 orang, UKT III sebanyak 3 orang, UKT IV sebanyak 22 orang, UKT V sebanyak 24 orang, UKT VI sebanyak 18 orang, UKT VII sebanyak 17 orang, dan UKT VIII sebanyak 14 orang. Dengan demikian metode *k-Means* dapat diterapkan untuk mendukung pengambilan keputusan penetapan kelompok UKT bagi mahasiswa baru.

5.2 Saran

1. Universitas Jambi dapat menerapkan metode *k-Means* dalam pengelompokan UKT mahasiswa baru menggunakan *tools data mining* ataupun pengembangan dari sistem registrasi *online* yang sudah ada.
2. Saran bagi penelitian selanjutnya antara lain:
 - a. Analisis, perancangan, dan/atau implementasi integrasi metode *k-Means* ke dalam sistem registrasi *online* Universitas Jambi.
 - b. Penerapan metode penentuan *centroid* awal.
 - c. Penerapan metode validasi hasil *clustering*.
 - d. Penerapan metode *data mining* lainnya, khususnya yang dapat mempertimbangkan skala prioritas atribut parameter.

6. Daftar Rujukan

- [1] Bhatia, P. 2019. *Data Mining and Data Warehousing, Principles and Practical Techniques*. Cambridge University Press.
- [2] IBM. 2022. *K-Means Cluster Analysis*. IBM Documentation. Available at: <https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=edition-k-means-cluster-analysis>. [Accessed 20 February 2023].
- [3] Keputusan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia Nomor 91/M/KPT/2018 tentang Biaya Kuliah Tunggal pada Perguruan Tinggi Negeri di Lingkungan Kementerian Riset, Teknologi, dan Pendidikan Tinggi Tahun Angkatan 2018, 2018.
- [4] Mauladi, Utomo, P. E. P., Hutabarat, B. F., & Putra, R. A. 2020. *Decision Support System to Determine Uang Kuliah Tunggal (UKT) by Combining Naïve Bayes Classifier and Fuzzy-TOPSIS*. Journal of Physics: Conference Series, 1566(1). <https://doi.org/10.1088/1742-6596/1566/1/012098>.
- [5] Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 25 Tahun 2020 tentang Standar Satuan Biaya Operasional Pendidikan Tinggi pada Perguruan Tinggi Negeri di Lingkungan Kementerian Pendidikan dan Kebudayaan, 2020.
- [6] Peraturan Pemerintah Republik Indonesia No. 25 Tahun 2002 tentang Penetapan Besarnya Nilai Jual Kena Pajak untuk Penghitungan Pajak Bumi dan Bangunan, 2002.
- [7] Puslapdik. 2022. *Pedoman Pendaftaran Kartu Indonesia Pintar Kuliah-KIP Kuliah Merdeka 2022 (Versi 1.0)*. Pusat Layanan Pembiayaan Pendidikan Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia. <https://kip-kuliah.kemdikbud.go.id/panduan>.
- [8] Rotondo, A., & Quilligan, F. 2020. *Evolution Path for Knowledge Discovery and Data Mining Process Model*. Review Article. SN Computer Science (2020) 1:109. Springer Nature Singapore Pte Ltd 2020.
- [9] Santoso, S. 2020. *Panduan Lengkap SPSS 26*. PT Elex Media Komputindo.
- [10] Sharda, R., Delen, D., & Turban, E. 2018. *Business Intelligence, Analytics, and Data Science: A Managerial Perspective* (Fourth Edition). Pearson.
- [11] Silveira, R., Oliveira, B., Carvalho, M., & Matos, T. 2020. *A Data Mining Approach for Customer Segmentation Using a SAF-T Based Business Intelligence System*. International Conferences

- Computer Graphics, Visualization, Computer Vision and Image Processing 2020; Big Data Analytics, Data Mining and Computational Intelligence 2020; and Theory and Practice in Modern Computing 2020. <https://www.oecd.org/tax/administration/45167181.pdf>.
- [12] Sugiyono. 2021. *Sistem Pendukung Keputusan Penetapan Uang Kuliah Tunggal Menggunakan Algoritma K-Means Clustering*. Jurnal Sains, Teknologi dan Industri, 18(2), 212–219.
- [13] Undang-Undang (UU) Nomor 1 Tahun 2022 tentang Hubungan Keuangan antara Pemerintah Pusat dan Pemerintah Daerah, 2022.
- [14] Universitas Jambi. 2021. *Sejarah Universitas Jambi*. Available at: <https://www.unja.ac.id/selayang-pandang/sejarah/>. [Accessed 22 September 2022].
- [15] Wierzchoń, S. T., & Kłopotek, M. A. 2018. *Modern Algorithms of Cluster Analysis* (J. Kacprzyk, Ed.; Vol. 34). Springer. <http://www.springer.com/series/11970>
- [16] Zulfadhilah, M., Riadi, I., & Prayudi, Y. 2016. *Log Classification using K-Means Clustering for Identify Internet User Behaviors*. International Journal of Computer Applications, 154(3), 34–39.