

Optimalisasi Jumlah Klaster Uang Kuliah Tunggal pada Data Sosial Ekonomi Mahasiswa

Khusniatul Fahriya¹, Wiyli Yustanti²

^{1,2}Jurusan Teknik Informatika/Sistem Informasi, Universitas Negeri Surabaya

khusniatul.18079@mhs.unesa.ac.id

wiyliyustanti@unesa.ac.id

Abstrak— UKT (Uang Kuliah Tunggal) adalah sistem pembayaran untuk biaya kuliah Perguruan Tinggi Negeri. Setiap perguruan negeri mempunyai ketentuan UKT yang berbeda-beda. Untuk menentukan tarif UKT mahasiswa dibutuhkan sebuah teknik pengelompokan data. Teknik ini bertujuan agar memudahkan pihak PTN dalam menentukan tarif UKT masing-masing mahasiswa. Teknik tersebut dinamakan dengan *clustering*. *Clustering* merupakan pengelompokan data berdasarkan kemiripan. Jika hasil *clustering* memiliki derajat kemiripan yang tinggi di kelompok klaster yang sama, tetapi derajat kemiripan yang lebih rendah di kelompok klaster yang berbeda maka dapat dikatakan *clustering* tersebut baik atau optimal. Dengan adanya *Clustering* maka akan memudahkan untuk mengetahui bagaimana data harus dikelompokkan dengan begitu akan memperoleh gambaran dari kondisi sosial ekonomi mahasiswa yang cukup relatif. Algoritma *clustering* yang digunakan pada penelitian ini adalah *K-Means Clustering*, *K-Medians Clustering*, *K-Modes Clustering* dan *Fuzzy C-Means Clustering*. Dengan adanya algoritma-algoritma *clustering* yang berbeda-beda maka didapatkan jumlah *cluster* yang optimal adalah 6 *cluster*. Untuk membuktikannya maka dibutuhkan uji validitas *clustering*. Pengujian tersebut menggunakan *Silhouette Coefficient*. Nilai *Silhouette Coefficient* jika mendekati 1, maka jumlah *clusternya* dapat dikatakan *representatif*. Dalam pengujian *Silhouette Coefficient* mendapatkan jumlah klaster optimalnya adalah 6 *cluster* dengan nilai sebesar 0.20212705. Sedangkan algoritma yang optimal adalah algoritma *K-Means Clustering* dengan nilai sebesar 0.1274176.

Kata Kunci— Uang Kuliah Tunggal), *Clustering*, *Silhouette Coefficient*

I. PENDAHULUAN

UKT (Uang Kuliah Tunggal) adalah sistem pembayaran untuk biaya kuliah Perguruan Tinggi Negeri. Setiap PTN memiliki persyaratan yang berbeda untuk besaran Uang Kuliah Kesatuan (UKT). Menurut Dirjen Dikti: 97 / E / KU / 2013 [1], dikeluarkan instruksi kepada pimpinan PTN untuk dilakukan penghapusan terhadap biaya masuk dan melaksanakan UKT sebagai biaya kuliah mahasiswa. Mahasiswa wajib membayar Uang Sekolah (UKT) setiap semester. Biaya UKT setiap mahasiswa juga berbeda-beda tergantung dari kondisi sosial ekonominya. Sesuai dengan Surat Edaran Dirjen Dikti No. 272 / E1.1 / KU / 2013 tentang cakupan biaya UKT [2].

Dalam penentuan tarif UKT untuk mahasiswa tentu bukanlah hal yang mudah. Karena tarif UKT harus mempertimbangkan masalah kondisi sosial ekonomi. Dan setiap mahasiswa mempunyai kedudukan serta pendapatan yang berbeda-beda. Untuk menentukan besaran UKT,

mahasiswa diwajibkan untuk memberikan data keuangan orang tua, termasuk pekerjaan orang tua, pendapatan orang tua, anggota keluarga, dll. Setelah mengumpulkan semua data, situasi keuangan siswa akan ditinjau untuk menentukan penerapan tarif UKT yang diterima siswa. Tentunya hal ini membutuhkan tingkat akurasi tinggi dan waktu yang cukup lama. Kesalahan dalam peninjauan dapat mempengaruhi kesalahan juga dalam administrasi pembayaran tarif UKT. Untuk itu dibutuhkan teknik yang dapat mengelompokkan (klaster) data. Penentuan jumlah klaster pada UKT adalah untuk menentukan kelompok UKT pada setiap mahasiswa.

Penelitian tentang penentuan Uang Kuliah Tunggal sebelumnya sudah dilakukan. Terdapat banyak perbedaan metode dalam penelitian sebelumnya, diantaranya Penggolongan UKT Menggunakan metode *Support Vector Machine* [3], Sistem Pendukung Keputusan untuk Penentuan Kategori UKT dengan Metode *Multifactor Evaluation Process* dalam penelitian [4], Penerapan *Fuzzy C-Means* untuk Penentuan Besar UKT Mahasiswa Baru [5]. Pada penelitian ini akan mengacu pada penelitian dari [6] tentang Klastering Perekonomian Mahasiswa Politeknik Negeri Bali Menggunakan *K-Means*. Pada penelitian tersebut Metode *K-Means*, data yang digunakan sebanyak 411 mahasiswa yang dilihat dari kondisi ekonominya. Saat memilih centroid, akan diasumsikan sendiri dengan hasil pada setiap uji coba jumlah *cluster* akan sama dengan hasil *cluster* yang sama juga. Pada penelitian tersebut juga dapat mengatasi perbedaan dari hasil pemilihan centroid yang dilakukan secara acak.

Oleh karena itu, penelitian ini mengambil topik penelitian yang berjudul “Optimalisasi Jumlah Klaster UKT pada Data Sosial Ekonomi Mahasiswa”. Dengan adanya modifikasi pengolahan data berdasarkan pada kondisi ekonomi maka akan sampai hasilnya sesuai dengan yang diinginkan dan akan lebih akurat. Kemudian dengan adanya kluster maka akan memudahkan untuk mengetahui bagaimana data harus dikelompokkan dengan begitu akan memperoleh gambaran dari kondisi sosial ekonomi mahasiswa yang cukup relatif. Klaster juga akan melakukan penemuan proses yang berulang. Untuk jumlah kelompok yang akan di klaster akan diasumsikan sendiri tanpa ada ketentuan terlebih dahulu. Klaster juga akan mengelompokkan data berdasarkan pada kemiripannya. Sedangkan data yang berbeda akan terpisah. Selain itu adanya klaster akan mencari berapa centroid yang paling optimal. Pemilihan algoritma *K-Means Clustering* dalam menentukan jumlah kluster dikarenakan algoritma ini mudah untuk dipelajari, memiliki tingkat efisiensi yang tinggi dalam

melakukan partisi data dan memiliki waktu komputasi yang relatif singkat. Selain itu, *K-Means Clustering* juga memiliki tingkat akurasi yang cukup tinggi untuk mengukur objek. Selain *K-Means Clustering* ada pemilihan algoritma yang lainnya. Pemilihan algoritma *K-Medians Clustering* dalam menentukan jumlah kluster dikarenakan algoritma tersebut data selalu mempunyai hasil yang tidak terlalu jauh jaraknya. Pemilihan algoritma *K-Modes Clustering* dalam menentukan jumlah kluster dikarenakan algoritma tersebut dapat melakukan *clustering* untuk tipe data yang kategorikal. Pemilihan algoritma *Fuzzy C-Means Clustering* dalam menentukan jumlah kluster dikarenakan algoritma tersebut memiliki tingkat akurasi yang tinggi dan waktu komputasi yang cepat. Data UKT mahasiswa ini nantinya akan diklasifikasi berdasarkan algoritma masing-masing. Hasil dari *clustering* menggunakan beberapa algoritma ini nantinya dapat dijadikan sebagai acuan oleh pihak PTN sebagai pertimbangan menentukan besarnya biaya UKT mahasiswa. Percobaan ini dengan menggunakan aplikasi RStudio dengan nilai *Silhouette Coefficient* sebagai acuan pengelompokan cluster. Pengelompokan ini dianalisa agar menjadi sebuah informasi. Selain itu diperoleh hasil algoritma terbaik dari uji validitas yang dilakukan pada setiap algoritma

II. METODE PENELITIAN

A. Sumber Data

Pada penelitian ini sumber datanya menggunakan data mahasiswa Universitas Negeri Surabaya pada tahun akademik 2016 dan 2017. Data yang digunakan sebanyak 39 atribut. Untuk melakukan klasterisasi, data tersebut harus dilakukan *praprocessing* terlebih dahulu. Karena penggunaan datanya untuk klasterisasi yang digunakan hanyalah atribut yang berpengaruh saja.

B. Metode

a. Praprocessing Data

Praprocessing data merupakan proses merubah skala. Pada data UKT ini dilakukan perubahan skala UKT golongan 0 atau K0 menjadi UKT golongan 1 atau K1. Dilakukan perubahan dikarenakan K0 merupakan jalur bidikmisi, sedangkan untuk *clustering* ini tidak menggunakan jalur bidikmisi. Setelah dilakukan perubahan skala maka untuk selanjutnya dilakukan normalisasi data dan mengganti data yang bernilai null. Karena *clustering* pada tiap-tiap algoritma tidak dapat di *cluster* jika terdapat data yang bernilai null.

b. Seleksi Variabel

Untuk menyeleksi variabel ini menggunakan weka. Data UKT mahasiswa sebelum dilakukan seleksi variabel terdapat 39 atribut. Atribut tersebut diambil dari data sosial ekonomi. 39 atribut tersebut akan dijelaskan dalam Tabel I.

TABEL I
 ATRIBUT DATA CLUSTERING

No	Nama Atribut
1	Jumlah Tanggungan
2	Pendidikan ayah
3	Pendidikan ibu
4	Pekerjaan ayah
5	Gaji ayah
6	Pekerjaan ibu
7	Gaji ibu
8	Kepemilikan rumah
9	Sumber listrik
10	Luas tanah
11	Bahan atap
12	Bahan lantai
13	Bahan dinding
14	Jarak dari kota
15	Kondisi dinding
16	Kondisi ruang tamu
17	Kondisi atap
18	Kondisi kamar mandi
19	Kondisi dapur
20	Kondisi balkon
21	Kondisi kamar tidur
22	Kepemilikan kamar mandi
23	Kepemilikan area cuci
24	Kepemilikan toilet
25	Sumber air
26	Tagihan air
27	Tagihan listrik
28	Tagihan telepon
29	Tagihan internet
30	Penghasilan Ayah lainnya
31	Penghasilan Ibu lainnya
32	Jumlah orang yang ada di rumah
33	Jumlah anak sedang bersekolah
34	Kepemilikan motor
35	Kepemilikan mobil
36	Nilai pajak rumah
37	Luas bangunan
38	Kondisi ruang utama
39	Kondisi ruang tengah

Kemudian dilakukan seleksi variabel yang berpengaruh dengan Algoritma *J48 Decision Tree*. Setelah dilakukan pemilihan algoritma maka klik start untuk dilakukan proses seleksi variabel. Hasil dari seleksi variabel diperoleh 37 atribut yang berpengaruh terhadap keputusan UKT sehingga 37 atribut tersebut digunakan untuk *clustering*. Adapun terdapat 2 atribut yang tidak berpengaruh sehingga 2 atribut tersebut tidak digunakan klasterisasi. 2 atribut tersebut yaitu sumber listrik dan kepemilikan kamar mandi.

c. Processing

Pada tahap *processing* dilakukan 4 algoritma untuk *clustering* yaitu *K-Means clustering*, *K-Medians Clustering*, *K-Modes Clustering* dan *Fuzzy*

C-Means Clustering. Setiap algoritma akan menghasilkan hasil kelompok *clustering* yang berbeda.

d. *Uji Validitas*

Pada tahap ini dilakukan pengujian validitas dari hasil 4 algoritma tersebut. Pengujian ini dilakukan untuk melihat kualitas *cluster* yang ada. Pengujian validitas kinerja *clustering* ini dengan menggunakan pengujian dengan kriteria internal. Uji validitas dilakukan dengan *Silhouette Coefficient*. Setelah dilakukan uji validitas maka dibuatlah sebuah perbandingan algoritma berdasarkan hasil *silhouette* tersebut. Dan dilakukan pemilihan algoritma yang baik dan jumlah *cluster* yang optimal berdasarkan nilai *index silhouette*.

III. HASIL DAN PEMBAHASAN

Pada bagian ini, akan dilakukan pembahasan hasil dari pengolahan data yang telah dilakukan dengan metode-metode penelitian yang telah dijelaskan pada bagian sebelumnya.

A. *Seleksi Variabel Data*

Untuk dapat melakukan *clustering* maka harus mempersiapkan datanya terlebih dahulu. Pada penelitian ini sumber datanya menggunakan data mahasiswa tahun 2016 dan 2017. Jumlah data yang digunakan sebanyak 6.065 dengan 39 atribut. Sedangkan tipe data yang digunakan adalah numerik. Data numerik membutuhkan perhitungan jarak *geometric* antar *cluster*. Pada data ini menggunakan jarak *Euclidean Distance*. Setelah data sudah ada maka dilakukan perubahan skala. Perubahan yang terjadi yaitu merubah UKT golongan 0 atau K0 menjadi UKT golongan 1 atau K1. Karena K0 merupakan jalur bidikmisi. Kemudian dilakukan pemilihan variabel. Pemilihan variabel ini adalah untuk mengetahui variabel yang berpengaruh di penentuan UKT. Untuk pemilihan variabelnya menggunakan metode *Decission Tree* dengan menggunakan aplikasi weka. Dari banyaknya data 39 atribut kemudian didapatkan hasil 37 atribut yang berpengaruh terhadap keputusan ukt dengan tingkat akurasi 62,78%. 37 atribut tersebut akan digunakan untuk klasterisasi. Namun sebelum dilakukan klasterisasi, data tersebut harus dilakukan normalisasi data terlebih dahulu terhadap data-data UKT yang ada. Normalisasi data digunakan untuk membuat variabel-variabel data tersebut berada di jangkauan yang sama serta dapat memperkecil perbedaan antar variabel. Untuk itu perlu dilakukan sebuah normalisasi data. Setelah dilakukan proses normalisasi data maka dilakukan penentuan sebuah *cluster center*. Penentuan ini dilakukan secara acak. Untuk jumlah *cluster center* sesuai dengan jumlah *cluster* yang ingin dibentuk. Pada penelitian ini jumlah *cluster* nya adalah 6 *cluster*.

Setelah itu dilakukan perhitungan jarak. Perhitungan jarak pada tiap-tiap algoritma yang akan digunakan adalah untuk memilih jarak terpendeknya pada suatu data dengan menggunakan *cluster center*. Sehingga akan mengetahui *cluster* mana yang masuk ke anggota-anggota tersebut. Jika

sudah dilakukan perhitungan jarak-jarak antara setiap data dengan *cluster center* nya, maka untuk selanjutnya data-data tersebut akan dikelompokkan berdasarkan jarak yang sudah ditentukan tersebut.

B. *Hasil Clustering*

Tahap selanjutnya adalah klasterisasi. Pada penelitian ini dilakukan 4 algoritma *clustering* yaitu *K-Means clustering*, *K-Medians Clustering*, *K-Modes Clustering* dan *Fuzzy C-Means Clustering*. Dilakukan *clustering* sebanyak 4 kali. Akan dilakukan percobaan pada K3, K4, K5, dan K6. Untuk mengetahui perubahan-perubahan pada setiap *cluster*.

TABEL II
 REKAPITULASI HASIL *CLUSTERING* K3

Cluster	Algoritma			
	K-Means	K-Medians	K-Modes	FCM
1	2411	2262	3523	0
2	1256	1799	1566	3220
3	2380	1995	967	2836

Pada tabel II merupakan rekapitulasi hasil *clustering* dengan jumlah *cluster* 3. Untuk hasil *cluster* setiap algoritma berbeda-beda.

TABEL III
 REKAPITULASI HASIL *CLUSTERING* K4

Cluster	Algoritma			
	K-Means	K-Medians	K-Modes	FCM
1	2352	1054	2652	2836
2	1108	2018	1568	0
3	1509	1590	906	0
4	1087	1394	930	3220

Pada tabel III merupakan rekapitulasi hasil *clustering* dengan jumlah *cluster* 4. Untuk hasil *cluster* setiap algoritma berbeda-beda sama dengan hasil dari *cluster* 3.

TABEL IV
 REKAPITULASI HASIL *CLUSTERING* K5

Cluster	Algoritma			
	K-Means	K-Medians	K-Modes	FCM
1	718	1606	2506	0
2	360	1240	653	0
3	959	1103	827	0
4	1973	841	932	2836
5	2046	1266	1183	3220

Pada tabel IV merupakan rekapitulasi hasil *clustering* dengan jumlah *cluster* 5. Untuk hasil *cluster* setiap algoritma berbeda-beda sama dengan hasil dari *cluster* 3 dan 4.

TABEL V
 REKAPITULASI HASIL *CLUSTERING* K6

Cluster	Algoritma			
	K-Means	K-Medians	K-Modes	FCM
1	1877	981	2137	2
2	1200	691	650	0
3	1223	1090	822	5
4	649	1556	668	3214

Cluster	Algoritma			
	K-Means	K-Medians	K-Modes	FCM
5	144	1013	1120	4
6	963	725	659	2931

Pada tabel V merupakan rekapitulasi hasil *clustering* dengan jumlah kluster 6. Dan terlihat bahwa hasil *clustering* setiap algoritma di setiap *clusternya* berbeda-beda. Hal ini dikarenakan rumus setiap algoritma berbeda. Untuk *K-Means Clustering* penentuan pusat *clusternya* berdasarkan rata-rata, *K-Medians Clustering* penentuan pusat *clusternya* berdasarkan median (nilai tengah), *K-Modes Clustering* penentuan pusat *clusternya* berdasarkan modus dan *Fuzzy C-Means Clustering* penentuan pusat *clusternya* berdasarkan *fuzzfier*.

C. Hasil Uji Validitas Clustering

Dari adanya hasil-hasil *clustering* pada masing-masing algoritma, maka dilakukan uji validitas *clustering*. Pengujian ini bertujuan untuk mengetahui pada titik mana pengelompokan ada di nilai yang optimal. Untuk metode yang digunakan untuk pengujian kualitas *cluster* ini adalah *Silhouette Coefficient*. Ada beberapa kriteria subjektif untuk mengukur pengelompokan berdasarkan *Silhouette Coefficient* [7]:

TABEL IV
KRITERIA SUBJEKTIF PENGUKURAN SILHOUETTE COEFFICIENT

Nilai Silhouette Coefficient	Interpretasi Silhouette Coefficient
$\leq 0,25$	Sangat Buruk
0,26 – 0,50	Buruk
0,51 – 0,70	Baik
0,71 – 1,00	Kuat

Kriteria subjektif terbagi menjadi 4 berdasarkan nilainya masing-masing. Untuk nilai *Silhouette Coefficient* $\leq 0,25$ masuk ke dalam interpretasi yang sangat buruk. Nilai *Silhouette Coefficient* 0,26 – 0,50 masuk ke dalam interpretasi yang buruk. Nilai *Silhouette Coefficient* 0,51 – 0,70 masuk ke dalam interpretasi yang baik. Dan nilai *Silhouette Coefficient* 0,71 – 1,00 masuk ke dalam interpretasi yang kuat. Pengujian ini dilakukan pada kluster 6 dikarenakan hasil dari algoritma *Fuzzy C-Means Clustering* banyak yang hasilnya 0. Dan pada kluster 6 hasil *Fuzzy C-Means Clustering* terbentuk kluster secara merata.

TABEL VII
REKAPITULASI HASIL UJI VALIDITAS CLUSTERING

Cluster	Algoritma			
	K-Means	K-Medians	K-Modes	FCM
1	0.19464841	0.14023565	0.048090073	- 0.003884 1312
2	0.07803176	0.11831838	0.020524951	0.000000 00

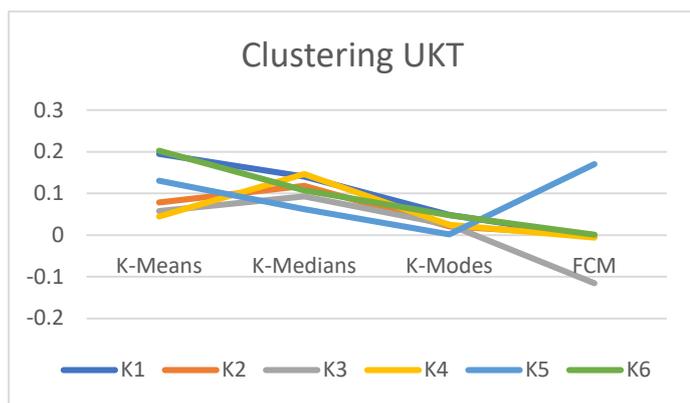
Cluster	Algoritma			
	K-Means	K-Medians	K-Modes	FCM
3	0.05740065	0.09245116	0.024578105	- 0.115811 5043
4	0.04468528	0.14672331	0.024486581	- 0.006269 7697
5	0.13053835	0.06202641	0.001420532	0.170440 5868
6	0.20212705	0.10745240	0.047260013	0.000687 2335

Pada tabel VII merupakan rekapitulasi hasil uji validitas *clustering*. Nilai *Silhouette Coefficient* sangat bervariasi, untuk nilai *silhouette* berada diantara -1 hingga 1. Jika nilai *silhouette* yang dihasilkan mendekati 1 maka jumlah *cluster* tersebut dapat dikatakan *representatif*. Dari hasil uji validitas, kluster terbaik adalah algoritma *K-Means Clustering* dengan nilai index sebesar 0.20212705 dan berada di *cluster* 6. Kluster terbaik algoritma *K-Medians Clustering* dengan nilai index sebesar 0.14672331 dan berada di *cluster* 4. Kluster terbaik algoritma *K-Modes Clustering* dengan nilai index sebesar 0.048090073 dan berada di *cluster* 1. Kluster terbaik algoritma *Fuzzy C-Means Clustering* dengan nilai index sebesar 0.1704405868 dan berada di *cluster* 5. Dari adanya perbedaan kluster terbaik pada setiap algoritma maka harus dilakukan penentuan algoritma terbaik agar mengetahui algoritma yang baik dan jumlah *cluster* yang optimum. Untuk penentuan algoritma terbaik maka hasil dari setiap algoritma harus dilakukan rata-rata uji validitas *Silhouette Coefficient*.

TABEL VIII
RATA-RATA SILHOUETTE COEFFICIENT

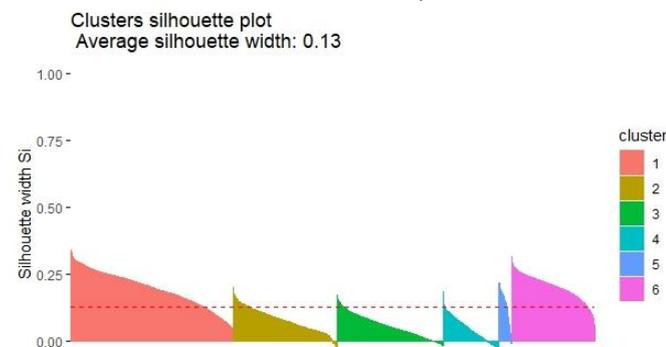
No	Algoritma	Rata-rata
1	K-Means	0.1274176
2	K-Medians	0.1137943
3	K-Modes	0.03061514
4	FCM	-0.002990513

Pada tabel VIII merupakan hasil rata-rata uji validitas dengan *Silhouette Coefficient*. Dibandingkan dengan *K-Medians Clustering*, *K-Modes Clustering* dan *Fuzzy C-Means Clustering*, algoritma *K-Means Clustering* memiliki kinerja yang lebih baik. Nilai *Silhouette Coefficient* *K-Means Clustering* lebih *representatif* karena memiliki nilai index yang mendekati 1 dengan nilai rata-rata 0.1274176. Jadi, algoritma pengelompokan yang optimal dari proses *clustering* UKT adalah algoritma *K-Means Clustering*.



Gbr. 1 Grafik Uji Validitas Silhouette Coefficient

Algoritma K-Means Clustering memiliki nilai index terbesar dibandingkan dengan algoritma yang lainnya. Pada kriteria subjektif Silhouette Coefficient K-Means Clustering termasuk pada Interpretasi Silhouette Coefficient yang buruk karena nilai Silhouette Coefficient $\leq 0,25$.



Gbr. 2 Hasil Silhouette Coefficient K-Means Clustering

IV. KESIMPULAN

Dari adanya pembahasan pada bagian sebelumnya maka didapatkan kesimpulan dari penelitian ini adalah dari 4 algoritma yang digunakan maka, algoritma pengelompokan yang optimal adalah algoritma K-Means Clustering dengan nilai index sebesar 0.1274176. Sedangkan jumlah cluster yang terbaik dengan nilai index sebesar 0.20212705 berada di cluster 6. Dibandingkan dengan K-Medians Clustering, K-Modes Clustering dan Fuzzy C-Means Clustering, algoritma K-Means Clustering memiliki kinerja yang lebih baik.

REFERENSI

[1] Direktorat Jenderal Pembelajaran dan Kemahasiswaan, Surat Edaran Dikjen Dikti Nomor 97/E1.1/KU/2013 tentang Uang Kuliah Tunggal, 2013.
 [2] Direktorat Jenderal Pembelajaran dan Kemahasiswaan, Surat Edaran Dirjen Dikti Nomor 272/E1.1/KU/2013 tentang Kisaran Tarif Uang Kuliah Tunggal, 2013.
 [3] Suyoga, I Gede Eka., I Putu Eka Nila Kencana., dan I Komang Gede Sukarsa, “Penggolonagn Uang Kuliah

Tuggal Menggunakan Support Vector Machine”, E-Jurnal Matematika, vol. 6, hal 220-225, Nov. 2017.
 [4] Noviandha, Friandy Dwi., Indah Fitri Astuti., dan Awang Harsa Kridalaksana, “Sistem Pengukung Keputusan Untuk Penentuan Kategori Uang Kuliah Tuggal dengan Metode Multifactor Evaluation Process (Studi Kasus: Universitas Mulawarman)”, Jurnal Ilmiah Ilmu Komputer, vol. 13, hal 88-96, Sept. 2018.
 [5] Muchsin, Ariyady Kurniawan dan Made Sudarma, “Penerapan Fuzzy C-Means Untuk Penentuan Besar Uang Kuliah Tunggol Mahasiswa Baru Bali”, Lontar Komputer, vol. 6, hal 175-184, Des. 2015.
 [6] Ciptayani, Putu Indah., I G N B Caturbawa., dan I Nyoman Eddy Indrayana, “Klastering Perekonomian Mahasiswa Politeknik Negeri Bali Menggunakan K-Means”, Jurnal Sistem dan Informatika, vol. 10, hal 32-40, Nov. 2015.
 [7] Anggara, et al., “Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member di Alvaro Fitness”, Jurnal Sistem dan Teknologi Informasi (JUSTIN), vol. 1, hal. 1-6, 2016.