



Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal

Haris Kurniawan¹, Sarjon Defit², Sumijan³

^{1,2,3}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Putra Indonesia “YPTK” Padang

¹hrs.kurniawan@gmail.com, ²sarjonde@yahoo.co.uk, ³sumijan@upi.pyk.ac.id

Abstract

Digitalization and automation in student services in Higher Education can produce big data. The mandate of the government in the Regulation of the Minister of Technology Research and Higher Education so that the amount of the Single Tuition (UKT) in State Universities is divided into 5 groups based on the level of parents' socioeconomic conditions. In the process of establishing UKT, there are so many socioeconomic indicators of parents that must be used as a reference, making it difficult to identify and find the right formula. To classify student data is done by data mining techniques using the K-Means Clustering method. This method classifies the size of UKT students based on the pattern or similarity of parents' socioeconomic data. The data used in this study is the data of prospective new students of Universitas Negeri Padang. This grouping aims to help determine the amount of UKT prospective new students at State Universities. The results of the study obtained 5 groups of UKT quantities, consisting of UKT category 1 IDR 500,000, UKT category 2 IDR 1,000,000, UKT category 3 IDR 2,000,000, UKT category 4 IDR 3,000,000 and UKT category 5 IDR 4,000,000.

Keywords: Data Mining, Clustering, K-Means, Single Tuition

Abstrak

Digitalisasi dan otomasi dalam pelayanan mahasiswa di Perguruan Tinggi dapat menghasilkan big data. Amanat pemerintah dalam Peraturan Menteri Riset Teknologi dan Pendidikan Tinggi agar besaran Uang Kuliah Tunggal (UKT) di Perguruan Tinggi Negeri dibagi ke dalam 5 kelompok berdasarkan tingkatan kondisi sosial ekonomi orang tua. Dalam proses menetapkan UKT begitu banyak indikator sosial ekonomi orang tua yang harus dijadikan acuan sehingga menyulitkan dalam mengidentifikasi dan mencari formula yang tepat. Untuk mengelompokkan data mahasiswa ini dilakukan dengan teknik data mining menggunakan metode K-Means Clustering. Metode ini mengelompokkan besaran UKT mahasiswa berdasarkan pola atau kemiripan data sosial ekonomi orang tua. Data yang digunakan dalam penelitian ini adalah data calon mahasiswa baru Universitas Negeri Padang. Pengelompokan ini bertujuan untuk membantu menetapkan besaran UKT calon mahasiswa baru pada Perguruan Tinggi Negeri. Hasil dari penelitian diperoleh 5 kelompok besaran UKT, terdiri dari UKT kategori 1 Rp. 500.000, UKT kategori 2 Rp. 1.000.000, UKT kategori 3 Rp. 2.000.000, UKT kategori 4 Rp. 3.000.000 dan UKT kategori 5 Rp. 4.000.000.

Kata kunci: Data Mining, Clustering, K-Means, Uang Kuliah Tunggal

1. Pendahuluan

Kemajuan teknologi informasi diberbagai bidang telah berdampak pada terjadinya revolusi industri 4.0. Pada masa ini hampir semua kegiatan dilakukan secara digitalisasi dan otomasi. Sistem digitalisasi dan otomasi ini telah meluas pada sistem pemerintahan, sehingga dapat memudahkan dan bahkan menggantikan pekerjaan aparatur negara. Seiring dengan itu, digitalisasi pada pelayanan mahasiswa baru di Perguruan Tinggi telah menghasilkan database yang besar (*big data*). *Big data* akan menyebabkan sulitnya memperoleh informasi atau *knowledge* yang dibutuhkan terkandung didalamnya.

Dalam proses menetapkan besaran UKT begitu banyak indikator sosial ekonomi yang harus menjadi acuan sehingga menyulitkan dalam mengidentifikasi dan

mencari formula yang tepat dalam mengelompokkan data sesuai dengan tingkatan sosial ekonomi orang tua mahasiswa. Pengelompokan yang tidak seimbang dapat menyebabkan besaran UKT yang diterima mahasiswa menjadi tumpang tindih dan kurang tepat sasaran.

Mahasiswa baru merupakan peserta seleksi penerimaan mahasiswa baru yang telah diterima di Perguruan Tinggi. Penerimaan mahasiswa baru diseleksi dengan berbagai jalur masuk seperti; Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) dan berbagai jalur masuk lainnya. Mahasiswa baru yang telah diterima diwajibkan untuk membayar UKT (Uang Kuliah Tunggal). UKT merupakan sebagian BKT (Biaya Kuliah Tunggal) yang ditanggungkan kepada setiap



Lisensi :

Internasional Creative Commons Attribution-ShareAlike 4.0.

mahasiswa berdasarkan kemampuan ekonomi dan sosial orang tuanya. BKT merupakan seluruh biaya operasional setiap mahasiswa yang dibayar setiap awal semester di Perguruan Tinggi Negeri. Sesuai dengan amanat pemerintah dalam Peraturan Menteri Ristek Dikti Nomor 22 tahun 2015 bahwa penerapan besaran UKT (studi kasus di Universitas Negeri Padang) dibagi kedalam 5 (lima) kelompok [8] .

Untuk menemukan sebuah informasi atau *knowledge* yang terkandung didalam database yang besar dapat dilakukan teknik *data mining*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan mesin learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait berbagai database besar [13] . Pada kesempatan ini penelitian *data mining* dilakukan dengan menggunakan metode *K-Means Clustering* untuk menentukan besaran UKT calon mahasiswa baru di Universitas Negeri Padang. *K-Means Clustering* merupakan salah satu metode *clustering* yang mempartisi data yang ada kedalam bentuk satu atau lebih kelompok. Metode *K-Means Clustering* akan mengelompokkan data yang memiliki karakteristik atau pola yang sama kedalam sebuah kelompok yang sama, dan data yang mempunyai karakteristik atau pola yang berbeda kedalam kelompok lainnya. Melalui *clustering* ini akan diketahui informasi atau *knowledge* dari sekumpulan data yang besar.

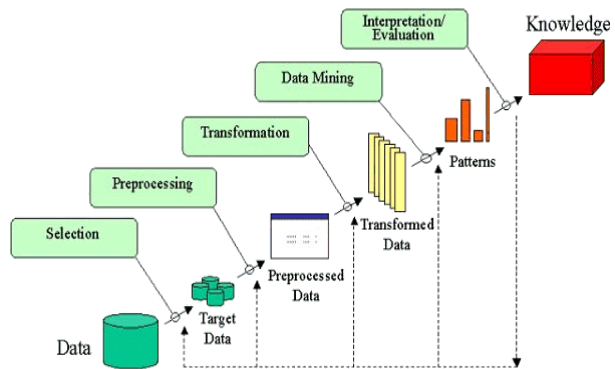
Untuk menentukan besaran UKT dengan metode *K-Means Clustering* ini data diambil dari database registrasi mahasiswa baru di Universitas Negeri Padang. Dari database yang besar itu dipilih data yang berhubungan dengan latar belakang kondisi ekonomi dan sosial orang tua. Kemudian data tersebut dilakukan pemodelan dan penggalan data untuk dikelompokkan berdasarkan pola kemiripan atribut-atributnya. Pada proses ini akan menghasilkan 5 (lima) pengelompokan besaran UKT mahasiswa berdasarkan kondisi ekonomi orang tua mahasiswa. Kelompok inilah nantinya akan menjadi dasar penentuan kategori besaran UKT calon mahasiswa baru.

K-Means Clustering ini telah diimplementasikan diberbagai bidang dalam kehidupan. Dalam bidang kesehatan digunakan untuk menemukan penyakit pasien, dibidang marketing digunakan untuk melihat kelompok target pemasaran, dibidang geografis digunakan untuk pemetaan wilayah potensial, dibidang sosial digunakan untuk mengelompokkan kesejahteraan masyarakat, dan masih banyak lagi bidang-bidang lain yang menerapkan metode *K-Means Clustering* ini.

Knowledge Discovery in Database (KDD)

KDD adalah proses yang tidak sepele yang digunakan untuk mengidentifikasi validitas data, potensi, guna, dan pada akhirnya menghasilkan pola data yang dapat dimengerti [12] . KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data. *Data*

mining sendiri adalah bagian dari tahapan proses KDD yang diilustrasikan seperti pada gambar 1 [3]:



Gambar 1. Proses KDD [11]

Berdasarkan gambar di atas dapat dijelaskan bahwa tahapan KDD adalah:

1. *Data Selection*

Data selection (pemilihan data) merupakan proses penyiapan data dengan memilih sekumpulan data yang menjadi target/atribut/indikator dari dalam data yang besar.

2. *Pre-processing/Cleaning*

Proses *cleaning* (pembersihan data) merupakan proses memperbaiki kesalahan data (*data error*), membuang data ganda, dan memeriksa data yang inkonsisten seperti kesalahan cetak (*tipografi*).

3. *Data Transformation*

Proses *Data Transformation* (transformasi data) merupakan proses perubahan data ataupun pengkodean data. *Data transformasi* dilakukan untuk mempermudah pengolahan data dan merupakan data yang telah siap untuk dilakukan proses *data mining*.

4. *Data mining*

Data mining merupakan proses pencarian pola atau informasi yang menarik/bermanfaat dalam data yang besar dengan menggunakan teknik atau metode tertentu.

5. *Interpretation* atau *Evaluation*

Interpretation atau *Evaluation* merupakan proses menampilkan pola/informasi/pengetahuan dari hasil proses *data mining* ke dalam bentuk yang mudah dimengerti oleh pihak-pihak membutuhkan.

KDD dan *data mining* sering sekali digunakan secara bergantian dalam menjelaskan proses ekstraksi/penggalan informasi yang tersembunyi dalam database yang besar. KDD dan *data mining* memiliki konsep yang berbeda, tetapi memiliki kaitan satu masa lain.

Data Mining

Data mining adalah proses mengekstraksi/menggali/mencari informasi yang tersembunyi dari dalam database yang besar serta bermanfaat bagi pihak yang berkepentingan. *Data mining* dapat membantu pemegang data untuk menganalisis dan menemukan

hubungan yang tidak terduga antar data, pada akhirnya akan membantu pengambilan keputusan.

Data mining atau penambangan data dapat didefinisikan sebagai proses eksplorasi, seleksi, dan pemodelan dari sejumlah data besar untuk menemukan pola atau kecenderungan yang biasanya tidak disadari keberadaannya [4]. Tujuan dari *data mining* ini adalah untuk menemukan pola yang sebelumnya tidak diketahui. Apabila pola sudah ditemukan, selanjutnya digunakan untuk membuat suatu keputusan untuk keperluan tertentu bagi yang berkepentingan. *Data mining* adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan, dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [2].

Clustering

Clustering merupakan proses pengelompokan sekumpulan objek ke dalam kelas/pola objek yang serupa. *Cluster* merupakan teknik yang sangat penting dalam analisis data. *Cluster* merupakan membagi data menjadi kelompok-kelompok atau *cluster* berdasarkan suatu kemiripan atribut-atribut diantara sekumpulan data, atribut-atribut yang mirip tersebut dipresentasikan sebagai titik-titik dalam ruang multidimensi [10].

K-Means Clustering

K-Means Clustering merupakan metode klaster dimana kumpulan data dibagi menjadi beberapa kelompok. Pada tahun 1967 Mac Queen pertama kali mengusulkan algoritma *K-Means*. Metode *K-Means* berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.

Langkah-langkah algoritma *K-Means Clustering* [9] :

1. Pilih secara acak k buah data sebagai pusat *cluster*
2. Jarak antara data dan pusat *cluster* dihitung menggunakan *Euclidian Distance*. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

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

dimana $D(i, j)$ adalah jarak data ke i ke pusat cluster j , X_{ki} adalah data ke i pada atribut data ke k , X_{kj} adalah titik pusat ke j pada atribut ke k

3. Data ditempatkan dalam *cluster* yang terdekat, dihitung dari tengah *cluster*
4. Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat. Rumus menghitung titik pusat *cluster* baru:

$$v = \frac{\sum_{i=1}^n x_i}{N} \quad ; \quad i = 1, 2, 3, \dots, n \quad (2)$$

dimana v adalah *centroid* pada *cluster*, X_i adalah objek ke- i , N adalah banyaknya objek/jumlah objek yang menjadi anggota *cluster*

5. Proses penentuan pusat cluster dan penempatan data dalam cluster diulangi sampai nilai *centroid* tidak berubah lagi

1.2.5 Uang Kuliah Tunggal (UKT)

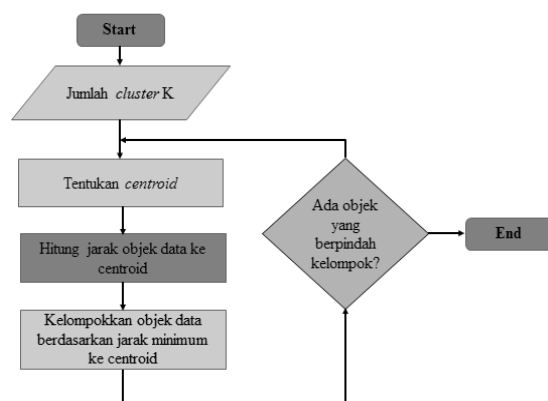
Uang Kuliah Tunggal (UKT) merupakan sebagian biaya kuliah tunggal yang ditanggung setiap mahasiswa pada setiap jurusan/program studi untuk program diploma dan program sarjana berdasarkan kemampuan ekonominya [6]. Pembayaran UKT merupakan implementasi dari pasal 88 ayat 1 sampai 5 pada Undang-Undang No. 12 Tahun 2012 tentang Pendidikan Tinggi. Pada ayat 1 menyatakan bahwa pemerintah menetapkan standar satuan biaya operasional pada pendidikan tinggi secara periodik dengan mempertimbangkan capaian standar nasional Pendidikan Tinggi, jenis program studi dan indeks kemahalan wilayah [5]. Penetapan UKT ini ditetapkan berdasarkan Peraturan Menteri Pendidikan Nasional No. 55 Tahun 2013 yang mengatur tentang Biaya Kuliah Tunggal (BKT) dan Uang Kuliah Tunggal (UKT) di Perguruan Tinggi Negeri dalam lingkungan Kementerian Pendidikan dan Kebudayaan RI [7].

2. Metode Penelitian

2.1 Metode *K-Means Clustering*

Metode *K-Means Clustering* termasuk dalam teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke k daerah bagian yang terpisah. Pada *K-Means*, setiap objek harus masuk dalam kelompok tertentu, tetapi dalam satu tahapan proses tertentu, objek yang sudah masuk dalam satu kelompok, pada satu tahapan berikutnya objek akan berpindah ke kelompok lain.

Pengelompokan data menggunakan algoritma *k-means clustering* ini dilakukan dengan langkah-langkah seperti pada Gambar 2.

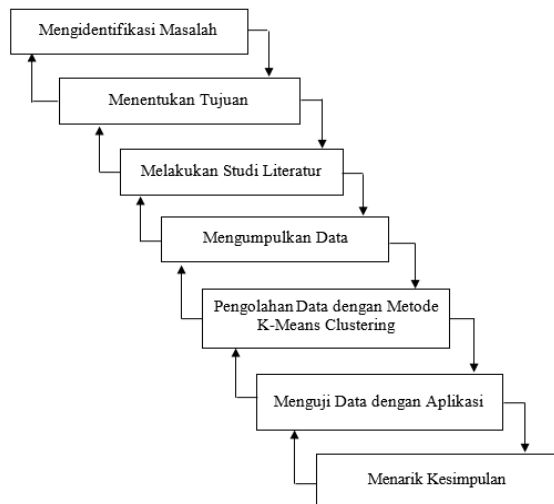


Gambar 2. Diagram Alir Proses *K-Means Clustering*

Untuk memulai algoritma *k-means clustering* diawali dengan menentukan jumlah *cluster*, selanjutnya tentukan pusat *cluster (centroid)*, kemudian dilanjutkan dengan menghitung jarak objek data ke *centroid* dan terakhir lakukan pengelompokan objek data berdasarkan jarak minimum terhadap *centroid*. Satu kali siklus proses *k-means clustering* disebut dengan 1 kali iterasi. Selanjutnya hitung pusat cluster yang baru, setelah itu lanjutkan dengan menghitung jarak data ke pusat cluster baru yang sama prosesnya dengan itersi pertama (iterasi-1) tadi. Jika hasil pengelompokan iterasi ke-2 dan iterasi ke-1 belum sama lanjutkan ke iterasi berikutnya. Namun jika hasil pengelompokan iterasi tersebut sudah sama, maka proses *k-means clustering* di hentikan dan itulah hasil akhir dari pengelompokan akhir.

2.2 Metode Riset

Kerangka penelitian merupakan pedoman dalam setiap langkah-langkah yang akan dilakukan dalam penelitian. Setiap langkah dilakukan sesuai dengan prosedur yang telah direncanakan. Langkah-langkah yang telah direncanakan pada kerangka penelitian ini dapat dilihat pada gambar 3.



Gambar 3. Kerangka Penelitian

Berdasarkan Gambar 3 dapat dilihat bahwa penelitian ini di awali dengan mengidentifikasi masalah yang terdiri dari perumusan masalah dan batasan masalah, selanjutnya menentukan tujuan yang ingin dicapai, melakukan studi literatur/referensi dan mengumpulkan data. Kemudian masuk ketahap pengolahan data dan pengujian dengan aplikasi RapidMiner. Tahap akhir adalah menarik kesimpulan untuk memperoleh intisari dari hasil penelitian

3. Hasil dan Pembahasan

Sistem penentuan besaran UKT calon mahasiswa baru menggunakan algoritma K-Means Clustering dilakukan dengan mengikuti tahap-tahap data mining. Berikut adalah proses data mining dalam menentukan besaran UKT calon mahasiswa baru menggunakan metode K-Means Clustering.

1. Menyiapkan Data Penelitian

Data penelitian yang dikumpulkan bersumber dari database Universitas Negeri Padang. Data yang dikumpulkan adalah data pendaftaran calon mahasiswa baru. Kemudian data yang akan diolah merupakan sampel yang terdiri 30 orang mahasiswa baru.

2. Pemilihan Data (Data Selection)

Pemilihan data merupakan proses pemilihan atribut yang sesuai dengan kebutuhan atau indikator untuk penentuan besaran UKT. Atribut diambil dari database pendaftaran calon mahasiswa baru Universitas Negeri Padang. Untuk kebutuhan *data mining* penentuan besaran UKT dengan algoritma *K-Means Clustering* ini, data yang diambil adalah data yang berkaitan dengan kondisi sosial ekonomi orang tua. Data tersebut terdiri dari nama mahasiswa, penghasilan orang tua, jumlah tanggungan, status rumah, tagihan PBB, tagihan Listrik, tagihan PDAM, pulsa HP, langganan TV, pajak mobil, pajak motor dan asset yang dimiliki.

3. Pembersihan Data (Data Cleaning)

Data yang telah dipilih pada proses *data selection* kemudian dilakukan proses pembersihan data. Proses pembersihan data bertujuan agar data yang tidak valid, seperti adanya data yang salah atau data yang tidak dibutuhkan dilakukan terlebih dahulu pembersihan data dengan cara pembenaran ataupun pembuangan data.

4. Transformasi Data (Data Transformation)

Agar mudah diolah menggunakan metode algoritma *K-Means Clustering*, data calon mahasiswa baru yang akan diolah dilakukan dulu proses transformasi data [1]. Kaedah/ketentuan dari proses transformasi data dibuat berdasarkan ketentuan di Universitas Negeri Padang. Proses transformasi ini dilakukan dengan *scoring* sesuai dengan tingkatan kondisi sosial ekonomi orang tua. Semakin tinggi tingkat kesejahteraan atau kondisi sosial ekonomi orang tua semakin tinggi pula skor yang diberikan. *Scoring* sebagai dasar untuk mentransformasi data penelitian ini adalah:

Tabel 1. Transformasi Data Penghasilan Orang Tua

| Penghasilan Ayah/Ibu | Skor |
|-----------------------------------|------|
| Rp. 0 - < Rp. 1.800.000 | 1 |
| Rp. 1.800.000 – < Rp. 2.000.000 | 2 |
| Rp. 2.000.000 – < Rp. 2.500.000 | 3 |
| Rp. 2.500.000 – < Rp. 3.600.000 | 4 |
| Rp. 3.600.000 – < Rp. 5.000.000 | 5 |
| Rp. 5.000.000 – < Rp. 6.000.000 | 6 |
| Rp. 6.000.000 – < Rp. 7.000.000 | 7 |
| Rp. 7.000.000 – < Rp. 8.500.000 | 8 |
| Rp. 8.500.000 – < Rp. 10.000.000 | 9 |
| Rp. 10.000.000 – < Rp. 12.000.000 | 10 |
| Rp. 12.000.000 – < Rp. 14.000.000 | 11 |
| Rp. 14.000.000 – < Rp. 16.000.000 | 12 |
| > Rp. 16.000.000 | 13 |

Tabel 2. Transformasi Data Tanggungan Orang Tua

| Tanggungan | Skor |
|------------|------|
| >=1 orang | 5 |
| 2 orang | 4 |
| 3 orang | 3 |
| 4 orang | 2 |
| >= 5 orang | 1 |

Tabel 3. Transformasi Data Status Rumah

| Kepemilikan Rumah | Skor |
|-------------------|------|
| Numpang | 1 |
| Sewa Bulanan | 2 |
| Sewa Tahunan | 3 |
| Rumah Dinas | 4 |
| Milik Sendiri | 5 |

Tabel 4. Transformasi Data Tagihan PBB

| Tagihan PBB | Skor |
|---------------------------|------|
| Rp. 0 | 1 |
| < Rp. 50.000 | 2 |
| Rp. 50.000 – Rp. 100.000 | 3 |
| Rp. 100.000 – Rp. 150.000 | 4 |
| Rp. 150.000 – Rp. 200.000 | 5 |
| Rp. 200.000 – Rp. 300.000 | 6 |
| Rp. 300.000 – Rp. 500.000 | 7 |
| > Rp. 500.000 | 8 |

Tabel 5. Transformasi Data Tagihan Listrik

| Tagihan Listrik | Skor |
|---------------------------|------|
| Rp. 0 | 1 |
| < Rp. 50.000 | 2 |
| Rp. 50.000 – Rp. 100.000 | 3 |
| Rp. 100.000 – Rp. 200.000 | 4 |
| Rp. 200.000 – Rp. 300.000 | 5 |
| Rp. 300.000 – Rp. 400.000 | 6 |
| Rp. 400.000 – Rp. 500.000 | 7 |
| > Rp. 500.000 | 8 |

Tabel 6. Transformasi Data Tagihan PDAM

| Tagihan PDAM | Skor |
|---------------------------|------|
| Rp. 0 | 1 |
| < Rp. 25.000 | 2 |
| Rp. 25.000 – Rp. 40.000 | 3 |
| Rp. 40.000 – Rp. 60.000 | 4 |
| Rp. 60.000 – Rp. 80.000 | 5 |
| Rp. 80.000 – Rp. 100.000 | 6 |
| Rp. 100.000 – Rp. 150.000 | 7 |
| > Rp. 150.000 | 8 |

Tabel 7. Transformasi Data Pulsa HP

| Pulsa HP | Skor |
|---------------------------|------|
| Rp.0 | 1 |
| < Rp. 40.000 | 2 |
| Rp. 40.000 – Rp. 60.000 | 3 |
| Rp. 60.000 – Rp. 80.000 | 4 |
| Rp. 80.000 – Rp. 100.000 | 5 |
| Rp. 100.000 – Rp. 150.000 | 6 |
| Rp. 150.000 – Rp. 250.000 | 7 |
| > Rp. 80.000 | 8 |

Tabel 8. Transformasi Data Langganan TV

| Langganan TV | Skor |
|---------------------------|------|
| Tidak Memiliki TV | 1 |
| Tidak Langganan/Punya TV | 2 |
| Rp. 0 - Rp. 100.000 | 3 |
| Rp. 100.000 – Rp. 150.000 | 4 |
| > Rp. 150.000 | 5 |

Tabel 9. Transformasi Data Pajak Mobil

| Pajak Mobil | Skor |
|-------------------------------|------|
| Rp. 0 | 1 |
| Rp. 1 – Rp. 250.000 | 2 |
| Rp. 250.000 – Rp. 750.000 | 3 |
| Rp. 750.000 – Rp. 1.250.000 | 4 |
| Rp. 1.250.000 – Rp. 2.000.000 | 5 |
| Rp. 2.000.000 – Rp. 2.500.000 | 6 |
| Rp. 2.500.000 – Rp. 3.000.000 | 7 |
| > Rp. 3.000.000 | 8 |

Tabel 10. Transformasi Data Pajak Motor

| Pajak Motor | Skor |
|---------------------------|------|
| Rp. 0 | 1 |
| Rp. 0 – Rp. 100.000 | 2 |
| Rp. 100.000 – Rp. 200.000 | 3 |
| Rp. 200.000 – Rp. 300.000 | 4 |
| Rp. 300.000 – Rp. 400.000 | 5 |
| Rp. 400.000 – Rp. 500.000 | 6 |
| Rp. 500.000 – Rp. 600.000 | 7 |
| > Rp. 600.000 | 8 |

Setelah semua data penelitian dirubah berdasarkan tabel tranformasi di atas, maka di peroleh hasil transformasi data seperti berikut:

Tabel 11. Data Hasil Transformasi

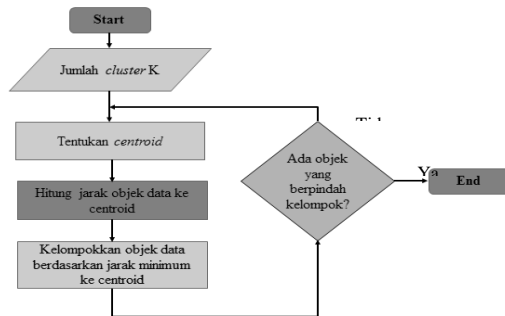
| Id. Mhs | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 |
|---------|----|----|----|----|----|----|----|----|----|-----|-----|-----|
| 1 | 9 | 2 | 1 | 5 | 2 | 6 | 8 | 8 | 2 | 4 | 3 | 2 |
| 2 | 5 | 3 | 1 | 5 | 4 | 5 | 7 | 2 | 2 | 8 | 4 | 1 |
| 3 | 5 | 4 | 6 | 5 | 2 | 4 | 2 | 3 | 2 | 5 | 4 | 1 |
| 4 | 8 | 3 | 1 | 4 | 2 | 5 | 2 | 7 | 3 | 1 | 5 | 1 |
| 5 | 8 | 3 | 1 | 4 | 2 | 5 | 2 | 7 | 3 | 1 | 5 | 1 |
| 6 | 3 | 4 | 1 | 5 | 3 | 8 | 1 | 8 | 2 | 1 | 4 | 1 |
| 7 | 6 | 2 | 1 | 1 | 2 | 3 | 6 | 7 | 2 | 1 | 4 | 1 |
| 8 | 7 | 2 | 1 | 5 | 2 | 4 | 5 | 2 | 2 | 1 | 3 | 2 |
| 9 | 3 | 2 | 1 | 1 | 5 | 8 | 2 | 4 | 2 | 1 | 4 | 1 |
| 10 | 2 | 5 | 1 | 5 | 2 | 4 | 2 | 2 | 2 | 1 | 4 | 1 |
| 11 | 5 | 3 | 5 | 5 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| 12 | 3 | 1 | 1 | 5 | 2 | 6 | 2 | 2 | 2 | 1 | 4 | 1 |
| 13 | 2 | 2 | 1 | 5 | 3 | 5 | 1 | 5 | 2 | 1 | 1 | 1 |
| 14 | 2 | 2 | 1 | 3 | 3 | 5 | 2 | 2 | 2 | 1 | 4 | 2 |
| 15 | 2 | 5 | 1 | 5 | 2 | 4 | 1 | 3 | 2 | 1 | 1 | 1 |
| 16 | 1 | 3 | 3 | 2 | 1 | 4 | 4 | 2 | 2 | 1 | 3 | 1 |
| 17 | 1 | 5 | 1 | 5 | 2 | 5 | 2 | 2 | 2 | 1 | 1 | 1 |
| 18 | 5 | 1 | 5 | 5 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 |
| 19 | 3 | 1 | 1 | 5 | 2 | 4 | 1 | 3 | 2 | 1 | 3 | 1 |
| 20 | 2 | 2 | 1 | 5 | 2 | 4 | 2 | 2 | 2 | 1 | 3 | 1 |
| 21 | 2 | 4 | 2 | 5 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| 22 | 2 | 3 | 1 | 1 | 2 | 3 | 2 | 2 | 2 | 1 | 5 | 1 |
| 23 | 2 | 1 | 1 | 5 | 2 | 3 | 2 | 2 | 3 | 1 | 1 | 1 |

| | | | | | | | | | | | | |
|----|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | 3 | 2 | 1 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 3 | 1 |
| 25 | 3 | 1 | 1 | 3 | 1 | 4 | 1 | 4 | 3 | 1 | 1 | 1 |
| 26 | 3 | 3 | 1 | 1 | 2 | 3 | 2 | 2 | 2 | 1 | 3 | 1 |
| 27 | 2 | 1 | 1 | 1 | 2 | 4 | 2 | 2 | 1 | 1 | 3 | 1 |
| 28 | 3 | 3 | 1 | 1 | 3 | 1 | 1 | 3 | 2 | 1 | 1 | 1 |
| 29 | 3 | 1 | 1 | 1 | 2 | 3 | 1 | 2 | 2 | 1 | 1 | 1 |
| 30 | 3 | 2 | 1 | 1 | 2 | 3 | 1 | 1 | 2 | 1 | 1 | 1 |

Berdasarkan tabel di atas, data pusat *cluster* awal C0 di ambil dari data mahasiswa dengan Id_Mahasiswa 11 pada Tabel. 12, C1 dari Id_Mahasiswa 8, C2 dari Id_Mahasiswa 6, C3 dari Id_Mahasiswa 16 dan C4 dari Id_Mahasiswa 17.

5. Proses Algoritma *K-Means* Clustering

Proses pengelompokan data menggunakan algoritma *K-Means Clustering* ini dilakukan dengan langkah-langkah seperti *flow chart* berikut:



Gambar 3. Flowcart *K-Means Clustering*

Pada penelitian ini ditentukan jumlah *cluster* sebanyak 5 kelompok. Jumlah *cluster* tersebut ditetapkan berdasarkan Peraturan Menristek Dikti No. 22 tahun 2015. Pengelompokan UKT di Universitas Negeri Padang terdiri dari 5 kategori [8]. Kategori I adalah UKT dengan besaran 500.000, kategori II UKT dengan besaran 1.000.000, kategori III UKT dengan besaran 2.000.000, kategori IV UKT dengan besaran 3.000.000 dan kategori V UKT dengan besaran 4.000.000. Pengelompokan UKT tersebut ditetapkan berdasarkan tingkat kondisi ekonomi orang tua calon mahasiswa baru.

5.1 Menentukan Pusat *Cluster* Awal Iterasi-1

Iterasi-1:

Untuk menentukan pusat *cluster* awal dilakukan dengan pemilihan data yang ingin dijadikan *centroid* awal. Karena pengelompokan besaran UKT yang telah ditentukan sebelumnya adalah sebanyak 5 kelompok, maka di alokasikan juga titik pusat *cluster* (*centroid*) sebanyak 5 data yang diambil dari data calon mahasiswa baru hasil transformasi pada Tabel 2. Pusat *cluster* yang dipilih sebanyak 5 data diberi nama C0, C1, C2, C3, C4 dengan data anggota/atributnya terdiri dari X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12 seperti terlihat pada Tabel berikut:

Table 12. Data Pusat *Cluster* Awal

| Pusat Cluster Awal | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 |
|--------------------|----|----|----|----|----|----|----|----|----|-----|-----|-----|
| C0 | 5 | 3 | 5 | 5 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| C1 | 7 | 2 | 1 | 5 | 2 | 4 | 5 | 2 | 2 | 1 | 3 | 2 |
| C2 | 3 | 4 | 1 | 5 | 3 | 8 | 1 | 8 | 2 | 1 | 4 | 1 |
| C3 | 1 | 3 | 3 | 2 | 1 | 4 | 4 | 2 | 2 | 1 | 3 | 1 |
| C4 | 1 | 5 | 1 | 5 | 2 | 5 | 2 | 2 | 2 | 1 | 1 | 1 |

5.2 Menghitung Jarak Data ke *Centroid* pada Iterasi-1

Setelah pusat *cluster* ditentukan kemudian hitung jarak setiap data mahasiswa ke masing-masing pusat *cluster* (*centrid*) menggunakan rumus *Euclidean Distance*, mulai dari jarak data mahasiswa 1 sampai dengan jarak data mahasiswa 30.

1. Menghitung jarak data mahasiswa 1:

$$DC0 = \sqrt{(9-5)^2 + (2-3)^2 + (1-5)^2 + (5-5)^2 + (2-2)^2 + (6-2)^2 + (8-2)^2 + (8-2)^2 + (2-2)^2 + (4-1)^2 + (3-1)^2 + (2-1)^2} = 11,619$$

$$DC1 = \sqrt{(9-7)^2 + (2-2)^2 + (1-1)^2 + (5-5)^2 + (2-2)^2 + (6-4)^2 + (8-5)^2 + (8-2)^2 + (2-2)^2 + (4-1)^2 + (3-3)^2 + (2-2)^2} = 7,874$$

$$DC2 = \sqrt{(9-3)^2 + (2-4)^2 + (1-1)^2 + (5-5)^2 + (2-3)^2 + (6-8)^2 + (8-1)^2 + (8-8)^2 + (2-2)^2 + (4-1)^2 + (3-4)^2 + (2-1)^2} = 10,247$$

$$DC3 = \sqrt{(9-1)^2 + (2-3)^2 + (1-3)^2 + (5-2)^2 + (2-1)^2 + (6-4)^2 + (8-4)^2 + (8-2)^2 + (2-2)^2 + (4-1)^2 + (3-3)^2 + (2-1)^2} = 12,042$$

$$DC4 = \sqrt{(9-1)^2 + (2-5)^2 + (1-1)^2 + (5-5)^2 + (2-2)^2 + (6-5)^2 + (8-2)^2 + (8-2)^2 + (2-2)^2 + (4-1)^2 + (3-1)^2 + (2-1)^2} = 12,649$$

Berdasarkan perhitungan jarak di atas, maka mahasiswa 1 jarak terdekat/terkecilnya terdapat pada DC1, sehingga mahasiswa 1 dimasukkan pada kelompok/*cluster* C1.

2. Menghitung jarak data mahasiswa 2:

Menghitung jarak data mahasiswa 2 ini sama caranya dengan menghitung jarak data mahasiswa 1 di atas. Begitulah seterusnya proses menghitung jarak sampai dengan data mahasiswa 30.

Setelah selesai melakukan proses perhitungan semua jarak data mahasiswa ke pusat *cluster* seperti proses di atas, maka diperoleh hasil perhitungan jarak yang di masukkan pada tabel berikut:

Tabel 13. Hasil Perhitungan Jarak Data

| Id Mhs | DC0 | DC1 | DC2 | DC3 | DC4 |
|--------|--------|--------|--------|--------|--------|
| 1 | 11,619 | 7,874 | 10,247 | 12,042 | 12,649 |
| 2 | 10,583 | 9,849 | 11,662 | 9,899 | 10,344 |
| 3 | 5,657 | 9,644 | 9,381 | 7,616 | 8,307 |
| 4 | 8,775 | 7,211 | 6,403 | 9,644 | 9,798 |
| 5 | 8,775 | 7,211 | 6,403 | 9,644 | 9,798 |
| 6 | 10,198 | 11,790 | 0,000 | 9,165 | 7,810 |
| 7 | 9,220 | 8,367 | 7,937 | 9,000 | 10,198 |
| 8 | 8,426 | 0,000 | 11,790 | 10,817 | 11,136 |
| 9 | 9,747 | 10,954 | 6,403 | 7,141 | 7,746 |
| 10 | 6,481 | 10,050 | 7,483 | 4,899 | 3,317 |
| 11 | 0,000 | 8,426 | 10,198 | 6,481 | 6,708 |
| 12 | 7,000 | 8,944 | 7,141 | 5,568 | 5,477 |
| 13 | 6,782 | 10,630 | 5,657 | 6,481 | 4,583 |
| 14 | 7,071 | 9,849 | 7,483 | 4,243 | 5,000 |

| | | | | | |
|----|-------|--------|--------|-------|-------|
| 15 | 5,916 | 10,583 | 7,280 | 5,745 | 2,000 |
| 16 | 6,481 | 10,817 | 9,165 | 0,000 | 5,196 |
| 17 | 6,708 | 11,136 | 7,810 | 5,196 | 0,000 |
| 18 | 2,449 | 9,110 | 11,136 | 7,483 | 8,062 |
| 19 | 5,831 | 9,110 | 7,211 | 5,657 | 5,196 |
| 20 | 5,831 | 9,539 | 7,746 | 4,472 | 3,873 |
| 21 | 4,472 | 10,536 | 9,165 | 5,477 | 3,606 |
| 22 | 7,616 | 10,630 | 9,055 | 4,000 | 6,403 |
| 23 | 5,568 | 9,899 | 9,110 | 5,385 | 4,690 |
| 24 | 6,633 | 9,950 | 9,055 | 4,899 | 6,245 |
| 25 | 6,245 | 9,798 | 7,681 | 5,568 | 5,657 |
| 26 | 6,403 | 9,592 | 9,000 | 3,873 | 5,657 |
| 27 | 7,348 | 10,440 | 9,055 | 4,000 | 6,245 |
| 28 | 6,325 | 10,630 | 10,000 | 6,000 | 6,557 |
| 29 | 6,481 | 10,149 | 9,798 | 5,292 | 6,403 |
| 30 | 6,325 | 10,149 | 10,198 | 5,099 | 5,916 |

| | | | | | | |
|----|-------|--------|--------|-------|-------|----|
| 29 | 6,481 | 10,149 | 9,798 | 5,292 | 6,403 | C3 |
| 30 | 6,325 | 10,149 | 10,198 | 5,099 | 5,916 | C3 |

Dari hasil pengelompokan pada Tabel di atas, maka pengelompokan data pada iterasi-1 adalah:

Tabel 15. Hasil Pengelompokan Data

| Kelompok | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 |
|----------|----|----|----|----|----|----|----|----|----|-----|-----|-----|
| C0 | 5 | 4 | 6 | 5 | 2 | 4 | 2 | 3 | 2 | 5 | 4 | 1 |
| | 5 | 3 | 5 | 5 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| | 5 | 1 | 5 | 5 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 |
| C1 | 9 | 2 | 1 | 5 | 2 | 6 | 8 | 8 | 2 | 4 | 3 | 2 |
| | 5 | 3 | 1 | 5 | 4 | 5 | 7 | 2 | 2 | 8 | 4 | 1 |
| | 7 | 2 | 1 | 5 | 2 | 4 | 5 | 2 | 2 | 1 | 3 | 2 |
| C2 | 8 | 3 | 1 | 4 | 2 | 5 | 2 | 7 | 3 | 1 | 5 | 1 |
| | 8 | 3 | 1 | 4 | 2 | 5 | 2 | 7 | 3 | 1 | 5 | 1 |
| | 3 | 4 | 1 | 5 | 3 | 8 | 1 | 8 | 2 | 1 | 4 | 1 |
| C3 | 6 | 2 | 1 | 1 | 2 | 3 | 6 | 7 | 2 | 1 | 4 | 1 |
| | 3 | 2 | 1 | 1 | 5 | 8 | 2 | 4 | 2 | 1 | 4 | 1 |
| | 2 | 2 | 1 | 3 | 3 | 5 | 2 | 2 | 2 | 1 | 4 | 2 |
| C4 | 1 | 3 | 3 | 2 | 1 | 4 | 4 | 2 | 2 | 1 | 3 | 1 |
| | 2 | 3 | 1 | 1 | 2 | 3 | 2 | 2 | 2 | 1 | 5 | 1 |
| | 3 | 2 | 1 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 3 | 1 |
| C4 | 3 | 1 | 1 | 3 | 1 | 4 | 1 | 4 | 3 | 1 | 1 | 1 |
| | 3 | 3 | 1 | 1 | 2 | 3 | 2 | 2 | 2 | 1 | 3 | 1 |
| | 2 | 1 | 1 | 1 | 2 | 4 | 2 | 2 | 1 | 1 | 3 | 1 |
| C4 | 3 | 3 | 1 | 1 | 3 | 1 | 1 | 3 | 2 | 1 | 1 | 1 |
| | 3 | 1 | 1 | 1 | 2 | 3 | 1 | 2 | 2 | 1 | 1 | 1 |
| | 3 | 2 | 1 | 1 | 2 | 3 | 1 | 1 | 2 | 1 | 1 | 1 |
| C4 | 2 | 5 | 1 | 5 | 2 | 4 | 2 | 2 | 2 | 1 | 4 | 1 |
| | 3 | 1 | 1 | 5 | 2 | 6 | 2 | 2 | 2 | 1 | 4 | 1 |
| | 2 | 2 | 1 | 5 | 3 | 5 | 1 | 5 | 2 | 1 | 1 | 1 |
| C4 | 2 | 5 | 1 | 5 | 2 | 4 | 1 | 3 | 2 | 1 | 1 | 1 |
| | 1 | 5 | 1 | 5 | 2 | 5 | 2 | 2 | 2 | 1 | 1 | 1 |
| | 3 | 1 | 1 | 5 | 2 | 4 | 1 | 3 | 2 | 1 | 3 | 1 |
| C4 | 2 | 2 | 1 | 5 | 2 | 4 | 2 | 2 | 2 | 1 | 3 | 1 |
| | 2 | 4 | 2 | 5 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| | 2 | 1 | 1 | 5 | 2 | 3 | 2 | 2 | 3 | 1 | 1 | 1 |

5.3 Mengelompokkan Data Kedalam Cluster

Kemudian berdasarkan perhitungan jarak (*Euclidean Distance*) pada Tabel. 13, dilakukan penentuan kelompok berdasarkan hasil perhitungan jarak terdekat/nilai minimum. Pada perhitungan jarak Mahasiswa 1 sebelumnya, jarak minimim terdapat pada DC1 maka Mahasiswa 1 masuk kedalam kelompok C1. Begitu proses selanjutnya untuk Mahasiswa 2 sampai dengan Mahasiswa 30. Setelah dilakukan pengelompokan semua data diperoleh hasil sebagai berikut:

Tabel 14. Pengalokasian Data ke Centroid

| Id. Mhs | DC0 | DC1 | DC2 | DC3 | DC4 | Cluser |
|---------|--------|--------|--------|--------|--------|--------|
| 1 | 11,619 | 7,874 | 10,247 | 12,042 | 12,649 | C1 |
| 2 | 10,583 | 9,849 | 11,662 | 9,899 | 10,344 | C1 |
| 3 | 5,657 | 9,644 | 9,381 | 7,616 | 8,307 | C0 |
| 4 | 8,775 | 7,211 | 6,403 | 9,644 | 9,798 | C2 |
| 5 | 8,775 | 7,211 | 6,403 | 9,644 | 9,798 | C2 |
| 6 | 10,198 | 11,790 | 0,000 | 9,165 | 7,810 | C2 |
| 7 | 9,220 | 8,367 | 7,937 | 9,000 | 10,198 | C2 |
| 8 | 8,426 | 0,000 | 11,790 | 10,817 | 11,136 | C1 |
| 9 | 9,747 | 10,954 | 6,403 | 7,141 | 7,746 | C2 |
| 10 | 6,481 | 10,050 | 7,483 | 4,899 | 3,317 | C4 |
| 11 | 0,000 | 8,426 | 10,198 | 6,481 | 6,708 | C0 |
| 12 | 7,000 | 8,944 | 7,141 | 5,568 | 5,477 | C4 |
| 13 | 6,782 | 10,630 | 5,657 | 6,481 | 4,583 | C4 |
| 14 | 7,071 | 9,849 | 7,483 | 4,243 | 5,000 | C3 |
| 15 | 5,916 | 10,583 | 7,280 | 5,745 | 2,000 | C4 |
| 16 | 6,481 | 10,817 | 9,165 | 0,000 | 5,196 | C3 |
| 17 | 6,708 | 11,136 | 7,810 | 5,196 | 0,000 | C4 |
| 18 | 2,449 | 9,110 | 11,136 | 7,483 | 8,062 | C0 |
| 19 | 5,831 | 9,110 | 7,211 | 5,657 | 5,196 | C4 |
| 20 | 5,831 | 9,539 | 7,746 | 4,472 | 3,873 | C4 |
| 21 | 4,472 | 10,536 | 9,165 | 5,477 | 3,606 | C4 |
| 22 | 7,616 | 10,630 | 9,055 | 4,000 | 6,403 | C3 |
| 23 | 5,568 | 9,899 | 9,110 | 5,385 | 4,690 | C4 |
| 25 | 6,633 | 9,950 | 9,055 | 4,899 | 6,245 | C3 |
| 24 | 6,245 | 9,798 | 7,681 | 5,568 | 5,657 | C3 |
| 26 | 6,403 | 9,592 | 9,000 | 3,873 | 5,657 | C3 |
| 27 | 7,348 | 10,440 | 9,055 | 4,000 | 6,245 | C3 |
| 28 | 6,325 | 10,630 | 10,000 | 6,000 | 6,557 | C3 |

5.4 Menghitung Pusat Cluster Baru Iterasi-2

Selanjutnya hitung pusat *cluster* baru untuk iterasi-2.

Iterasi-2:

Menghitung pusat *cluster* baru C0:

$$C0 (X1) = \frac{5+5+5}{3} = 5,000$$

$$(X2) = \frac{4+3+1}{3} = 2,667$$

$$(X3) = \frac{6+5+5}{3} = 5,333$$

demikian seterusnya sampai menghitung (X12)

Menghitung pusat *cluster* baru C1:

$$C1 (X1) = \frac{9+5+7}{3} = 7,000$$

$$(X2) = \frac{2+3+2}{3} = 2,333$$

$$(X3) = \frac{1+1+1}{3} = 1,000$$

demikian seterusnya sampai menghitung (X12)

Menghitung pusat *cluster* baru C2:

$$C2 (X1) = \frac{8+8+3+3}{4} = 5,600$$

$$(X2) = \frac{3+3+4+2}{4} = 2,800$$

$$(X3) = \frac{1+1+1+1}{4} = 1,000$$

demikian seterusnya sampai menghitung (X12)
Menghitung pusat *cluster* baru C3:

$$C3 (X1) = \frac{6+2+1+2+3+3+3+2+3+3+3}{11} = 2,500$$

$$(X2) = \frac{2+2+3+3+2+1+3+1+3+1+2}{11} = 2,100$$

$$(X3) = \frac{1+1+3+1+1+1+1+1+1+1+1}{11} = 1,200$$

demikian seterusnya sampai (X12)
Menghitung pusat *cluster* baru C4:

$$C4 (X1) = \frac{2+3+2+2+1+3+2+2+2}{9} = 2,111$$

$$(X2) = \frac{5+1+2+5+5+1+2+4+1}{9} = 2,889$$

$$(X3) = \frac{1+1+1+1+1+1+1+2+1}{9} = 1,111$$

demikian seterusnya sampai menghitung (X12).
Setelah dilakukan proses perhitungan semua pusat *cluster* baru, maka diperoleh hasil pusat *cluster* baru yang dapat dilihat pada Tabel 16.

Tabel 16. Hasil Pusat *Cluster* Baru

| Pusat cluster baru | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 |
|--------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| C0 | 5,00 | 2,66 | 5,33 | 5,00 | 2,00 | 2,33 | 1,66 | 2,33 | 2,00 | 2,33 | 2,00 | 1,00 |
| C1 | 7,00 | 2,33 | 1,00 | 5,00 | 2,66 | 5,00 | 6,66 | 4,00 | 2,00 | 4,33 | 3,33 | 1,66 |
| C2 | 5,60 | 2,80 | 1,00 | 3,00 | 2,80 | 5,80 | 2,60 | 6,60 | 2,40 | 1,00 | 4,40 | 1,00 |
| C3 | 2,50 | 2,10 | 1,20 | 1,50 | 2,10 | 3,30 | 1,70 | 2,20 | 2,00 | 1,00 | 2,50 | 1,10 |
| C4 | 2,11 | 2,88 | 1,11 | 5,00 | 2,11 | 4,11 | 1,55 | 2,55 | 2,11 | 1,00 | 2,11 | 1,00 |
| | 1 | 9 | 1 | 0 | 1 | 1 | 6 | 6 | 1 | 0 | 1 | 0 |

5.5 Menghitung Jarak Data ke *Centroid* pada Iterasi-2
Pada tahap ini kembali dihitung jarak setiap data mahasiswa ke masing-masing pusat *cluster (centrid)* menggunakan rumus *Euclidean Distance*. Perhitungan dimulai dari jarak data mahasiswa 1 sampai dengan jarak data mahasiswa 30.

Menghitung jarak data mahasiswa 1

$$DC0 = \sqrt{(9-5,000)^2 + (2-2,667)^2 + (1-5,333)^2 + (5-5,000)^2 + (2-2,000)^2 + (6-2,333)^2 + (8-1,667)^2 + (8-2,333)^2 + (2-2,000)^2 + (4-2,333)^2 + (3-2,000)^2 + (2-1,000)^2} = 11,210$$

$$DC1 = \sqrt{(9-7,000)^2 + (2-2,333)^2 + (1-1,000)^2 + (5-5,000)^2 + (2-2,667)^2 + (6-5,000)^2 + (8-6,667)^2 + (8-4,000)^2 + (2-2,000)^2 + (4-4,333)^2 + (3-3,333)^2 + (2-1,667)^2} = 4,485$$

$$DC2 = \sqrt{(9-5,600)^2 + (2-2,800)^2 + (1-1,000)^2 + (5-3,000)^2 + (2-2,800)^2 + (6-5,1,000)^2 + (3-4,400)^2 + (2-1,000)^2} = 7,754$$

$$DC3 = \sqrt{(9-2,500)^2 + (2-2,100)^2 + (1-1,200)^2 + (5-1,500)^2 + (2-2,100)^2 + (6-3,300)^2 + (8-1,700)^2 + (8-2,200)^2 + (2-2,000)^2 + (4-2,200)^2 + (3-2,000)^2 + (2-1,000)^2} = 12,052$$

$$(8-1,700)^2 + (8-2,200)^2 + (2-2,000)^2 + (4-1,000)^2 + (3-2,500)^2 + (2-1,100)^2 = 12,052$$

$$DC4 = \sqrt{(9-2,111)^2 + (2-2,889)^2 + (1-1,111)^2 + (5-5,000)^2 + (2-2,111)^2 + (6-4,111)^2 + (8-1,556)^2 + (8-2,556)^2 + (2-2,111)^2 + (4-1,000)^2 + (3-2,111)^2 + (2-1,000)^2} = 11,568$$

Berdasarkan perhitungan jarak di atas, maka mahasiswa 1 jarak terdekat/terkecilnya terdapat pada DC1, sehingga mahasiswa 1 dimasukkan pada kelompok/*cluster* C1.

5.6 Menghitung jarak data mahasiswa 2:

Menghitung jarak data mahasiswa 2 ini sama caranya dengan menghitung jarak pada data mahasiswa 1 di atas. Begitulah seterusnya proses menghitung jarak sampai dengan data mahasiswa 30.

Setelah selesai dilakukan proses perhitungan semua jarak data mahasiswa ke pusat *cluster* seperti proses di atas, maka diperoleh hasil perhitungan jarak yang di masukkan pada Tabel berikut:

Tabel 17. Hasil Perhitungan Jarak Data

| Id. Mhs | DC0 | DC1 | DC2 | DC3 | DC4 |
|---------|--------|-------|-------|--------|--------|
| 1 | 11,210 | 4,865 | 7,754 | 12,052 | 11,568 |
| 2 | 9,730 | 4,967 | 9,814 | 10,259 | 9,760 |
| 3 | 4,082 | 7,572 | 8,020 | 8,040 | 7,313 |
| 4 | 8,347 | 7,000 | 3,053 | 8,405 | 8,099 |
| 5 | 8,347 | 7,000 | 3,053 | 8,405 | 8,099 |
| 6 | 9,798 | 9,381 | 4,682 | 8,686 | 7,175 |
| 7 | 9,110 | 6,557 | 5,032 | 7,513 | 8,733 |
| 8 | 6,403 | 4,435 | 6,366 | 6,711 | 6,221 |
| 9 | 9,434 | 8,963 | 5,340 | 6,053 | 6,850 |
| 10 | 6,481 | 8,446 | 6,893 | 4,883 | 2,931 |
| 11 | 1,826 | 8,524 | 8,290 | 6,151 | 5,450 |
| 12 | 6,708 | 7,550 | 6,043 | 4,841 | 3,469 |
| 13 | 6,831 | 8,679 | 5,926 | 5,181 | 3,151 |
| 14 | 6,880 | 8,165 | 6,093 | 3,072 | 3,388 |
| 15 | 6,298 | 9,147 | 7,275 | 4,984 | 2,502 |
| 16 | 6,532 | 8,718 | 7,611 | 3,720 | 4,706 |
| 17 | 7,095 | 9,292 | 8,032 | 5,314 | 2,874 |
| 18 | 2,828 | 9,557 | 9,095 | 6,499 | 6,212 |
| 19 | 5,657 | 8,000 | 5,960 | 3,955 | 2,388 |
| 20 | 5,831 | 8,000 | 6,717 | 3,666 | 1,466 |
| 21 | 5,033 | 9,557 | 8,362 | 4,609 | 2,893 |
| 22 | 7,439 | 9,274 | 6,893 | 2,800 | 5,114 |
| 23 | 5,916 | 8,660 | 7,882 | 4,152 | 2,715 |
| 24 | 6,633 | 9,165 | 6,717 | 1,497 | 4,587 |
| 25 | 6,351 | 8,699 | 6,142 | 3,527 | 3,747 |
| 26 | 6,506 | 8,622 | 6,551 | 1,356 | 4,401 |
| 27 | 7,303 | 9,092 | 7,037 | 1,908 | 4,706 |
| 28 | 6,733 | 9,933 | 7,818 | 3,292 | 5,389 |
| 29 | 6,733 | 9,557 | 7,611 | 2,154 | 4,846 |
| 30 | 6,683 | 9,730 | 8,094 | 2,200 | 4,776 |

5.7 Mengelompokkan Data Kedalam *Cluster*

Selanjutnya berdasarkan perhitungan jarak (*Euclidean Distance*) pada Tabel. 17, lakukan penentuan kelompok berdasarkan hasil perhitungan jarak terdekat/nilai minimum. Pada perhitungan jarak Mahasiswa 1 sebelumnya, perhitungan jarak (DC0,DC1,DC2,DC3, DC4) dimana jarak minimum terdapat pada DC1 maka Mahasiswa 1 masuk kedalam kelompok C1. Begitu

proses selanjutnya untuk Mahasiswa 2 sampai dengan Mahasiswa 30. Setelah dilakukan pengelompokan semua data diperoleh hasil seperti terlihat pada tabel 18.

Selanjutnya lihat dan bandingkan hasil pengelompokan iterasi-2. Dari perbandingan pengelompokan iterasi-2 dengan iterasi-1, kelompok dan anggota kelompok *cluster* C0,C1,C2,C3 dan C4 hasilnya sudah sama. Dengan demikian proses algoritma *K-Means Clustering* dihentikan pada iterasi-2 karena tidak ada lagi terjadi perubahan/perpindahan anggota kelompok objek antara iterasi-2 dengan iterasi-1.

Tabel 18. Hasil Pengelompokan Data

| Id. Mh | DC0 | DC1 | DC2 | DC3 | DC4 | DC4 | Cluser |
|--------|--------|-------|-------|--------|--------|--------|--------|
| 1 | 11,210 | 4,865 | 7,754 | 12,052 | 11,568 | 11,568 | C1 |
| 2 | 9,730 | 4,967 | 9,814 | 10,259 | 9,760 | 9,760 | C1 |
| 3 | 4,082 | 7,572 | 8,020 | 8,040 | 7,313 | 7,313 | C0 |
| 4 | 8,347 | 7,000 | 3,053 | 8,405 | 8,099 | 8,099 | C2 |
| 5 | 8,347 | 7,000 | 3,053 | 8,405 | 8,099 | 8,099 | C2 |
| 6 | 9,798 | 9,381 | 4,682 | 8,686 | 7,175 | 7,175 | C2 |
| 7 | 9,110 | 6,557 | 5,032 | 7,513 | 8,733 | 8,733 | C2 |
| 8 | 6,403 | 4,435 | 6,366 | 6,711 | 6,221 | 6,221 | C1 |
| 9 | 9,434 | 8,963 | 5,340 | 6,053 | 6,850 | 6,850 | C2 |
| 10 | 6,481 | 8,446 | 6,893 | 4,883 | 2,931 | 2,931 | C4 |
| 11 | 1,826 | 8,524 | 8,290 | 6,151 | 5,450 | 5,450 | C0 |
| 12 | 6,708 | 7,550 | 6,043 | 4,841 | 3,469 | 3,469 | C4 |
| 13 | 6,831 | 8,679 | 5,926 | 5,181 | 3,151 | 3,151 | C4 |
| 14 | 6,880 | 8,165 | 6,093 | 3,072 | 3,388 | 3,388 | C3 |
| 15 | 6,298 | 9,147 | 7,275 | 4,984 | 2,502 | 2,502 | C4 |
| 16 | 6,532 | 8,718 | 7,611 | 3,720 | 4,706 | 4,706 | C3 |
| 17 | 7,095 | 9,292 | 8,032 | 5,314 | 2,874 | 2,874 | C4 |
| 18 | 2,828 | 9,557 | 9,095 | 6,499 | 6,212 | 6,212 | C0 |
| 19 | 5,657 | 8,000 | 5,960 | 3,955 | 2,388 | 2,388 | C4 |
| 20 | 5,831 | 8,000 | 6,717 | 3,666 | 1,466 | 1,466 | C4 |
| 21 | 5,033 | 9,557 | 8,362 | 4,609 | 2,893 | 2,893 | C4 |
| 22 | 7,439 | 9,274 | 6,893 | 2,800 | 5,114 | 5,114 | C3 |
| 23 | 5,916 | 8,660 | 7,882 | 4,152 | 2,715 | 2,715 | C4 |
| 24 | 6,633 | 9,165 | 6,717 | 1,497 | 4,587 | 4,587 | C3 |
| 25 | 6,351 | 8,699 | 6,142 | 3,527 | 3,747 | 3,747 | C3 |
| 26 | 6,506 | 8,622 | 6,551 | 1,356 | 4,401 | 4,401 | C3 |
| 27 | 7,303 | 9,092 | 7,037 | 1,908 | 4,706 | 4,706 | C3 |
| 28 | 6,733 | 9,933 | 7,818 | 3,292 | 5,389 | 5,389 | C3 |
| 29 | 6,733 | 9,557 | 7,611 | 2,154 | 4,846 | 4,846 | C3 |
| 30 | 6,683 | 9,730 | 8,094 | 2,200 | 4,776 | 4,776 | C3 |

5.8 Proses Evaluasi Hasil

Berdasarkan Proses *K-Means Clustering* yang telah dilakukan diperoleh pengetahuan bahwa terdapat 5 kelompok besaran UKT calon mahasiswa baru yang terdiri dari:

1. *Cluster* C0 adalah kelompok UKT kategori V dengan besaran biaya 4.000.000,00 terdiri dari mahasiswa 3,11,18 sebanyak 3 orang
2. *Cluster* C1 adalah kelompok UKT kategori VI dengan besaran biaya 3.000.000,00 terdiri dari mahasiswa 1,2,8 sebanyak 3 orang
3. *Cluster* C2 adalah kelompok UKT kategori III dengan besaran biaya 2.000.000,00 terdiri dari mahasiswa 4,5,6,7,9 sebanyak 5 orang
4. *Cluster* C3 adalah kelompok UKT kategori II dengan besaran biaya 1.000.000,00 terdiri dari mahasiswa 14,16,22,24,25,26,27,28,29,30 sebanyak 10 orang

5. *Cluster* C4 adalah kelompok UKT kategori I dengan besaran biaya 500.000,00 terdiri dari mahasiswa 10,12,13,15,17,19,20,21,23 sebanyak 9 orang.

Dengan demikian algoritma *K-Means Clustering* ini telah mampu mengelompokkan besaran UKT calon mahasiswa baru ke dalam 5 kelompok berdasarkan tingkatan kondisi sosial ekonomi orang tua. Pengelompokan inilah yang digunakan sebagai dasar untuk membantu pengambilan keputusan dalam menentukan besaran UKT calon mahasiswa baru di Universitas Negeri Padang.

4. Kesimpulan

Setelah dilakukan penelitian *data mining* dengan metode *K-Means Clustering* untuk menentukan kelompok besaran UKT calon mahasiswa baru studi kasus di Universitas Negeri Padang, maka dapat diambil kesimpulan bahwa :

1. Dari proses *K-Means Clustering* yang telah dilakukan, pengelompokan besaran UKT calon mahasiswa baru terdiri dari 5 kelompok yaitu; kelompok UKT kategori 1 dengan besaran UKT 500.000,00, kelompok UKT kategori 2 dengan besaran UKT 1.000.000,00, kelompok UKT kategori 3 dengan besaran UKT 2.000.000,00, kelompok UKT kategori 4 dengan besaran UKT 3.000.000,00, kelompok UKT kategori 5 dengan besaran UKT 4.000.000,00.
2. Penerapan *data mining* dengan menggunakan metode *K-Means Clustering* dapat membantu dalam menentukan besaran UKT calon mahasiswa baru.

Saran

Setelah melakukan penelitian, penulis memberikan beberapa saran antara lain:

1. Penulis menyarankan adanya penelitian *data mining* dalam menentukan kelompok besaran UKT ini dengan membandingkan secara langsung beberapa metode *clustering*, guna menguji metode yang paling tepat digunakan.
2. Penulis mengharapkan sistem *data mining* dengan metode *K-Means Clustering* ini dapat diimplementasikan berupa aplikasi agar dapat dimanfaatkan langsung oleh Perguruan Tinggi.

Daftar Rujukan

- [1] Aranda, J., dan Natasya, W.A.G., 2016. Penerapan Metode K-Means Cluster Pada Sistem Pendukung Keputusan Pemilihan Konsentrasi Untuk Mahasiswa International Class STMIK AMIKOM Yogyakarta, *Seminar Nasional Teknologi Informasi dan Multimedia 2016 STMIK AMIKOM Yogyakarta*. Yogyakarta, Indonesia 6-7 Februari 2016. STMIK AMIKOM: Yogyakarta.
- [2] Larose D, T., 2005. *Discovering knowledge in data: an introduction to data mining*, Jhon Wiley & Sons Inc.
- [3] Metisen, B. M., dan Sari, H. L., 2015. Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokkan

- Penjualan Produk Pada Swalayan Fadhila, *Jurnal Media Infotama*. Vol. 11 No. 2, September 2015.
- [4] Pramadhani, A. E., dan Setiadi, T., 2014. Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit ISPA (Infeksi Saluran Pernapasan Akut) Dengan Algoritma Decision Tree (ID3), *Jurnal Sarjana Teknik Informatika*. Volume 2 Nomor 1, Februari 2014.
- [5] Republik Indonesia., 2012. Undang Undang Republik Indonesia. No.12. tentang Pendidikan Tinggi. *Sekretariat Negara*. Jakarta.
- [6] Republik Indonesia., 2013. Permenag No.12. Tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal Bagi Mahasiswa Baru Pada Perguruan Tinggi Negeri Agama RI Tahun Akademik 2013/2014. *Sekretariat Negara*. Jakarta.
- [7] Republik Indonesia., 2013. Permendikbud No.55 Tahun 2013 tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal Pada Perguruan Tinggi Negeri di Lingkungan Kemdikbud. *Menteri Pendidikan dan Kebudayaan*. Jakarta.
- [8] Republik Indonesia., 2015. Permenristek Dikti No.22 Tahun 2015 tentang Biaya Kuliah Tunggal dan Uang Kuliah Tunggal Pada Perguruan Tinggi Negeri di Lingkungan Kemristek Dikti. *Menteri Pendidikan dan Kebudayaan*. Jakarta.
- [9] Riano, K., dan Sedyono, E., (2013). Penentuan Alih Fungsi Lahan Marginal Menjadi Lahan Pangan Berbasis Algoritma K-means di Wilayah Kabupaten Boyolali. *Jdc*. 2. 20
- [10] Setiawan, R., 2016. Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru, *Jurnal Lentera ICT*. Vol.3 No.1.
- [11] Singh, M., dan Bansal, M., 2015. A Survey on Various K-Means algorithms for Clustering, *IJCSNS*. VOL.15 No.6, June 2015.
- [12] Taslim., dan Fajrijal., 2016. Penerapan Algorithma K-Mean untuk Clustering Data Obat pada Puskesmas Rumbai, *Jurnal Teknologi Informasi & Komunikasi Digital Zone*. Vol. 7, No. 2, November 2016:108-114.
- [13] Turban., 2005. Decision Support Systems and Intelligent Systems 7th Ed. New Jersey. *Pearson Education*. No.5 (2013), pp. 241-266