

# Klastering Perekonomian Mahasiswa Politeknik Negeri Bali Menggunakan K-Means

Putu Indah Ciptayani\*, I G N B Caturbawa, I Nyoman Eddy Indrayana  
Politeknik Negeri Bali  
Email: putuindah@pnb.ac.id\*

## Abstrak

Politeknik Negeri Bali (PNB) adalah salah satu sekolah vokasi di Indonesia yang menggunakan uang kuliah tunggal (UKT). UKT merupakan biaya yang dibayarkan mahasiswa berdasarkan pada kemampuan ekonominya. Di PNB, mahasiswa memilih level UKT yang mereka inginkan berdasarkan pada kondisi ekonominya, dan kemudian PNB akan meninjau kembali kondisi ekonomi mahasiswa untuk menentukan apakah mahasiswa tersebut memperoleh UKT sesuai dengan level yang mereka pilih atau tidak. Proses ini tentu saja membutuhkan waktu yang lama dan tambahan sumber daya manusia, demikian juga akurasi yang tinggi sangat dibutuhkan guna menghindari kesalahan dalam pemberian UKT untuk mahasiswa. Penelitian ini melakukan klastering berdasarkan pada kondisi ekonomi mahasiswa menggunakan metode K-Means. Dalam penelitian ini akan diperkenalkan sebuah metode untuk pemilihan centroid awal guna mendapatkan hasil klaster yang terbaik. Adapun klaster yang dihasilkan dapat digunakan untuk menganalisis kondisi ekonomi mahasiswa PNB sehingga nantinya pihak manajemen dapat menentukan kuota untuk setiap level UKT. Kesimpulan dari penelitian ini adalah kebanyakan mahasiswa berada dalam ekonomi rendah dan beberapa pencilon data berada pada klaster ekonomi sangat tinggi. Hasil penelitian ini dapat dipergunakan sebagai data untuk pembuatan sistem pendukung keputusan penentuan UKT.

**Kata kunci:** k-means, klastering, pemilihan centroid, klastering perekonomian mahasiswa

## Abstract

Bali State Polytechnic is one of the vocational schools in Indonesia that use Single tuition fee (UKT). UKT is a student paid tuition based on the ability of its economy. In Bali State Polytechnic, student choose the level of UKT they want based on their economic condition, and then Bali State Polytechnic will review student's economic condition to define whether the students will get the UKT they've been chosen or not. This process requires a lot of time and human resource, as well as high accuracy to avoid mistakes in determining the right UKT for students. This research did student clustering based on their economic condition using k-means. In this research we introduce the custom algorithm to select initial centroid to get best cluster result. The conclusion of this study is that most students are in poor economy and a few outliers of data are at a very high economic cluster. The result of this cluster can be used to analyze the economic condition of Bali State Polytechnic in order to define the quota of every UKT level. For next usage, this result can be used as data for UKT decision support system.

**Keywords:** k-means, clustering, centroid selection, student's economy clustering

## 1. Pendahuluan

Penentuan biaya kuliah untuk universitas di Indonesia telah diatur dalam Peraturan Menteri Pendidikan dan Kebudayaan No. 55, tahun 2013. Dalam aturan tersebut terdapat pengaturan uang kuliah tunggal. Uang Kuliah Tunggal (UKT) merupakan biaya yang dibayar mahasiswa berdasarkan pada kemampuan ekonominya. Dalam Peraturan Menteri Pendidikan dan Kebudayaan No. 73 tahun 2014, terdapat pengaturan UKT untuk setiap universitas negeri di Indonesia, yang mana salah satunya adalah Politeknik Negeri Bali (PNB). UKT di PNB dibagi menjadi lima level. Dalam penentuan level UKT untuk mahasiswa, tentu saja akan dilakukan peninjauan terhadap kondisi ekonomi mahasiswa. Proses ini dibutuhkan untuk menentukan level UKT yang akan diterima mahasiswa.

Di PNB, penentuan UKT mahasiswa dilakukan dengan memberikan data perekonomian orang tua mahasiswa seperti penghasilan ayah, penghasilan ibu, kebutuhan bulanan, dan jumlah tanggungan

keluarga. Selanjutnya, mahasiswa akan memilih level UKT yang diinginkannya. Setelah semua data terkumpul, maka dilakukan peninjauan terhadap kondisi ekonomi mahasiswa untuk menentukan apakah mahasiswa akan mendapatkan UKT sesuai pilihannya atau akan dipindahkan ke level lainnya yang lebih sesuai. Tentu saja, proses ini akan membutuhkan waktu yang cukup panjang dan tingkat akurasi yang cukup tinggi dalam membandingkan kondisi ekonomi untuk setiap mahasiswa. Kesalahan dalam peninjauan akan menyebabkan kesalahan dalam administrasi level UKT. Salah satu alternatif yang bisa dilakukan untuk menghindari kesalahan tersebut adalah dengan melakukan klastering berdasarkan pada kondisi ekonomi mahasiswa. Dalam penelitian ini, klastering mahasiswa dilakukan berdasarkan pada kondisi ekonominya dengan menggunakan metode K-Means. Proses pengelompokan sekumpulan objek baik fisik maupun abstrak, ke dalam kelas yang sama disebut dengan klastering. Sebuah klaster pada dasarnya merupakan sekumpulan objek data yang memiliki kesamaan satu sama lain dan ketidaksamaan dengan objek lain di luar klasternya [1].

Klastering dibedakan menjadi dua yaitu, hirarki dan non-hirarki. Demikian juga dengan analisis klaster dapat dibedakan menjadi dua yaitu teknik klastering hirarki dan non-hirarki. Contoh dari teknik hirarki yaitu *single linkage*, *complete linkage*, *average linkage*, *median*, dan *Ward*. Sedangkan teknik non-hirarki misalnya *k-means clustering*, *adaptive k-means*, *k-medoids*, dan *fuzzy clustering* [2]. Metode klastering K-means telah digunakan secara luas dan menjadi metode yang populer untuk melakukan klastering. Metode ini sangat sederhana dan cepat. Oyelade dkk [2] menggunakan metode K-Means untuk memprediksi kemampuan akademis mahasiswa dengan melakukan klastering mahasiswa berdasarkan pada hasil ujiannya. Untuk menghitung nilai kesamaan diantara objek digunakan *Euclidean distance*. Semakin kecil jarak maka semakin besar derajat kesamaan objek. Dubey dkk [3] melakukan deteksi keruakaan pada apel menggunakan K-Means, dimana K-Means digunakan sebagai langkah awal dalam klastering berdasarkan pada warna dan fitur spasialnya. Shovon [4] menggunakan klastering K-Means dan *decision tree* untuk memprediksi IPK, dimana K-Means digunakan untuk membagi mahasiswa berdasarkan pada nilai evaluasi sebelum ujian akhir dilakukan. Hasil dari K-Means tergantung pada pemilihan centroid inisial. Untuk metode K-Means klasik, centroid inisial dipilih secara acak, sehingga hasil dari setiap percobaan akan berbeda [5]. Pemilihan centroid awal sangatlah penting, sehingga beberapa penelitian melakukan riset untuk memilih centroid awal, seperti pada [5,6].

Dalam penelitian ini, digunakan klastering K-Means untuk melakukan klaster pada kondisi ekonomi mahasiswa berdasarkan pada total penghasilan orang tua, pengeluaran bulanan dan jumlah tanggungan keluarga. Untuk setiap kelompok yang dibentuk oleh pengguna akan dipilih sebuah centroid secara otomatis. Dalam penelitian ini digunakan algoritma tersendiri untuk memilih centroid dengan cara mengurutkan data secara menaik, dan kemudian memilih data yang memiliki posisi pada kelipatan dari jumlah data dibagi dengan total klaster. Dengan menggunakan algoritma ini, klaster yang terbentuk akan selalu sama pada setiap percobaan dan centroid awal diharapkan merupakan centroid yang terbaik. *Euclidean distance* digunakan untuk mengukur jarak antara mahasiswa dengan centroid awal. Jarak terpendek akan menjadi klaster yang ditempati oleh mahasiswa tersebut. Nilai centroid yang baru akan dihitung berdasarkan pada anggota baru setiap klaster. Proses ini akan dilakukan secara berulang sampai setiap centroid tidak mengalami perubahan nilai. Karena sederhana dan cepatnya proses ini, maka hasil akan diperoleh dalam waktu yang singkat. Pihak manajemen tingkat atas dapat menggunakan hasil dari klastering ini untuk membuat keputusan dalam rangka menentukan level UKT setiap mahasiswa. Di samping itu, hasil penelitian ini juga dapat digunakan sebagai data untuk sistem pendukung keputusan berbasis komputer.

## 2. Metode Penelitian

### 2.1 Metode Klastering K-Means

Klastering merupakan proses pengelompokan sekumpulan objek abstrak maupun fisik ke dalam kelas-kelas yang sama. Semua data di dalam sebuah klaster memiliki persamaan yang besar dan perbedaan yang besar dengan objek dalam klaster lainnya. Meskipun klasifikasi adalah cara yang efektif untuk membedakan kelompok atau kelas obyek, namun klasifikasi membutuhkan waktu yang lebih lama dan pelabelan dari sekelompok besar data atau pola pelatihan Kadangkala akan lebih cepat apabila dilakkan dengan cara sebaliknya. Partisi pertama yaitu mengelompokkan data ke dalam kelompok berdasarkan kesamaan (misalnya, menggunakan klastering), dan kemudian menetapkan label grup dalam jumlah yang relatif kecil [1].

Algoritma K-Means memiliki parameter input berupa jumlah klaster dan data yang ingin diklaster. *Klaster similarity* adalah ukuran untuk rerata nilai objek di dalam klaster, yang dapat dilihat pada centroid klaster tersebut. Centroid dapat dikatakan sebagai nilai tengah dari suatu klaster. Di bawah ini adalah algoritma klastering K-Means [1] :

**Input:**

k: jumlah kluster,  
D: data dari n objek

**Output:** Sekumpulan kluster.

**Metode:**

- (1) Pilih k objek secara acak dari D sebagai centroid;
- (2) Ulangi
- (3) Tempatkan setiap objek pada kluster yang memiliki kesamaan terbesar dengan centroidnya;
- (4) Perbarui centroid, misalnya dengan melakukan perhitungan rata-rata nilai objek di dalam kluster;
- (5) Sampai tidak ada perubahan;

Metode K-Means merupakan salah satu teknik klustering yang populer. Metode ini sangat sederhana dan cepat. K-Means juga dikenal sebagai teknik *centroid-based*. Kompleksitas algoritma adalah  $O(nkt)$ , dimana n adalah jumlah data, k adalah jumlah kluster dan t adalah jumlah iterasi. Kualitas dari klustering akhir dari algoritma K-Means sangat tergantung pada pemilihan acak centroid awal. Dalam algoritma K-Means asli, nilai centroid awal dipilih secara random dan akhirnya didapatkan kluster yang berbeda untuk setiap percobaan meskipun datanya sama [7].

Terdapat beberapa keuntungan metode K-Means [8] :

1. Algoritma K-Means adalah algoritma klasik untuk menyelesaikan masalah kluster; algoritma ini relatif sederhana dan cepat.
2. Untuk pengumpulan data yang besar, algoritma ini relatif fleksibel dengan efisien tinggi, karena kompleksitas adalah  $O(nkt)$ , dengan n adalah kali iterasi, k adalah jumlah kluster, t adalah jumlah iterasi.
3. Menyediakan hasil yang relatif baik untuk kluster cembung.

*Euclidean distance* digunakan untuk mengukur ketidaksamaan. Semakin besar jaraknya maka semakin tinggi pula tingkat perbedaanya. *Euclidean distance* ditunjukkan pada Persamaan 1

$$d(X,Y) = ((x_1+y_1)^2+(x_2+y_2)^2+\dots+(x_n+y_n)^2)^{0.5} \tag{1}$$

dimana X dan Y adalah vektor dengan n elemen.

**2.2 Research Design**

**2.2.1. Tempat dan Waktu Penelitian**

Penelitian ini dilakukan di Politeknik Negeri Bali. Penelitian ini dilakukan selama 3 bulan yaitu dimulai dari Juli 2015 sampai dengan September 2015.

**2.2.2. Sumber Data**

Adapun sumber data dalam penelitian ini berasal dari database Unit Management Information System (MIS) Politeknik Negeri Bali. Jumlah mahasiswa yang dikluster dalam penelitian ini adalah 411 mahasiswa untuk semua jurusan. Data mahasiswa yang digunakan adalah data mahasiswa yang diterima pada tahun ajaran 2013/2014.

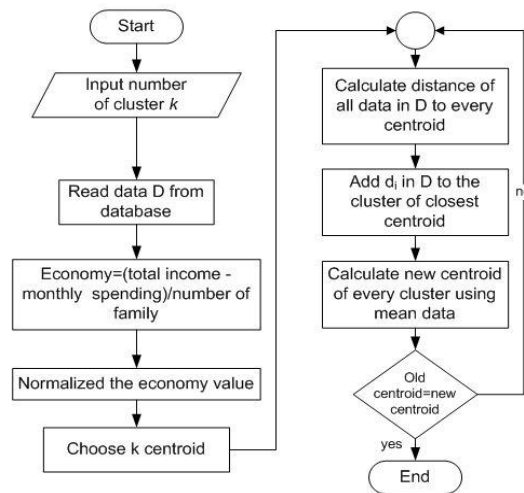
**2.2.3. Alur Sistem**

Penelitian dilakukan dengan menggunakan K-Means untuk mengkluster mahasiswa PNB berdasarkan pada kondisi ekonomi mereka. Parameter yang digunakan untuk menentukan kondisi ekonomi mahasiswa adalah total penghasilan orang tua, pengeluaran bulanan dan jumlah tanggungan keluarga. Adapun kondisi ekonomi mahasiswa akan dihitung dengan Persamaan 2 :

$$\text{Economy} = (\text{total pendapatan-pengeluaran bulanan})/\text{jumlah tanggungan} \tag{2}$$

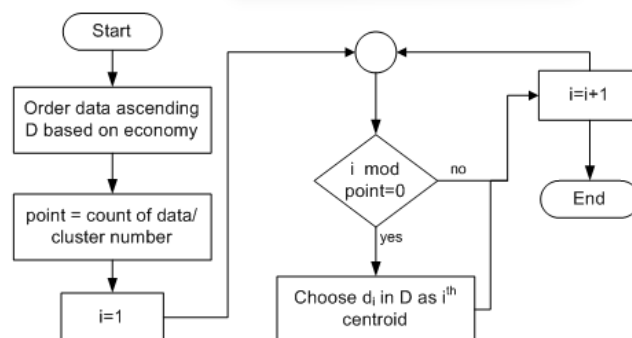
Flowchart design sistem ditunjukkan pada Gambar 1. Pengguna sistem akan menentukan k yaitu jumlah kluster yang diinginkan. Kemudian sistem akan membaca semua data mahasiswa dari database. Sebelum menuju ke proses berikutnya akan dilakukan perhitungan kondisi ekonomi setiap mahasiswa menggunakan Persamaan 2. Setelah semua kondisi ekonomi dihitung, maka dilakukan normalisasi agar nilai ekonomi menjadi dalam interval 0 hingga 1. Hal ini akan memudahkan pemrosesan selanjutnya. Langkah berikutnya adalah pemilihan centroid awal dari setiap kluster. Pemilihan centroid akan dilakukan dengan menggunakan algoritma yang ditunjukkan pada Gambar 2. Langkah berikutnya adalah menghitung jarak dari setiap data dengan setiap centroid. Perhitungan jarak dilakukan dengan menggunakan *Euclidean distance* pada Persamaan 1. Setiap data akan dipindahkan ke dalam kluster baru,

yaitu kluster yang memiliki jarak centroid terkecil dengan data tersebut. Jika jarak semua data dengan setiap centroid sudah dihitung dan data telah menempati kluster yang paling sesuai dengannya, maka akan dilakukan proses perhitungan ulang centroid. Perhitungan ulang centroid akan dilakukan dengan menghitung rata-rata data dalam kluster tersebut. Jika terdapat nilai centroid yang berbeda dengan nilai sebelumnya, maka klastering dianggap belum konvergen dan belum menemui kluster yang maksimal. Dengan demikian, proses klastering harus dilangi kembali dengan menghitung jarak setiap data di seluruh kluster dengan masing-masing centroid yang baru. Proses ini akan dilakukan berulang kali sampai tidak ada perubahan pada setiap centroid. Jika setiap centroid kluster telah memiliki nilai yang sama, maka proses klastering sudah konvergen dan dapat dihentikan. Kluster-kluster inilah yang dianggap sebagai hasil akhir dari algoritma klastering.



Gambar 1. Flowchart Sistem Klastering

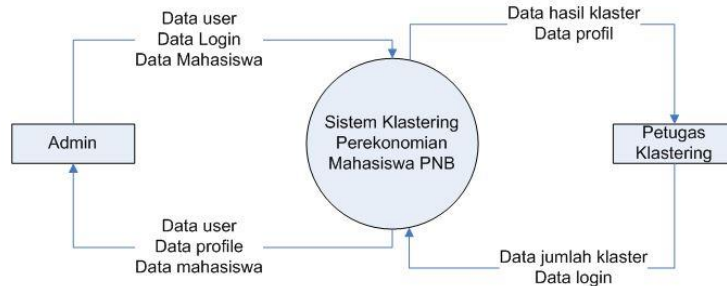
Pemilihan centroid dalam penelitian ini tidak menggunakan teknik acak. Pemilihan centroid secara acak akan memberikan hasil kluster yang berbeda pada setiap percobaan. Dengan teknik pemilihan centroid ini, maka hasil dari setiap percobaan akan konsisten setiap kali program klastering dijalankan. Adapun flowchart pemilihan centroid yang dilakukan dalam penelitian ini ditunjukkan pada Gambar 2. Data yang menjadi input dalam algoritma ini adalah data dengan yang nilai ekonominya sudah dihitung dengan Persamaan 2. Sebelum proses pemilihan centroid dilakukan maka, semua data D akan diurutkan secara menaik berdasarkan pada kondisi ekonominya. Langkah berikutnya adalah menghitung sebuah nilai bernama point dengan melakukan pembagian jumlah data dibagi dengan jumlah kluster atau  $n/k$ . Nilai ini digunakan untuk menentukan titik yang akan dijadikan centroid. Data yang akan menjadi centroid adalah data yang posisinya berada pada kelipatan dari variabel point. Dengan demikian, centroid awal merupakan nilai tengah dari setiap kluster, dan centroid ini akan menjadi pusat dari setiap data yang paling sesuai dengan klasternya. Untuk mengetahui apakah suatu data menempati posisi sesuai dengan kelipatan point, akan digunakan operator aritmatika modulo. Jika nomor urut dapat dibagi oleh variabel point, maka berarti data tersebut menempati posisi pada kelipatan variabel point.



Gambar 2. Flowchart Pemilihan Centroid

**3. Hasil dan Pembahasan**  
**3.1. Perancangan Sistem**

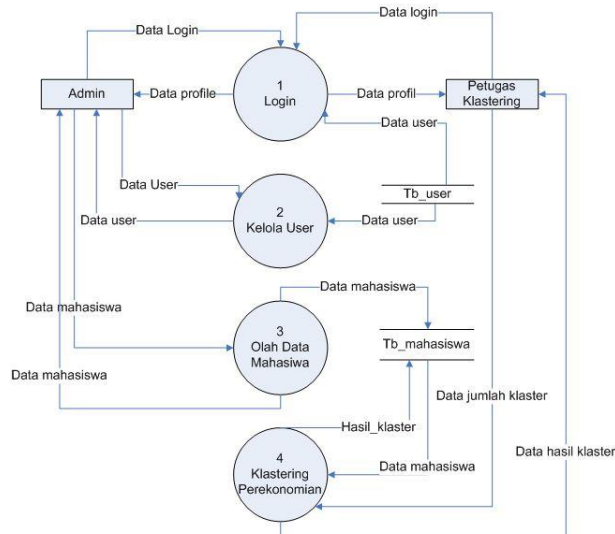
Perancangan sistem pada penelitian ini dilakukan dengan diagram konteks, diagram aliran data dan perancangan tabel. Diagram konteks ditunjukkan pada Gambar 3, sedangkan diagram aliran data ditunjukkan pada Gambar 4.



Gambar 3. Diagram Konteks Sistem Klustering Perekonomian Mahasiswa

Dari diagram konteks pada Gambar 3, terlihat ada dua pengguna sistem, yaitu admin dan petugas klustering. Admin akan mengelola data pengguna sistem, mengolah data master mahasiswa. Sedangkan petugas klustering dapat melakukan klustering perekonomian mahasiswa dan melihat hasilnya. Baik admin maupun petugas klustering, harus login terlebih dahulu untuk dapat melakukan akses data ke dalam sistem. Kedua jenis hak akses dapat melakukan akses terhadap profilnya masing-masing.

Diagram alir data ditunjukkan pada Gambar 4. Terdapat empat proses utama di dalam sistem klustering ini, yaitu login, pengelolaan pengguna, pengolahan data mahasiswa, dan klustering perekonomian. Proses login dilakukan oleh kedua pengguna di dalam sistem. Dalam proses login, kedua pengguna melakukan input data loginnya masing-masing, dan sistem akan melakukan pengambilan data di dalam penyimpanan data bernama tb\_user. Penyimpanan data tb\_user menyimpan data pengguna sistem beserta hak akses yang dimilikinya. Proses kedua yaitu pengelolaan data pengguna hanya dapat dilakukan oleh admin sistem. Pengelolaan data pengguna dalam hal ini adalah penambahan pengguna baru dalam sistem. Adapun setiap pengguna baru akan disimpan ke dalam penyimpanan data tb\_user. Proses ketiga adalah proses pengelolaan data mahasiswa. Pengolahan data mahasiswa hanya dapat dilakukan oleh admin sistem. Pengolahan data mahasiswa meliputi penyisipan, pengubahan dan penghapusan data mahasiswa. Semua data mahasiswa akan disimpan ke dalam penyimpanan data tb\_mahasiswa. Proses terakhir yaitu proses klustering perekonomian, yang merupakan inti dari sistem ini. Proses ini dapat dilakukan oleh petugas klustering, dengan melakukan input jumlah kluster terlebih dahulu. Sistem akan melakukan pembacaan data seluruh mahasiswa pada penyimpanan data tb\_mahasiswa. Setelah sistem melakukan proses klustering sesuai dengan algoritma yang sudah dijelaskan pada Gambar 1 dan 2, maka hasil klustering akan disimpan kembali sebagai perubahan data ke dalam penyimpanan data tb\_mahasiswa. Hasil klustering ini juga akan diberikan kepada petugas klustering.



Gambar 4. Diagram Alir Data Sistem Klustering Perekonomian Mahasiswa

Dalam sistem klustering perekonomian ini terdapat dua buah tabel, seperti yang terlihat pada diagram alir data. Kedua tabel tersebut adalah tabel pengguna dan tabel mahasiswa. Adapun struktur tabel pengguna ditunjukkan pada Tabel 1, sedangkan tabel mahasiswa ditunjukkan pada Tabel 2. Tabel tb\_pengguna terdiri dari empat kolom, dimana username sekaligus dijadikan sebagai kunci primer tabel. Password pengguna di dalam sistem ini menggunakan fungsi hash MD5. Kolom hak akses menyimpan hak setiap pengguna, yaitu sebagai admin atau petugas klustering. Kolom hak akses inilah yang akan digunakan sebagai pengecekan untuk menampilkan menu ketika pengguna melakukan login ke dalam sistem. Tabel mahasiswa terdiri dari 15 kolom. Kunci primer dari tabel tersebut adalah kolom NIM. Kolom jincome merupakan atribut turunan yang menyimpan penjumlahan dari kolom inc\_ibu dan inc\_ayah. Alasan penyimpanan atribut turunan ini adalah untuk meningkatkan kecepatan dalam proses klustering. Netincome didapatkan dari pengurangan antara jincome dengan costbulanan. Kolom ekonomi didapatkan dengan menerapkan Persamaan 2 yaitu membagi netincome dengan jtanggungan. Ekonomi\_norm merupakan hasil normalisasi dari kolom ekonomi. Kolom kluster akan menyimpan hasil kluster dari setiap mahasiswa. Kolom ini akan diupdate setelah dilakukan proses klustering oleh petugas.

Tabel 1. Tb\_Pengguna

Field	Tipe Data	Keterangan
username	VARCHAR(30)	Primary Key
nama	VARCHAR(50)	
password	CHAR(32)	MD5 Hash Function
hak_akses	ENUM('Admin', 'Petugas')	

Tabel 2. Tb\_Mahasiswa

Field	Tipe Data	Keterangan
tahun	INT(11)	
namaprodi	VARCHAR(40)	
NIM	VARCHAR(10)	Primary Key
namamhs	VARCHAR(50)	
kdprod	VARCHAR(7)	
nmsekolah	VARCHAR(40)	
inc_ibu	INT(11)	
inc_ayah	INT(11)	
jincome	FLOAT	
costbulanan	FLOAT	
netincome	INT(11)	
jtanggungan	FLOAT	
ekonomi	FLOAT	
ekonomi_norm	FLOAT	
kluster	TINYINT(4)	

### 3.2. Implementasi Sistem

Implementasi sistem dilakukan dengan program berbasis website menggunakan Bahasa PHP dan pengolah basis data MySQL. Adapun Gambar 5 menunjukkan halaman klustering. Petugas akan memasukkan jumlah kluster yang diinginkan, kemudian, sistem akan melakukan klustering dan menampilkan hasilnya berupa centroid inisial, kemudian jumlah iterasi yang dilakukan hingga hasil centroid konvergen, centroid akhir dan tabel ringkasan klustering. Tabel ringkasan klustering dalam halaman tersebut menyajikan informasi jumlah anggota masing-masing kluster, nilai terkecil dan terbesar dari setiap kluster.

### 3.3. Hasil Pengujian Sistem

Metode K-Means diaplikasikan pada 411 data mahasiswa di PNB. Percobaan dilakukan sebanyak tiga kali. Percobaan pertama mengelompokkan data menjadi tiga kluster, percobaan kedua mengelompokkan data menjadi empat kluster dan percobaan terakhir mengelompokkan data menjadi 5

klaster. Tabel 3, 4 dan 5 menampilkan hasil dari klastering tersebut secara berturut-turut. Semua centroid dipilih dengan algoritma yang ditunjukkan pada Gambar 2. Untuk setiap percobaan, akan diambil jumlah anggota klaster, perekonomian minimum dan maksimum, baik itu nilai sebenarnya ataupun nilai yang telah dinormalisasi.

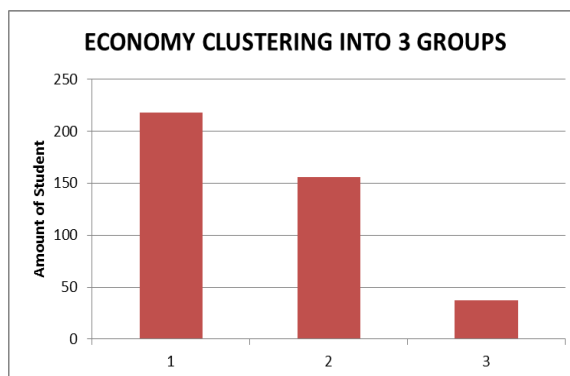


Gambar 5. Halaman Klastering Perekonomian

Tabel 3. Hasil Perobaan K-means Pertama

Klaster	Jumlah Data	Min	Min norm	Max	Max norm
1	218	-1,300,000	0	500,000	0.245455
2	156	512,500	0.247159	2.088.720	0.462098
3	37	2.162.020	0.472094	6,033,330	1

Nilai centroid awal untuk percobaan pertama adalah 0.186, 0.293 dan 1. Setelah melakukan 9 kali iterasi, maka semua centroid konvergen ke dalam nilai 0.168, 0.324, dan 0.606. Klaster pertama memiliki anggota terbanyak, sedangkan klaster terakhir memiliki anggota paling sedikit. Gambar 6 menampilkan grafik untuk jumlah anggota pada masing-masing klaster. Klaster 1 untuk ekonomi lemah, klaster 2 untuk ekonomi sedang, dan klaster 3 untuk ekonomi tinggi. Dari Gambar 6, klaster yang memiliki anggota terbanyak adalah klaster pertama dan klaster dengan anggota paling sedikit adalah klaster ketiga. Grafik tersebut menunjukkan bahwa kebanyakan mahasiswa PNB berada pada ekonomi menengah atau rendah. Hanya 37 mahasiswa yang memiliki perekonomian yang tinggi jika dilihat dalam Tabel 3.



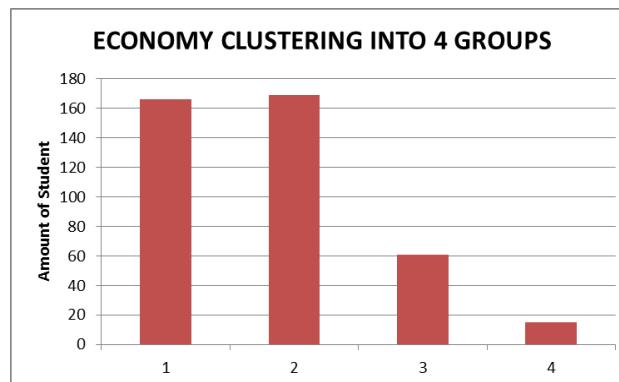
Gambar 6. Grafik Kluster Mahasiswa Ke Dalam Tiga Kelompok

Pada percobaan kedua, nilai centroid untuk setiap kluster adalah 0.164, 0.236, 0.320455 dan 0.773. Nilai centroid tetap pada 0.149, 0.277, 0.446 dan 0.736 setelah 10 iterasi. Berdasarkan pada Tabel 4, kebanyakan mahasiswa berada pada kluster kedua (ekonomi rendah) dengan anggota 169 dan kluster dengan anggota paling sedikit adalah kluster terakhir dengan 15 anggota. Berdasarkan Gambar 7, jumlah anggota kluster pertama (ekonomi sangat rendah) dan kluster kedua (ekonomi rendah) hampir sama. Dari hasil percobaan, dapat disimpulkan bahwa hasil percobaan masih sama dengan percobaan pertama yaitu kebanyakan mahasiswa PNB memiliki tingkat ekonomi menengah atau rendah.

Percobaan terakhir melakukan pengelompokan mahasiswa ke dalam lima kluster. Kluster pertama untuk ekonomi yang sangat, kluster kedua untuk ekonomi rendah, kluster ketiga untuk ekonomi menengah, kluster keempat untuk ekonomi tinggi, sedangkan kluster terakhir untuk ekonomi sangat tinggi. Nilai centroid awal dimulai dari kluster pertama adalah 0.151, 0.208, 0.264, 0.3375 dan 0.648. Pada Gambar 8, dapat dilihat bahwa kelompok dengan anggota terbanyak adalah kluster pertama. Kebanyakan mahasiswa berada pada kluster 1 dan 2. Kluster 3 masih memiliki jumlah yang cukup banyak jika dibandingkan dengan kluster 4 dan 5. Dari hasil percobaan ini dapat dilihat bahwa kebanyakan mahasiswa berada dalam kondisi ekonomi rendah dan perekonomian menengah dalam jumlah sedang. Nilai centroid konvergen pada 0.136, 0.226, 0.317, 0.464 dan 0.736 setelah 9 iterasi.

Tabel 4. Hasil K-means untuk Percobaan Kedua

Kluster	Jumlah Data	Min	Min norm	Max	Max norm
1	166	(1,300,000)	0	250,000	0.211364
2	169	275,000	0.214773	1,331,110	0.358788
3	61	1,350,000	0.361364	2,893,050	0.57178
4	15	3,222,380	0.616688	6,033,330	1



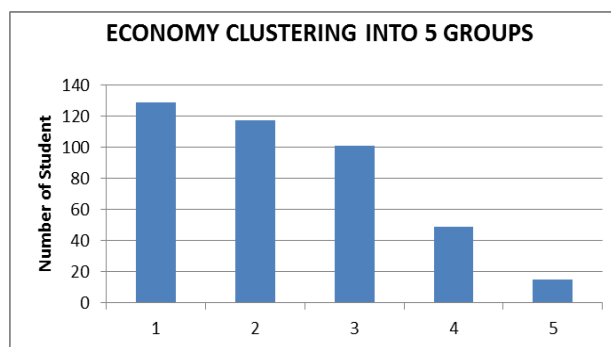
Gambar 7. Grafik Kluster Mahasiswa Ke Dalam Empat Kelompok

Tabel 5. Hasil K-means untuk Percobaan Ketiga

Kluster	Jumlah Data	Min	Min norm	Max	Max norm
1	129	(1,300,000)	0	0	0.177273
2	117	30,668	0.181455	666,667	0.268182
3	101	697,386	0.272371	1,524,040	0.385096
4	49	1,567,400	0.391009	2,893,050	0.57178
5	15	3,222,380	0.616688	6,033,330	1

Dari ketiga percobaan yang telah dilakukan, dapat dilihat bahwa kebanyakan mahasiswa berada pada ekonomi rendah. Jumlah mahasiswa dalam kluster ekonomi tinggi sangat sedikit. Rata-rata iterasi untuk setiap percobaan adalah 9 iterasi. Metode pemilihan centroid dalam Gambar 2, mampu menghasilkan hasil yang sama untuk setiap percobaan dengan jumlah kluster yang sama.





Gambar 8. Grafik Kluster Mahasiswa Ke Dalam Lima Kelompok

#### 4. Kesimpulan dan Saran

Metode K-means digunakan dalam penelitian untuk mengkluster 411 mahasiswa berdasarkan pada kondisi ekonominya. Dengan metode pemilihan centroid yang dibuat sendiri dalam penelitian ini, hasil dari setiap percobaan jumlah kluster yang sama akan menghasilkan kluster yang sama. Ini dapat mengatasi permasalahan perbedaan hasil yang dihasilkan oleh metode pemilihan centroid acak. Hasil dari percobaan ini adalah kebanyakan mahasiswa berada dalam ekonomi rendah. Jumlah mahasiswa dalam ekonomi sangat tinggi adalah sangat sedikit jika dibandingkan dengan jumlah semua mahasiswa. Beberapa pencilan data berada pada kluster ekonomi sangat tinggi. Kelemahan metode ini adalah bahwa metode ini mungkin berakhir pada optimum lokal. Untuk mengatasi hal ini, dibutuhkan metode heuristik seperti *genetic algorithm* yang dapat ditambahkan agar diraih optimum global.

Untuk penelitian selanjutnya, hasil kluster ini dapat digunakan untuk membangun sistem pendukung keputusan untuk menentukan level UKT bagi mahasiswa PNB, sehingga setiap mahasiswa akan memperoleh UKT terbaiknya. Dengan melakukan analisa terhadap hasil kluster ini, pihak manajemen tingkat atas PNB, akan memperoleh gambaran umum dari kondisi ekonomi mahasiswa. Di samping itu, mereka dapat menentukan kuota untuk setiap UKT.

#### Daftar Pustaka

- [1] Han, J., & Kamber, M. Data Mining: Concepts and Techniques. Soft Computing (Vol. 54). 2006
- [2] Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. Application of k Means Klustering algorithm for prediction of Students Academic Performance, 7, 292–295. 2010
- [3] Dubey, S. R., Dixit, P., Singh, N., & Gupta, J. P. Infected Fruit Part Detection using K-Means Klustering Segmentation Technique. International Journal of Interactive Multimedia and Artificial Intelligence, 2(2), 65. 2013
- [4] Shovon, H. I. An Approach of Improving Student 's Academic Performance by using K-means klustering algorithm and Decision tree, 3(8), 146–149. 2012
- [5] Baswade, A. M., Nalwade, P. S. Selection of Initial Centroids for k-Means Algorithm. International Journal of Computer Science and Mobile Computing . IJCSMC, Vol. 2, Issue. 7, July 2013, pg.161 – 164. 2013
- [6] Shakti, M., Thanamani, A., S. An Effective Determination of Initial Centroids in K-Means Klustering Using Kernel PCA, 2(3), 955-959. 2011
- [7] Yedla, M., Pathakota, S. R., Srinivasa, T. M. Enhancing K-means Klustering Algorithm with Improved Initial Center. 2010
- [8] Zhang, C., Fang, Z. An Improved K-means Klustering Algorithm. Journal of Information & Computational Science, 10(1), 193-199. 2013