

PEMODELAN *K-MEANS* PADA PENENTUAN PREDIKAT KELULUSAN MAHASISWA STMIK PALANGKARAYA

Lili Rusdiana dan Sam'ani

Teknik Informatika dan STMIK Palangkaraya

Jl. G.Obos No.114 Palangka Raya

Teknik Informatika dan STMIK Palangkaraya

Jl. G.Obos No.114 Palangka Raya

Email: sohib.sahabat@gmail.com , sam.stmikplk@gmail.com

ABSTRACT

Data mining by applying the cluster model can use the K-Means concept as a clustering method. K -Means model application can be used to classify data such as the title of graduation students based on the amount of load studies, GPA, and graduation thesis. This research was conducted to find out the results of the use of the concept of K-Means on the determination of student graduation in order to get the results in the form of an analysis of the modeling.

The problem of this research is to seek to model the concept of K-Means on the determination of graduation Students using Student Data of STMIK Palangkaraya. K-Means algorithm is used through three stages: initialization , the first iteration and the second iteration.

The results of the modeling is the use of K-Means algorithm to determine student graduation to obtain analysis of the results in the form of 70 % could determine the appropriate graduation of 10 data were used as a sample. Modeling using the K-Means is one of the concepts to be able to classify the data, so the use of K-Means algorithm can be a reference for the development of a modeling study , especially regarding data mining.

Keyword : Data Mining, K-Means, The Title Of Graduation

PENDAHULUAN

Salah satu penerapan cabang pengetahuan dalam pengambilan keputusan yaitu penggunaan dalam kumpulan data untuk mendapatkan informasi baru sebagai pendukung dalam pemecahan masalah seperti menggunakan pemodelan pada pengelompokan data. Pengelompokan data termasuk ke dalam *data mining* yaitu seperti analisis kelompok (*cluster*

analysis) yang mengelompokkan data-data ke dalam jumlah kelompok tertentu (*cluster*) berdasarkan karakteristik yang terdapat pada masing-masing data dalam kelompok yang ada. Data sebagai *input-output* dikelompokkan dalam *cluster* dan *output* yaitu berupa sebuah daftar *cluster* (Naba, 2009). *Data mining* digunakan dalam menelusuri kumpulan data agar dapat membangun

sebuah model untuk mengenali pola dalam pengelompokan data (Prasetyo, 2012).

Pengelompokan melakukan pemisahan, pemecahan, maupun segmentasi data ke dalam sejumlah kelompok (*cluster*) berdasarkan karakteristik. Pada *clustering* maka didapat hasil berupa keanggotaan setiap data pada setiap *cluster* dan dominan suatu data untuk masuk ke dalam suatu *cluster* tertentu.

Data mining dengan menerapkan model *cluster* dapat menggunakan konsep *K-Means* sebagai salah satu metode pengelompokan yang membagi data ke dalam dua kelompok atau lebih dengan tujuan untuk meminimalkan fungsi objektif yang telah diatur dalam proses pengelompokan (Prasetyo, 2012). Dalam pemodelan *K-Means*, data yang memiliki karakteristik sama akan dikelompokkan dalam kelompok yang sama dan data yang memiliki karakteristik yang berbeda akan masuk ke kelompok yang lainnya.

Penerapan model *K-Means* dapat digunakan untuk mengelompokkan data seperti predikat kelulusan Mahasiswa yang didapat berdasarkan jumlah beban studi yang telah ditempuh, IPK yang diperoleh Mahasiswa, dan kelulusan skripsi. Salah satu perguruan tinggi

yang menerapkan peraturan perhitungan jumlah beban studi, nilai IPK dan kelulusan skripsi sebagai identifikasi kelulusan untuk Mahasiswa yaitu STMIK Palangkaraya.

Adanya informasi yang perlu diketahui oleh Mahasiswa agar Mahasiswa dapat memperbaiki nilai IPK maupun menambah beban studi agar dapat memenuhi syarat untuk lulus. Data kelulusan Mahasiswa tersebut dapat digunakan sebagai sampel penerapan pemodelan *K-Means* untuk memperoleh hasil analisis penggunaan *K-Means*.

Maka dari itu, penulis bermaksud untuk menggunakan konsep yang ada pada *K-Means* sebagai model dalam pengelompokan data dari data acak Mahasiswa STMIK Palangkaraya sebagai sampel dalam penggunaan *K-Means* berdasarkan beban studi, jumlah IPK Mahasiswa, dan kelulusan skripsi. Berdasarkan latar belakang di atas, maka permasalahan yang akan dibahas dalam penulisan ini adalah bagaimana memodelkan konsep *K-Means* pada penentuan predikat kelulusan Mahasiswa berdasarkan beban studi, jumlah IPK, dan kelulusan skripsi?

Adapun tujuan Penelitian ini untuk mengetahui hasil dari penggunaan konsep *K-Means* pada penentuan

kelulusan Mahasiswa berdasarkan beban studi, jumlah IPK, dan kelulusan skripsi pada data Mahasiswa STMIK Palangkaraya.

Tinjauan Pustaka

Pengelompokan data terhadap suatu kumpulan data dapat dilakukan berdasarkan karakteristik tertentu untuk dilakukan analisis seperti penggunaan *K-Means* yang termasuk dalam analisis kelompok pada data mining.

Data mining mendukung proses dalam bisnis cerdas (*business intelligence*) karena *data mining* merupakan salah satu bidang yang terkait juga dengan gudang data (*data warehouse*) dalam bisnis cerdas. Selain bisnis cerdas, *data mining* juga memiliki kesamaan dengan bidang ilmu kecerdasan buatan (*Artificial Intelligence*), pengenalan pola (*pattern recognition*), dan pembelajaran mesin (*machine learning*) seperti algoritma pencarian, teknik pemodelan, dan teori pembelajaran (Prasetyo, 2012:4).

Teknik pengelompokan data sering diterapkan dalam berbagai bidang seperti model pengelompokan untuk mengantisipasi potensi pembelian berdasarkan perilaku pelanggan yang didasarkan pada analisis *clustering*.

Penelitian ini menggunakan data pembelian pada masa lalu untuk mengetahui perilaku pelanggan. Dengan mengumpulkan karakteristik kunci dari informasi pribadi pelanggan setia digunakan untuk mencari potensi pelanggan lain. Analisis aturan asosiasi mengekstrak pengetahuan tentang perilaku pembelian pelanggan setia, yang digunakan untuk mendeteksi dan menyaring potensi pelanggan berdasarkan loyalitas (chang, 2007).

Metode *K-Means* dalam hal *clustering* juga digunakan dalam pelacakan wajah yang telah terintegrasi dan pengenalan ekspresi wajah dalam sebuah sistem. Pelacakan dan pengenalan dilakukan pada video untuk menangkap ekspresi dari frame yang dihasilkan sehingga dapat dianalisis dengan beberapa tahapan yang dilakukan untuk dapat mengelompokkan data dalam video tersebut (Geetha, 2014).

Algoritma *K-Means* juga digunakan dalam metode ANFIS (*Adaptive Neuro Fuzzy Inference System*) untuk tahap inisialisasi pengacakan data untuk masuk dalam kelompok seperti pada penelitian yang menggunakan ANFIS dengan algoritma *K-Means* sebagai pendukung dalam inisialisasi data penentuan normalitas kehamilan.

Dihasilkan bahwa ANFIS dapat diimplementasikan untuk menentukan normalitas kehamilan. Penelitian ini dengan menggunakan 40 data sampel dan 16 data latih serta menggunakan pemrograman MATLAB untuk mengimplementasikan ANFIS. Diperoleh 97.5 % diagnosis sesuai dengan data latih yang digunakan (Rusdiana dkk , 2015).

Dasar Teori

Data Mining

Data mining merupakan salah satu disiplin ilmu yang digunakan untuk menentukan suatu informasi tertentu dalam sekumpulan data sebagai pendukung pengambilan keputusan. *Data mining* juga sebagai salah satu proses untuk memperoleh informasi yang memiliki nilai guna dari sekumpulan data (Tan, 2006).

Data mining merupakan proses pengekstrakan informasi dari jumlah kumpulan data dengan menggunakan algoritma dan tehnik gambar dari statistik, mesin pembelajaran dan sistem manajemen *database* (Han dan Kamber, 2011).

Menurut Prasetyo (2012), *data mining* terbagi menjadi beberapa kelompok yaitu sebagai berikut:

1. Model prediksi
Pemodelan yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang telah didapat.
2. Analisis kelompok
Mengelompokkan sekumpulan data ke dalam kelompok (*cluster*) tertentu berdasarkan karakteristik yang sama pada masing-masing data.
3. Analisis asosiasi
Menentukan pola yang menggambarkan kekuatan hubungan fitur dalam data.
4. Deteksi anomaly
Pengamatan suatu data dari sekumpulan data yang secara signifikan memiliki karakteristik yang berbeda dari sisa data yang lain. Terdapat beberapa jenis tipologi aplikasi *Data Mining*, yaitu :
 1. Metode prediksi
Memprediksi nilai yang akan datang berdasarkan data-data yang telah memiliki variabel seperti klasifikasi, regresi, detikasi anomali, dan lainnya.
 2. Metode deskriptif

Membantu *user* agar mudah melihat pola berdasarkan dari data yang telah ada.

Konsep Pengelompokan

Teknik pengelompokan melakukan pemisahan, pemecahan, maupun segmentasi data ke dalam sejumlah kelompok (*cluster*) berdasarkan karakteristik. Tujuan dari analisis kelompok yaitu untuk mengetahui informasi dan hubungan dari kesamaan data yang telah dikelompokkan berdasarkan karakteristik tertentu yang memiliki kemiripan atau hubungan satu sama lainnya dan jika yang berbeda atau tidak memiliki hubungan dengan lainnya dalam kelompok lainnya. Dalam konsep pengelompokan, diperlukan kemiripan yang lebih besar dalam kelompok dan perbedaan yang lebih besar di antara kelompok yang lainnya.

Tujuan dari pengelompokan yaitu untuk dapat membedakan pengelompokan untuk pemahaman dan pengelompokan untuk penggunaan. Dalam hal untuk pemahaman, kelompok yang terbentuk harus menangkap struktur alami data dan biasanya proses ini sebagai proses awal yang akan dilanjutkan dengan pekerjaan utama seperti peringkasan,

pelabelan pada setiap kelompok yang akan digunakan sebagai data latih klasifikasi dan sebagainya. Sedangkan dalam hal tujuan untuk penggunaan, agar pengelompokan yang dilakukan dapat digunakan untuk mencari *prototype* kelompok yang paling representative terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok pada sebuah data.

Metode K-Means

Ada banyak metode yang dapat digunakan dalam pengelompokan contohnya metode *K-Means*. Pengelompokan yang dapat digunakan seperti pengelompokan nonhierarki yang membagi data ke dalam bentuk dua atau lebih kelompok. *K-Means* merupakan metode analisis kelompok yang mengarah pada pembagian N objek pengamatan ke dalam K kelompok (*cluster*) dan setiap objek pengamatan dimiliki oleh suatu kelompok dengan rata-rata (*mean*) terdekat (prasetyo, 2012).

Pada *clustering* maka didapat hasil berupa keanggotaan setiap data pada setiap *cluster* dan dominan suatu data untuk masuk ke dalam suatu *cluster* tertentu.

Algoritma *K-Means* dalam pengelompokan data seperti berikut ini (prasetyo, 2012):

1. Tentukan jumlah kelompok.
2. Alokasikan data ke dalam kelompok secara acak.
3. Hitung pusat kelompok (rata-rata) dari data yang terdapat pada masing-masing kelompok.

Lokasi titik pusat pada setiap kelompok yang diambil dari rata-rata (*mean*) semua nilai data pada setiap fiturnya harus dihitung kembali. Untuk menghitung sentroid fitur ke-*i* dapat digunakan seperti pada persamaan 1.

$$C_i = \frac{1}{M} \sum_{i=1}^M x_i \dots\dots\dots (1)$$

Keterangan :

C = Centroid

M = jumlah data dalam sebuah kelompok

i = fitur ke-*i* dalam sebuah kelompok

x = matrik set data M x N.

N = jumlah fitur

persamaan 1 dilakukan sebanyak *p* (dimensi data) dimensi sehingga *i* mulai 1 sampai *p*.

4. Alokasikan masing-masing data ke sentroid/rata-rata terdekat.
5. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok

atau apabila ada perubahan nilai sentroid di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

Dalam mengukur jarak pada ruang jarak (*distance space*) dengan cara Euclidean yaitu menggunakan formula seperti pada persamaan 2.

$$D(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{i=1}^p |x_{2i} - x_{1i}|^2} \dots\dots(2)$$

Keterangan :

D = Jarak antara data x_2 dan x_1

$|\cdot|$ = Nilai mutlak

p = dimensi data

i = fitur ke-*i* dalam sebuah kelompok

x = matrik set data M x N.

N = jumlah fitur

Predikat kelulusan

Program sarjana regular merupakan program pendidikan akademik setelah pendidikan menengah, yang memiliki beban studi sekurang-kurangnya 144 SKS dan sebanyak-banyaknya 160 SKS yang dijadwalkan untuk 8 semester dan ditempuh dalam waktu paling cepat 8 semester dan paling lama 14 semester.

Perolehan nilai dan SKS digunakan untuk perhitungan beban studi. Sistem Kredit Semester (SKS) merupakan suatu sistem pendidikan yang menggunakan satuan kredit untuk menyatakan beban studi Mahasiswa (Permendikbud, 2014).

Predikat kelulusan dapat diterima Mahasiswa dengan memenuhi syarat antara lain sebagai berikut (Mendiknas, 2013):

1. Indeks prestasi kumulatif
2. Jumlah SKS yang telah ditentukan.
3. Lulus ujian skripsi.

Syarat diatas merupakan beberapa syarat yang terdapat pada perguruan tinggi, mengingat masing-masing perguruan tinggi memiliki syarat yang berbeda-beda.

Perguruan tinggi seperti STMIK Palangkaraya pada jurusan Teknik Informatika (TI) memiliki beban studi untuk Mahasiswa yaitu ditentukan antara 145 sampai 155 SKS yang dapat ditempuh selama 8 (delapan) semester. Beban studi tersebut juga berlaku untuk Mahasiswa jenjang Strata 1 pada jurusan lainnya seperti Sistem Informasi. Mahasiswa STMIK Palangkaraya pada jenjang Strata 1 dapat menyelesaikan studinya lebih

tepat yaitu dalam waktu 7 (tujuh) semester (STMIK Palangkaraya, 2014)

Penerapan pemodelan K-Means untuk menentukan predikat kelulusan Mahasiswa pada STMIK Palangkaraya Pemodelan merupakan penggunaan teknik ataupun prinsip tertentu dalam suatu rancangan sistem. Konsep dari pemodelan *K-Means* akan digunakan dalam pengelompokan dari data Mahasiswa yang diambil secara acak dan akan dijadikan sampel untuk menerapkan metode *K-Means*.

Metode *K-Means* menggunakan pengelompokan berdasarkan struktur yaitu termasuk ke dalam pengelompokan sekatan yang membagi set data ke dalam sejumlah kelompok yang tidak tumpang tindih (*non overlap*) antara satu kelompok dengan kelompok yang lainnya yaitu setiap data hanya akan menjadi anggota satu kelompok. Sedangkan pengelompokan berdasarkan keanggotaan data dalam kelompok, metode *K-Means* termasuk dalam kategori eksklusif yaitu sebuah data hanya dapat dipastikan akan menjadi anggota satu kelompok dan tidak menjadi anggota kelompok yang lainnya. Perhitungan menggunakan program sederhana sebagai

implementasi dari konsep *K-Means* agar metode yang digunakan mudah dipahami dan dapat dipelajari serta dapat dikembangkan untuk penelitian selanjutnya ataupun untuk kasus yang berbeda.

Penerapan konsep *K-Means* terhadap data Mahasiswa yang digunakan untuk menentukan pengelompokan predikat kelulusan. Data mahasiswa yang akan digunakan yaitu data Mahasiswa pada jurusan TI di STMIK Palangkaraya.

Dalam penulisan ini, penulis akan menerapkan pemodelan *K-Means* untuk menentukan predikat lulus Mahasiswa berdasarkan jumlah beban studi yang telah ditempuh, Indeks Prestasi Kumulatif (IPK) yang diperoleh oleh Mahasiswa, dan keterangan kelulusan skripsi.

Analisis

Analisis Kebutuhan

Pada analisis kebutuhan membahas beberapa kebutuhan dan atau persyaratan terkait dengan data masukan, proses dan hasil keluaran. Kebutuhan atau persyaratan ini diperoleh berdasarkan data yang digunakan untuk predikat kelulusan pada Mahasiswa dan data yang dapat diperoleh sebagai sampel dari Jurusan

karena data yang digunakan merupakan data pada Mahasiswa jurusan TI di STMIK Palangkaraya. Berdasarkan data tersebut diperoleh hasil analisa kebutuhan sistem berikut:

a. Kebutuhan *input*

Program yang akan dibangun membutuhkan data *input*, antara lain : Data Mahasiswa seperti NIM, Nama, jurusan, SKS yang dikumpulkan, IPK, dan kelulusan skripsi.

b. Kebutuhan Proses

Proses yang digunakan untuk mengolah data *input* adalah teknik clustering dengan menggunakan algoritma *K-Means*.

c. Kebutuhan *Output*

Output yang diharapkan berupa analisis dari hasil klustering sehingga dapat diketahui informasi / pola dari data *input*.

Analisis Fungsional

Untuk mendukung penelitian ini maka diperlukan alat seperti berikut :

a. Perangkat lunak

Perangkat lunak yang digunakan adalah *Microsoft Office Excel 2007* untuk simulasi program menggunakan pemodelan *K-Means*

serta sistem operasi yang digunakan adalah *Microsoft Windows XP*.

b. Perangkat keras

Perangkat keras yang digunakan adalah Processor Pentium dual, RAM 1 Gb, HDD 320 Gb, VGA Card NVidia, dan Monitor VGA.

Penelitian ini untuk mengetahui dan menganalisis hasil klustering sehingga dapat diketahui informasi / pola dari data *input*. Model yang digunakan dalam klustering yaitu *K-Means* yang akan dip roses menggunakan *software Microsoft Office Excel 2007*.

Desain

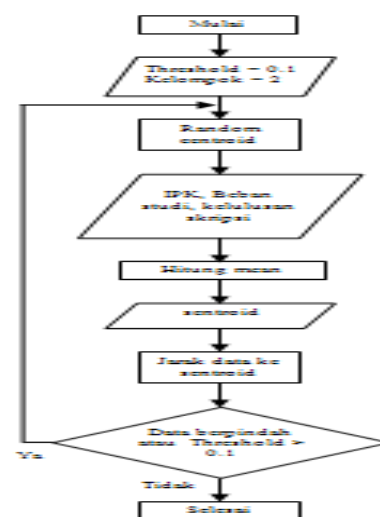
Proses dalam perancangan / desain meliputi penyusunan blok-blok program untuk memudahkan pada saat menyusun program sederhana untuk keperluan penggunaan model *K-Means*. Secara umum, desain perangkat lunak mempertimbangkan dari segi pengguna, yaitu : tampilan antar muka dan proses perhitungan/komputasi dengan benar. Pada penelitian ini, desain perangkat lunak hanya berupa pada proses perhitungan/komputasi menggunakan konsep *K-Means* menggunakan data Mahasiswa untuk mengetahui predikat kelulusan Mahasiswa, penulis

menggunakan antara muka untuk proses perhitungan yang akan digunakan dalam analisis.

Model *K-Means* untuk menentukan predikat kelulusan Mahasiswa pada STMIK Palangkaraya

Alur kerja pemodelan yang digunakan dimulai dengan menentukan *input*, *output*, penyimpanan, pemrosesan, dan basis pengetahuan. Pemodelan menggunakan alur dari konsep *K-Means*.

Pemodelan seperti pada Gambar 1 untuk pelatihan dan pengujian yang dibangun menggunakan antarmuka pada *Microsoft Office Excel* berdasarkan pada algoritma *K-Means*



Gambar 1. Model *K-Means* untuk menentukan predikat kelulusan Mahasiswa

1. Mengalokasikan semua data pada satu kelompok secara acak sebagai nilai centroid sementara.
2. Membaca data jumlah beban studi, IPK, dan kelulusan skripsi.
3. Mengitung lokasi sentroid atau pusat kelompok dengan menggunakan rata-rata (mean) semua data yang bergabung dalam setiap kelompok.
4. Menghitung jarak setiap data dengan centroid, jarak terdekat akan menjadi anggota pada kluster tersebut.
5. Rata-rata ini akan menjadi titik pusat untuk iterasi selanjutnya. Hal ini dilakukan sampai nilai rata-rata tidak berubah.
6. Jika proses perhitungan jarak dan pengelompokan pada langkah 6 sudah selesai dan ada perubahan anggota kluster atau nilai ambang batas (threshold) di atas 0.1, maka mengambil lagi nilai centroid secara random dari data Mahasiswa yakni akan membaca kembali ke langkah 3 data jumlah SKS, IPK, dan kelulusan skripsi. Demikian dilakukan sampai semua data terhitung.

Rancangan antarmuka

Rancangan antarmuka berdasarkan pada data *input* dan *output* yang

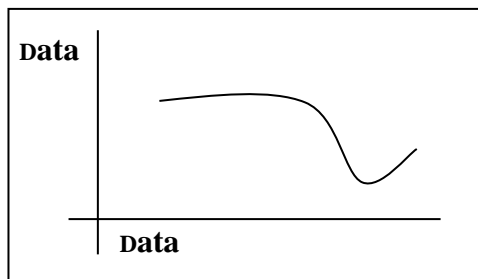
diperlukan dalam pembangunan aplikasi untuk mendukung penggunaan konsep *K-Means*. Rancangan antarmuka menggunakan *blok diagram* yang ditujukan untuk penulis dalam mengetahui perhitungan dan untuk analisis data dari hasil perhitungan yang diselesaikan. Dari antarmuka, penulis dapat melakukan perubahan data yang dapat diubah sebelum data masuk ke konsep *K-Means* seperti melakukan inisialisasi dengan mengalokasikan semua data pada satu kelompok secara acak oleh penulis.

Rancangan antarmuka yang digunakan seperti pada Gambar 2 dengan beberapa kolom yang digunakan dalam penggunaan konsep *K-Means* untuk menentukan predikat kelulusan.

Data ke-	IPK	SKS	Skripsi	Kelompok 1	Kelompok 2

Gambar 2. Rancangan antarmuka inisialisasi data

Rancangan antarmuka juga disajikan dalam bentuk grafik yang bertujuan untuk mempermudah penyajian data yang akan dianalisis. Rancangan antarmuka dalam bentuk grafik seperti pada Gambar 3.



Gambar 3. Rancangan antarmuka dalam bentuk grafik

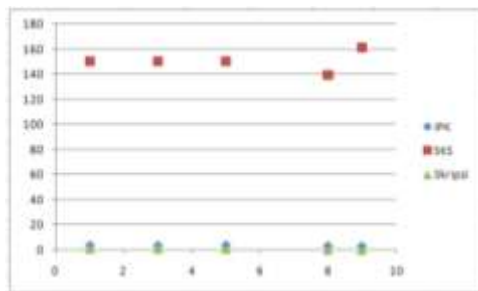
Hasil Penelitian

Penggunaan aplikasi dengan model *K-Means* untuk menentukan predikat kelulusan Mahasiswa berdasarkan rancangan alur pada model seperti pada gambar 1 dan rancangan antarmuka. Data Mahasiswa yang digunakan dalam penelitian sebagai bahan dalam implementasi model dari konsep *K-Means* untuk mengetahui akurasi penggunaan model tersebut, maka dilakukan perhitungan menggunakan software untuk mendapatkan hasil berupa analisis dari pemodelan yang digunakan. Hasil penelitian ini ditujukan untuk mengetahui hasil dari penggunaan metode *K-Means* dalam menentukan predikat kelulusan pada Mahasiswa STMIK Palangkaraya, khususnya pada Jurusan TI angkatan tahun 2011 sebagai sampel data. Penggunaan *software* dalam mengenal data dan kemudian disimulasikan untuk mendapatkan hasil yang maksimal.

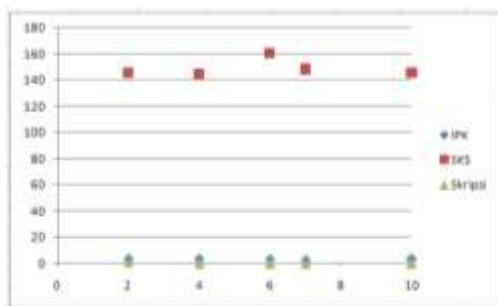
Pemodelan K-Means

Pada pemodelan *K-Means* dalam menentukan predikat kelulusan Mahasiswa menggunakan beberapa *input* penting yang diperlukan yaitu jumlah beban studi, IPK, dan kelulusan skripsi dari sampel yang digunakan sebanyak 10 data Mahasiswa serta *output* berupa analisis dari program yang digunakan untuk pemodelan *K-Means* pada data Mahasiswa. Data *input* dan *output* dapat digunakan dalam perancangan algoritma *K-Means* yang didasarkan pada algoritma yang telah didefinisikan sebelumnya seperti algoritma pada persamaan 1 dan persamaan 2. Jadi, pendekatan yang digunakan dapat diasumsikan bahwa dalam inisialisasi pada *K-Means* yang digunakan, data acak awal dilakukan oleh penulis atau sudah mengetahui lebih dulu data masuk dalam masing-masing kelompok dari data yang dimodelkan. Hal ini bertujuan untuk mengetahui perubahan data akhir yang akan dianalisis berdasarkan konsep *K-Means*. Dari semua data sampel yang digunakan dalam penelitian ini maka didapat kecenderungan data masuk ke *cluster* seperti pada tabel 2. Gambar 4. menunjukkan grafik keanggotaan data pada kelompok 1 dan Gambar 5.

enunjukkan grafik keanggotaan data pada kelompok 2.



Gambar 4. Grafik keanggotaan kelompok 1



Gambar 5. Grafik keanggotaan kelompok 2

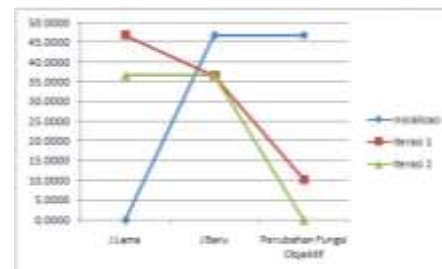
Tabel 1. menunjukkan hasil perbandingan nilai yang diperoleh yaitu perubahan fungsi objektif. Perbandingan diperoleh dari penggunaan algoritma *K-Means* terhadap data yang digunakan. Nilai *threshold* yang diberikan yaitu 0.1 dan didapat hasil berupa perubahan fungsi objektif bernilai dibawah nilai *threshold* pada saat iterasi ke-2 sehingga pengulangan yang dilakukan hanya sampai pada iterasi ke-2. Setiap langkah terdapat nilai jumlah lama dan nilai jumlah baru yang dibandingkan sehingga diperoleh nilai perubahan

fungsi objektif berdasarkan penggunaan algoritma *K-Means*.

Tabel 1. hasil perbandingan nilai pada setiap langkah

Langkah	J Lama	J Baru	Perubahan Fungsi Objektif
Inisialisasi	0.0000	46.9047	46.9047
Iterasi 1	46.9047	36.6228	10.2819
Iterasi 2	36.6228	36.6228	0.0000

Perbandingan nilai pada jumlah lama, jumlah baru, dan perubahan fungsi objektif tampak seperti pada gambar 6. yang menunjukkan grafik terhadap perbandingan data pada setiap langkahnya berdasarkan data pada tabel 1.



Gambar 6. grafik terhadap perbandingan data

Analisis algoritma *K-Means* untuk menentukan predikat kelulusan Mahasiswa pada STMIK Palangkaraya Pada tahap analisis pada penggunaan metode *K-Means*, data yang dijadikan sebagai sampel pada penelitian ini yaitu gabungan antara *input* dan *output* yang dikombinasikan dalam bentuk matrik dengan jumlah kolom sebanyak 3 kolom. Data yang digunakan pada

pemodelan *K-Means* dalam predikat kelulusan yaitu seperti pada lampiran 1.

Pada algoritma *K-Means*, terdapat algoritma yang digunakan untuk melakukan *clustering* sehingga didapat kecenderungan data masuk ke suatu *cluster* dengan membagi menjadi 2 kelompok yaitu lulus dan belum lulus dengan inisialisasi kelompok secara acak. Tabel 2 menunjukkan data yang cenderung masuk ke *cluster* yang digunakan pada langkah inisialisasi.

Tabel 2. kecenderungan data masuk ke *cluster*

Data ke-	IPK	SKS	Strata	Kelompok 1	Kelompok 2
1	3.35	130	1	*	
2	3.01	145	1		*
3	3.28	130	1	*	
4	3.01	144	0		*
5	3.47	130	1	*	
6	3.51	160	0		*
7	2.77	148	0		*
8	2.81	139	0	*	
9	2.71	161	0	*	
10	2.99	145	0		*

Dari persamaan 1 digunakan untuk menghitung sentroid yang hasilnya seperti pada lampiran 2. Sentroid yang didapat, diperoleh dari setiap langkah dimulai dari inisialisasi sampai dengan iterasi ke-2.

Persamaan 2 menunjukkan jarak yang dihitung pada setiap data yang ada terhadap setiap pusat klaster. Perhitungannya dengan menggunakan persamaan *Euclidean Distance Space* seperti pada contoh berikut ini :

a. Jarak antara data Mahasiswa pertama dengan pusat *cluster* pertama

$$c_{11} = \sqrt{(3.35 - 3.1240)^2 + (150 - 150.000)^2 + (1 - 0.6)^2}$$

$$c_{11} = 0.4594$$

b. Jarak antara data Mahasiswa pertama dengan pusat *cluster* kedua

$$c_{21} = \sqrt{(3.01 - 3.1240)^2 + (145 - 150.000)^2 + (1 - 0.6)^2}$$

$$c_{11} = 5.0173$$

Hasil dari perhitungan jarak pada setiap data dapat dilihat pada lampiran 3 dengan didapat bahwa jumlah lama adalah 0 (nol) dan jumlah baru adalah 46.9047 seperti tampak pada tabel 1. Dari lampiran 3, maka dimulai iterasi 1 seperti pada lampiran 4 untuk menentukan pengelompokan berikutnya dan menghasilkan pengelompokan data yang baru yaitu adanya perubahan dari pengelompokan awal. Pengelompokan baru tampak seperti pada lampiran 4. Penentuan kelompok baru dilihat dari nilai terkecil dari kolom kelompok 1 dan kolom kelompok 2 sehingga didapat hasil seperti pada kolom min.

Lampiran 5 pada iterasi 1 mengalokasikan setiap data pada sentroid terdekat yang sudah dipilih. Kemudian dilanjutkan dengan menghitung centroid seperti pada lampiran 6 beserta centroid yang didapat. Lampiran 7 menunjukkan Jarak data ke sentroid pada iterasi 1 dan 8 menunjukkan Hasil perhitungan

kembali jarak terpendek setiap data pada kelompok terdekat pada iterasi 2. Untuk iterasi 2 ditunjukkan seperti pada Lampiran 5 sampai lampiran 8 yang menunjukkan hasil sama dengan iterasi 1. Sehingga didapat bahwa nilai dari iterasi 2 berada dibawah nilai threshold seperti hasil yang ditunjukkan pada tabel 1. Lampiran 9 menunjukkan hasil data akhir pada iterasi ke-2.

Pembahasan

Pada *clustering* maka didapat hasil berupa keanggotaan setiap data pada setiap *cluster* dan dominan suatu data untuk masuk ke dalam suatu *cluster* tertentu. Jenis tipologi aplikasi *Data Mining* pada penelitian ini yaitu metode deskriptif Metode ini bertujuan untuk membantu *user* agar mudah melihat pola berdasarkan dari data yang telah ada.

Penentuan kelulusan Mahasiswa menggunakan metode K-Means dengan data *input* yaitu berupa jumlah beban studi yang telah ditempuh, IPK yang diperoleh Mahasiswa, dan kelulusan skripsi. Terdapat 10 data yang digunakan untuk 10 data Mahasiswa serta *output* berupa analisis dari program yang digunakan untuk pemodelan *K-Means* pada data Mahasiswa. Aplikasi sederhana

digunakan untuk mendukung pemodelan yang digunakan terhadap sampel data dalam mendapatkan hasil untuk dianalisis.

Perhitungan perbandingan sebagai perhitungan akurasi menggunakan rumus MAPE seperti pada tabel 3 yang menunjukkan perbandingan perhitungan akurasi dengan nilai yaitu 70% dapat mengenali hasil real dan 30% tidak dapat mengenali hasil real.

SIMPULAN

Pada penelitian penentuan predikat kelulusan Mahasiswa menggunakan pemodelan *K-Means* dapat diambil kesimpulan sebagai berikut:

1. *Input* data yang digunakan dalam penelitian ini yaitu jumlah beban studi yang telah ditempuh, IPK yang diperoleh Mahasiswa, dan kelulusan skripsi dan *output* ditampilkan dari data yang digunakan yaitu berupa pengelompokan data berdasarkan algoritma *K-Means*.
2. Minimnya data yang digunakan sebagai sampel yaitu 10 data yang sederhana agar memudahkan dalam proses analisis dan penjelasan terhadap perhitungan dalam menggunakan algoritma *K-Means*

3. terhadap data yang digunakan serta untuk memudahkan pemahaman penggunaan pemodelan *K-Means*
4. Pemodelan *K-Means* yang digunakan dalam penelitian ini memiliki keakuratan yang cukup baik terhadap permasalahan yang terjadi terkait predikat kelulusan Mahasiswa dengan hasil 70% dapat mengenali data pada 10 data yang digunakan sebagai sampel.
5. Metode lain dapat digabungkan atau dikembangkan dalam membangun sistem agar dapat diketahui perbandingan hasilnya seperti menggabungkan dengan metode *subtractive clustering* atau juga *fuzzy* dalam *fuzzy C-Means*.
6. Sistem ini dapat menjadi rekomendasi untuk pihak-pihak yang membutuhkan dengan sistem ini seperti lembaga pendidikan lainnya baik untuk penelitian, pengembangan maupun penggunaan sistem itu sendiri seperti dikembangkan menjadi sistem pakar

DAFTAR PUSTAKA

Chang, HJ., Lun-Ping H., dan Chia-Ling H. 2007. Expert Systems, An anticipation model of potential customers' purchasing behavior based on clustering analysis and association rules analysis, *Expert*

Systems with Applications(32), 753–764

Geetha, A., Ramalingam, V., dan Palanivel, S. 2014. An Integrated Face Tracking and Facial Expression Recognition System, *proquest*, 1-9

Han, J. dan Kamber, M. 2001. *Data Mining Concepts and Techniques Second Edition*. Morgan Kauffman. San Francisco

Kementerian Pendidikan dan Kebudayaan an Direktorat Jenderal Pendidikan Tinggi dan Badan Standar Nasional Pendidikan. 2013.

Peraturan Menteri Pendidikan dan

n
Kebudayaan Republik Indonesia Tentang Standar Nasional Pendidikan Tinggi (SNPT). Mendiknas

Naba, EA. 2009. *Belajar Cepat Fuzzy Logic Menggunakan Matlab*. Andi. Yogyakarta

Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia. 2014. Permendikbud RI Nomor 49 Tahun 2014 Tentang Standar Nasional Pendidikan Tinggi. Permendikbud

Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Andi. Yogyakarta

Rusdiana, L., E.Sediyono, dan B.Surarso. 2015. *Studi Implementasi Adaptive Neuro Fuzzy Inference System Untuk Menentukan Normalitas Kehamilan*. Universitas Diponegoro. Semarang

STMIK Palangkaraya. 2015. *Panduan AKADEMIK*, STMIK Palangkaraya, Palangka Raya

Tan, P. 2006. *Introduction to Data Mining*. Pearson Education. Boston