

PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN FUZZY C-MEANS DALAM MENENTUKAN PREDIKAT KELULUSAN MAHASISWA

Lili Rusdiana

Teknik Informatika, STMIK Palangkaraya
Jl. G.Obos No.114, Palangkaraya, 73111
E-mail : fasliiana7@gmail.com

ABSTRAK

The research was purposed to compare the method performance for knowing the accuracy of K-Nearest Neighbor and Fuzzy C-Means in determining of students graduation in system information study program STMIK Palangka Raya. Extreme programming was used in this research because it was devoted to development to software appropriately. Some steps which used in research were planning by preparing the data and other necessary needs, designing, encoding on Matlab software, and testing. Based on the calculation of accuracy using MAPE for value the result of testing percentage in comparing the method obtained that Fuzzy C-Means method could predict by increasing the accuracy value along with increasing the training data, while the K-Nearest Neighbor method got the decreasing of accuracy value although the training data was in creased. But, the K-Nearest Neighbor method was still better on accuracy which the accuracy was above 90% and Fuzzy C-Means method had the accuracy below 90% in this research.

Kata Kunci: *Extreme Programming, Fuzzy C-Means, K-Nearest Neighbor, Predicate of students graduations*

1. PENDAHULUAN

Metode pengelompokan data dapat diterapkan dalam berbagai bidang, salah satunya yakni Perguruan Tinggi dengan menggunakan data seperti data Mahasiswa yang lulus pendaftaran Mahasiswa baru, klasifikasi dosen yang berprestasi, ataupun klasifikasi untuk menentukan predikat kelulusan Mahasiswa. Beragam metode yang dapat digunakan dalam pengelompokan seperti *Naïve Bayes*, *K-Nearest Neighbor* (K-NN), *Artificial Neural Network* (ANN), *K-Means*, *Fuzzy C-Means* dan metode lainnya. Metode *K-Means* digunakan untuk memprediksi kelulusan Mahasiswa (Rusdiana dan Sam'ani; 2016) dan penggunaan data yang sama namun berbeda metode yakni *Fuzzy C-Means* (Rosmiati dan Rusdiana; 2016). H

Penggunaan penerapan pada metode yang berbeda dengan suatu data yang sama dapat menghasilkan suatu nilai keakuratan yang berbeda pula seperti penggunaan metode *Naïve Bayes* dan *K-Nearest Neighbor* pada data status kerja di Kabupaten Demak tahun 2012. Hasil penggunaan metode tersebut diketahui bahwa metode *Naïve Bayes* memiliki nilai keakuratan sebesar 94.09% sedangkan *K-Nearest Neighbor* memiliki nilai keakuratan sebesar 96,06 % (Putri, dkk ; 2014). Penelitian tersebut menghasilkan nilai keakuratan yang lebih unggul pada penggunaan metode *K-Nearest Neighbor* daripada metode *Naïve Bayes*. Namun, penelitian saat ini untuk mengetahui hasil perbandingan menggunakan metode *K-Nearest Neighbor* dengan *Fuzzy*

C-Means, akan mendapatkan hasil yang sama atau berbeda dari penelitian sebelumnya yang membandingkan metode *Naïve Bayes* dengan *K-Nearest Neighbor*.

Metode *Fuzzy C-Means* dan *K-Nearest Neighbor* yang memerlukan data latih dan data uji. *Fuzzy C-Means* merupakan metode untuk klasifikasi data berdasarkan pada nilai keanggotaan tiap-tiap data dalam suatu *cluster*. Sedangkan metode *K-Nearest Neighbor* yang melakukan klasifikasi dengan menghitung kedekatan jarak antardata. Nilai pada suatu jarak akan digunakan untuk mencari nilai terdekat. Data latih pada metode *Fuzzy C-Means* haruslah melalui proses pelatihan, sedangkan metode *K-Nearest Neighbor* tidak memerlukan pelatihan.

Penggunaan metode yang berbeda dapat dibandingkan untuk mengetahui kelebihan ataupun keakuratan dalam mengklasifikasi dari masing-masing metode pada penggunaan data yang sama yaitu perbandingan metode *K-Nearest Neighbor* dan *Fuzzy C-Means* dengan menggunakan data Mahasiswa untuk menentukan prediksi kelulusan.

2. RUANG LINGKUP PENELITIAN

Dalam penelitian ini permasalahan mencakup sebagai berikut:

a. Cakupan permasalahan.

Cakupan permasalahan yaitu analisis hasil perbandingan pada metode yang berbeda yaitu

metode *K-Nearest Neighbor* dengan *Fuzzy C-Means* pada penggunaan data Mahasiswa Program Studi Sistem Informasi angkatan tahun 2011 STMIK Palangkaraya.

b. Batasan-batasan penelitian.

Penelitian ini memiliki batasan-batasan sebagai berikut :

- 1) Metode yang menjadi perbandingan hanyalah metode *K-Nearest Neighbor* dengan *Fuzzy C-Means*.
- 2) Data yang digunakan pada penelitian ini yaitu data Mahasiswa Program Studi Sistem Informasi, angkatan tahun 2011 yang direkap pada tahun 2016 pada STMIK Palangkaraya.
- 3) Aplikasi pendukung yang digunakan yaitu *Microsoft Office Excel 2007* dan *Matlab 7*.
- 4) Penelitian ini menghasilkan analisis perbandingan nilai keakuratan penggunaan 2 metode yang berbeda yaitu metode *K-Nearest Neighbor* dengan *Fuzzy C-Means* pada data dan aplikasi yang sama.

3. BAHAN DAN METODE [Huruf Kapital]

Bahan dan metode yang digunakan dalam penelitian ini yaitu sebagai berikut :

3.1 Teknik Pengelompokan

Teknik pengelompokan memiliki beberapa cara seperti melakukan pemisahan, pemecahan, maupun segmentasi data ke dalam sejumlah kelompok (*cluster*) berdasarkan suatu karakteristik tertentu. Analisis kelompok dilakukan untuk mengetahui informasi dan hubungan dari kesamaan data yang telah dikelompokkan berdasarkan karakteristik tertentu yang memiliki kemiripan atau hubungan satu sama lainnya dan tidak memiliki hubungan atau berbeda dengan lainnya pada kelompok lainnya. Dalam konsep pengelompokan, diperlukan kemiripan yang lebih besar dalam kelompok dan perbedaan yang lebih besar di antara kelompok yang lainnya. Pengelompokan data dapat dibagi menjadi kelompok data latih dan data uji, baik data latih tersebut dilakukan pelatihan maupun tidak. Dengan tujuan untuk mencari *prototype* kelompok yang paling representative terhadap data agar dapat memberikan abstraksi dari setiap objek data dalam suatu kelompok.

3.2 Metode *Fuzzy C-Mean*

Metode pengelompokan *Fuzzy C-Means* merupakan pengelompokan dengan satu data tunggal yang dapat diasumsikan sebagai suatu kelompok, beberapa kelompok kecil yang bergabung menjadi kelompok yang besar, sehingga semua data yang digunakan bergabung ke dalam salah satu kelompok tersebut.

Metode *Fuzzy C-Means* yang termasuk dalam *Fuzzy Clustering* yakni *Fuzzy Clustering* yang mengelompokkan suatu data sebagai *input-output* ke dalam beberapa *cluster* dan *output* yakni berupa daftar *cluster*. Metode *Fuzzy C-Means* umumnya telah memiliki jumlah *cluster* yang dapat ditebak pada sekumpulan data.

Fuzzy C-Means yang digunakan untuk mengelompokkan data seperti dalam bidang biologi. Menggunakan metode pengelompokan untuk data ekspresi gen dalam algoritma *clustering* probabilistik. Menempatkan gen ke beberapa *cluster* dan representasi perilaku gen yang lebih tepat. Penelitian ini menunjukkan bahwa metode *Fuzzy c-means* dapat memprediksi fungsi gen secara efektif (Tari, 2009).

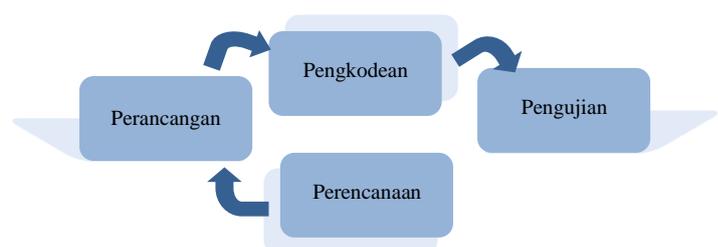
3.3 Metode *K-Nearest Neighbor*

Metode *K-Nearest Neighbor* memerlukan data uji dan data latih untuk menghitung jarak untuk melakukan klasifikasi dengan menghitung kedekatan jarak antara suatu data dengan data yang lainnya (Prasetyo, 2012). Nilai pada suatu jarak akan digunakan untuk mencari nilai kedekatan ataupun kemiripan antara data uji dan data latih yang sebelumnya data latih tidak memerlukan pelatihan. Tetapi, jika terdapat hasil yang memiliki perbedaan maka akan dipilih data dengan jumlah yang terbanyak.

Penggunaan metode *K-Nearest Neighbor* yang pernah dilakukan untuk menganalisis data medis (Khamis, 2014). Penelitian ini dilakukan karena adanya perkembangan data pasien yang disimpan pada suatu rumah sakit. Data yang tersimpan dapat bersifat tersembunyi sementara data tersembunyi tersebut dapat membantu untuk menganalisis sebagai hasil diagnosis yang lebih baik. Algoritma *K-Nearest Neighbor* digunakan untuk mengapatkan hasil diagnosis yang lebih akurat. Hasil penelitian ini yakni aplikasi yang dapat mengatasi kesalahan dalam diagnosis.

3.4 *Extreme Programming*

Extreme Programming merupakan salah satu cabang dari metode *agile* dengan suatu pendekatan untuk pengembangan perangkat lunak secara cepat dengan melalui beberapa tahapan seperti pada gambar 1 (kendall dan kendall, 2011).



Gambar 1. Metode pengembangan perangkat lunak

Langkah-langkah *Extreme Programming* yang digunakan dalam penulisan ini seperti pada Gambar 1, yaitu sebagai berikut:

a. Perencanaan (*Planning*)

Mengetahui ruang lingkup pekerjaan sebagai rencana untuk memodelkan sistem, meliputi pencarian informasi data dan situs yang berhubungan dengan keperluan data kelulusan Mahasiswa seperti nama, NIM, IPK, jumlah SKS, dan keterangan kelulusan skripsi yang akan dijadikan sebagai *input* dan *output* dalam aplikasi

seederhana, fitur dan fungsi disesuaikan dengan kebutuhan pengembangan.

b. Perancangan (*Design*)

Tahapan ini membuat perancangan yang diperlukan dalam penelitian yaitu Lembar Kerja Tampilan (LKT) untuk merancang aplikasi dan *software* Microsoft Office Excel untuk merancang data latih yang kemudian diolah kembali menggunakan *software* Matlab.

c. Pengkodean (*Coding*)

Tahapan pengkodean menggunakan pengkodean yang terdapat pada *software* Matlab untuk menggunakan dan mengolah data latih dan data uji.

d. Pengujian (*Testing*)

Pengujian menggunakan aplikasi sederhana untuk mengetahui hasil dari data, metode ataupun aplikasi yang digunakan sehingga didapat suatu kesimpulan mengenai layanan atau fitur dan fungsionalitas yang terdapat pada aplikasi yang dibangun menggunakan metode yang berbeda. Validasi dari pengujian menggunakan perhitungan dari MAPE (*Mean Absolute Percentage Error*).

3.5 Predikat kelulusan

Perguruan tinggi STMIK Palangkaraya mempunyai program sarjana yang merupakan program pendidikan akademik yang memiliki beban studi sekurang-kurangnya 144 SKS dan sebanyak-banyaknya 160 SKS yang dijadwalkan untuk 8 semester dan ditempuh minimal dalam waktu 8 semester dan maksimal 14 semester.

Predikat kelulusan Mahasiswa pada suatu Perguruan Tinggi dapat berbeda karena adanya peraturan atau syarat yang berlaku pada Perguruan Tinggi tersebut. Beberapa syarat tersebut antara lain sebagai berikut :

- a. Indeks prestasi kumulatif menunjukkan paling sedikit yaitu nilai minimal yang telah ditetapkan oleh kampus.
- b. Jumlah SKS yang telah ditentukan.
- c. Lulus ujian skripsi.

Syarat diatas merupakan beberapa syarat yang terdapat pada perguruan tinggi, mengingat masing-masing perguruan tinggi memiliki syarat yang berbeda-beda. Mahasiswa dapat dinyatakan lulus berdasarkan syarat yang telah ditentukan oleh masing-masing perguruan tinggi tersebut, salah satunya STMIK Palangkaraya. Walaupun masih banyak syarat lainnya yang diberlakukan seperti pembayaran dan kelulusan ujkikom serta lainnya. Dalam penelitian ini hanya menggunakan 3 syarat tersebut yaitu IPK yang diperoleh, jumlah SKS yang ditempuh, dan keterangan kelulusan skripsi untuk perbandingan metode yang digunakan dan ditambahkan dengan data nama serta nim dalam penggunaan aplikasi yang dibangun.

4. RANCANGAN SISTEM/APLIKASI

Setelah semua data yang diperlukan telah terkumpul maka tahap selanjutnya yaitu mempersiapkan data tersebut agar dapat digunakan dalam proses

perbandingan metode. Data mentah selanjutnya digunakan untuk proses perancangan / desain meliputi penyusunan bagian-bagian aplikasi untuk memudahkan pada saat penggunaan perhitungan pada aplikasi dan perancangan antar muka (*user interface*) untuk keperluan interaksi aplikasi dalam penggunaannya untuk analisis.

Rancangan antarmuka menggunakan Lembar Kerja Tampilan (LKT) untuk rancangan tampilan perhitungan dan analisis data dari hasil perhitungan yang diselesaikan menggunakan masing-masing metode. LKT pada penelitian ini menggunakan aplikasi Microsoft Office word yang LKT sendiri terdiri dari nomor lembar kerja, bagian tampilan, navigator, dan keterangan. Dari antarmuka, dapat melakukan perubahan data yang dapat diubah sebelum data masuk ke masing-masing metode seperti melakukan inisialisasi dengan mengalokasikan semua data pada satu kelompok secara acak seperti bagian tampilan yang ditunjukkan pada Gambar 2.

| | | | |
|---------|-----|---------|---|
| No. : 1 | | | Navigator : Data sebagai data latih |
| SKS | IPK | Skripsi | Keterangan : > SKS untuk jumlah SKS yang telah ditempuh > IPK untuk nilai IPK > Skripsi untuk keterangan kelulusan skripsi |
| | | | |
| | | | |
| | | | |
| | | | |

Gambar 2. Rancangan antarmuka inisialisasi data

Rancangan antarmuka untuk penggunaan aplikasi guna membandingkan penggunaan metode *K-Nearest Neighbor* dengan *Fuzzy C-Means* terhadap data Mahasiswa dalam penentuan predikat kelulusan dengan menggunakan data latih dan data uji seperti pada gambar 3.

| | | | |
|---------|---|---|---|
| N | | Aplikasi Penentuan Predikat Kelulusan Mahasiswa | Navigator : Klik tombol "Proses" untuk mengetahui hasil dari data yang telah dimasukkan. |
| NIM | : | <input type="text"/> | Keterangan : > NIM untuk nomor induk yang dimiliki Mahasiswa > Nama untuk nama Mahasiswa > SKS untuk jumlah SKS yang telah ditempuh > IPK untuk nilai IPK > Skripsi untuk keterangan kelulusan skripsi |
| Nama | : | <input type="text"/> | |
| SKS | : | <input type="text"/> | |
| IPK | : | <input type="text"/> | |
| Skripsi | : | <input type="text"/> | |
| | | Keterangan | <input type="button" value="Proses"/> |

Hasil :

Gambar 3. Rancangan antarmuka aplikasi

5. IMPLEMENTASI

Penggunaan aplikasi untuk membandingkan kinerja dari metode *K-Nearest Neighbor* dengan *Fuzzy C-Means* dalam menentukan predikat kelulusan Mahasiswa berdasarkan rancangan pada desain antarmuka yaitu menggunakan data yang diperoleh dari bagian akademik dan kemahasiswaan yaitu data Mahasiswa pada STMIK Palangkaraya, Program Studi Sistem Informasi, angkatan 2011, dan data yang digunakan merupakan data terakhir pada tahun 2016 dan tidak diperbaharui guna penggunaan data pada aplikasi.

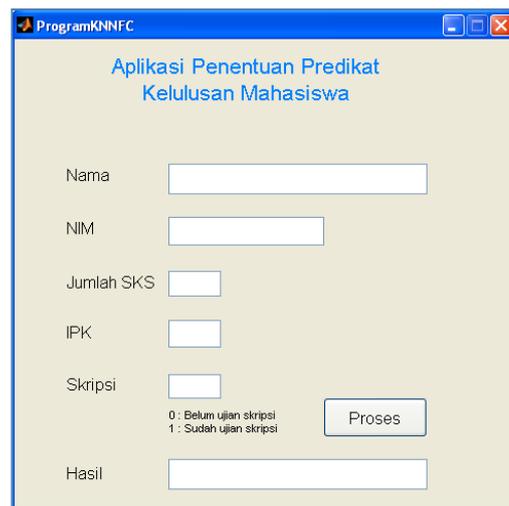
Penggunaan sejumlah data awal sebanyak 136 data dan data setelah mengalami proses awal (*pre-processing*) data menjadi 94 data yang dijadikan sebagai data dalam aplikasi dengan menghilangkan beberapa data yang tidak dapat digunakan seperti adanya kerangkapan data.

Antarmuka berdasarkan pada data *input* dan *output* yang diperlukan untuk mengimplementasikan dalam mendukung penggunaan metode *K-Nearest Neighbor* dengan *Fuzzy C-Means*. Gambar 4 menunjukkan data dari data latih yang berjumlah 10 data dari seluruh data yang digunakan secara acak. Data ini yang akan dijadikan sebagai pencarian nilai terdekat yaitu 0 dan 1.

| NIM | SKS | IPK | Skripsi |
|-----|------|-----|---------|
| 112 | 2.09 | 0 | 0 |
| 148 | 2.64 | 0 | 0 |
| 151 | 2.56 | 1 | 0 |
| 151 | 3.36 | 1 | 1 |
| 2 | 2 | 0 | 0 |
| 109 | 2.82 | 0 | 0 |
| 143 | 2.57 | 1 | 0 |
| 151 | 2.83 | 1 | 1 |
| 145 | 3.23 | 1 | 1 |
| 143 | 2.70 | 1 | 0 |

Gambar 4. Data latih yang berjumlah 10 data

Gambar 5 menunjukkan tampilan untuk data *input* dan *output* pada aplikasi yang digunakan dalam membandingkan penggunaan metode *K-Nearest Neighbor* dan *Fuzzy C-Means*.



Gambar 5. Tampilan *interface* aplikasi

Berdasarkan proses dengan teknik *clustering* dan menggunakan metode *K-Nearest Neighbor* dalam menentukan kelulusan mahasiswa dengan data *input* yang digunakan dalam perhitungan yaitu berupa jumlah beban studi (SKS) yang telah ditempuh, IPK yang diperoleh Mahasiswa, dan keterangan kelulusan skripsi maka diperoleh hasil yaitu untuk 10 data latih dan 50 data latih pada masing-masing dilakukan 2 tahap data uji yaitu menguji data latih dan menguji semua data.

Pengujian dengan menguji data latih berdasarkan 10 data latih maka didapat bahwa dari 10 data latih yang sesuai sebanyak 10 data dan tidak ada data yang tidak sesuai sehingga disimpulkan bahwa semua data latih dapat dikenali. Sedangkan untuk pengujian pada semua data yaitu 94 berdasarkan 10 data latih maka yang sesuai sebanyak 88 data dan yang tidak sesuai sebanyak 6 data. Pengujian dengan menguji data latih berdasarkan 50 data latih maka didapat bahwa data yang sesuai sebanyak 45 data dan 5 data lainnya tidak sesuai. Sedangkan pengujian pada semua data yaitu 94 data maka yang sesuai sebanyak 87 data dan yang tidak sesuai sebanyak 7 data.

Perbandingan data pengujian pada data latih dan seluruh data dapat dilihat seperti pada tabel 1.

Tabel 1. Perbandingan data pengujian

| Data latih | 10 | | 50 | |
|--------------|----|----|----|----|
| Data uji | 10 | 94 | 50 | 94 |
| Sesuai | 10 | 88 | 45 | 87 |
| Tidak sesuai | 0 | 6 | 5 | 7 |

Berdasarkan tabel 1 maka didapat prosentase perbandingan yaitu seperti pada tabel 2.

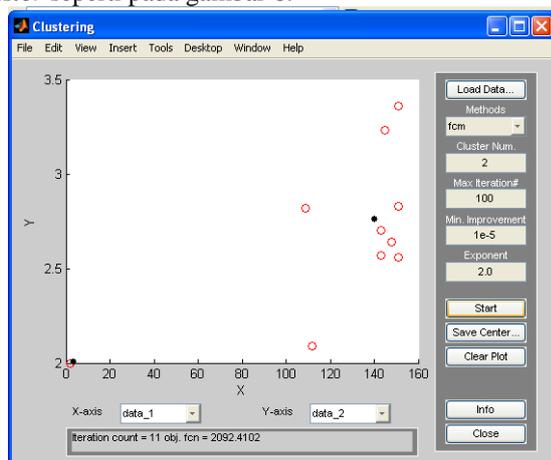
Tabel 2. Prosentase perbandingan data pengujian

| Data latih | 10 | | 50 | |
|--------------|------|-------|-----|-------|
| Data uji | 10 | 94 | 50 | 94 |
| Sesuai | 100% | 93,6% | 90% | 92,5% |
| Tidak sesuai | 0% | 6,4% | 10% | 7,5% |

Tabel 1 menunjukkan perbandingan dari 2 pengujian menggunakan data latih dan data latih ini tidak dilakukan pelatihan karena pada metode *K-Nearest Neighbor* data latih tidak perlu dilakukan pelatihan. Nilai yang tidak

sesuai semakin besar dan nilai yang sesuai semakin kecil dari tahapan-tahapan pengujian yang dilakukan pada data latih yang semakin banyak. Perhitungan perbandingan antara pengujian standar dengan pengujian sistem sebagai perhitungan akurasi menggunakan rumus MAPE dan hasilnya seperti pada tabel 2 yang menunjukkan perbandingan pelatihan dan pengujian dalam bentuk prosentase dengan nilai tertinggi yaitu 100% sistem dapat mengenali seluruh data uji dan 0% sistem dapat mengenali data dari seluruh data dengan menggunakan data latih yang minimum dalam penelitian ini. Namun dengan data latih yang meningkat, jumlah prosentasi kesesuaian menurun tetapi tidak menurun secara drastis, peningkatan tetap ada pada pengujian data latih yang lebih banyak yaitu dari 50 data latih pada 94 data uji. Dari hasil ini didapat bahwa semakin banyak data latih, maka akan menghasilkan kemunduran prosentase kesesuaian data pengujian dalam menentukan predikat kelulusan Mahasiswa pada data latih, namun terjadi peningkatan prosentasi prediksi semua data jika menggunakan 50 data latih yaitu dari 90% menjadi 92,5%.

Hasil dari penggunaan metode *Fuzzy C-Means* berdasarkan pelatihan data yang digunakan, terdapat 11 iterasi dengan menampilkan hasil *output* pada pembagian *cluster* seperti pada gambar 6.



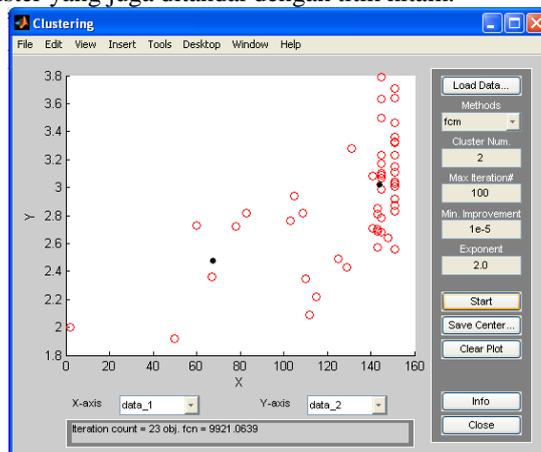
Gambar 6. Clustering dari 10 data latih

Pada gambar 6 nampak jendela *clustering* yang digunakan untuk mencari titik pusat kluster, terdapat 2 titik kluster yaitu kluster lulus dan kluster belum lulus. Data yang digunakan dalam hal ini yaitu data latih seperti pada gambar 4.

Gambar 6 memperlihatkan adanya dua titik hitam yang ditampilkan sebagai pusat kluster. Setelah dua titik pusat kluster telah ditemukan, maka data tertentu dapat dikelompokkan dalam kluster yang tertentu pula berdasarkan dari jaraknya terhadap titik-titik pusat tersebut (Widodo dkk, 2013).

Gambar 6 juga menunjukkan jumlah iterasi sebanyak 11 iterasi untuk data latih yang berjumlah 10 data berdasarkan *clustering* yang dilakukan. Sedangkan untuk data latih yang berjumlah 50 data berdasarkan *clustering* yang dilakukan, terdapat 23 iterasi dengan menampilkan hasil *output* yang sama pada pembagian *cluster* seperti

pada gambar 7. Pada gambar 7 nampak jendela *clustering* yang digunakan untuk mencari titik pusat kluster yang juga ditandai dengan titik hitam.



Gambar 7. Clustering dari 50 data latih

Pengujian dengan menguji data latih berdasarkan 10 data latih maka didapat bahwa dari 10 data latih yang sesuai sebanyak 3 data dan 7 data lainnya tidak sesuai. Sedangkan pengujian pada semua data yaitu 94 data maka yang sesuai sebanyak 70 data dan yang tidak sesuai sebanyak 28 data. Pengujian dengan menguji data latih berdasarkan 50 data latih maka didapat bahwa dari 50 data latih, yang sesuai sebanyak 43 data dan 7 data lainnya tidak sesuai. Sedangkan pengujian pada semua data yaitu 94 data maka yang sesuai sebanyak 80 data dan yang tidak sesuai sebanyak 18 data.

Perbandingan data pengujian pada data latih dan seluruh data dapat dilihat seperti pada tabel 3.

Tabel 3. Perbandingan data pengujian

| Data latih | 10 | 50 |
|--------------|----|----|
| Data uji | 10 | 94 |
| Sesuai | 3 | 43 |
| Tidak sesuai | 7 | 28 |

Berdasarkan tabel 3 maka didapat prosentase perbandingan yaitu seperti pada tabel 4.

Tabel 4. Prosentase perbandingan data pengujian

| Data latih | 10 | 50 |
|--------------|-----|-----|
| Data uji | 10 | 94 |
| Sesuai | 30% | 86% |
| Tidak sesuai | 70% | 14% |

Tabel 3 menunjukkan perbandingan dari 2 tahap pelatihan dan pengujian, nilai yang tidak sesuai semakin kecil dan nilai yang sesuai semakin besar dari tahapan-tahapan pelatihan yang dilakukan pada data yang semakin banyak. Perhitungan perbandingan antara pengujian standar dengan pengujian sistem sebagai perhitungan akurasi menggunakan rumus MAPE dan hasilnya seperti pada tabel 4 yang menunjukkan perbandingan pelatihan dan pengujian dalam bentuk prosentase dengan nilai tertinggi yaitu 86% sistem dapat mengenali seluruh data latih dan 80.8% sistem dapat mengenali data dari seluruh data. Dari hasil ini didapat bahwa semakin banyak data yang dilatih, maka akan menghasilkan kesesuaian data pengujian yang membaik

dengan meningkatnya kinerja sistem dalam menentukan predikat kelulusan Mahasiswa.

Hasil perhitungan seperti yang ditunjukkan pada tabel 1 dan tabel 2 menunjukkan bahwa penggunaan metode *K-Nearest Neighbor* menunjukkan prosentase tertinggi yaitu 100% dapat mengenali data latih yang minim digunakan dalam penelitian ini. Namun, dengan metode *K-Nearest Neighbor* akan terjadi penurunan nilai pengujian pada penambahan data latih, sedangkan pada penggunaan metode *Fuzzy C-Means* seiring dengan meningkatnya data latih maka akan meningkatkan pula penilaian pada data uji seperti pada tabel 3 dan tabel 4. Hal ini menunjukkan bahwa adanya peningkatan nilai yang membaik pada penggunaan metode *Fuzzy C-Means* daripada penggunaan metode *K-Nearest Neighbor* yg terjadi penurunan nilai pengujian seiring bertambahnya data latih.

Berdasarkan prosentase pada tabel 2 dan 3 dapat diketahui bahwa nilai akurasi tertinggi dari metode *K-Nearest Neighbor* dengan nilai prosentase diatas 90% sedangkan metode *Fuzzy C-Means* menunjukkan hasil masih dibawah 90%.

Perbandingan antara metode *K-Nearest Neighbor* dengan metode *Fuzzy C-Means* dengan hasil bahwa metode *Fuzzy C-Means* lebih baik akurasi daripada metode *K-Nearest Neighbor* jika dilihat dari penggunaan data latih sedangkan untuk nilai akurasi tertinggi yaitu dari penggunaan metode *K-Nearest Neighbor* dalam penentuan predikat kelulusan Mahasiswa yang menggunakan data Mahasiswa Program Studi Sistem Informasi angkatan tahun 2011.

6. KESIMPULAN

Pada penelitian perbandingan metode *K-Nearest Neighbor* dan *Fuzzy C-Means* dalam menentukan predikat kelulusan Mahasiswa dapat diambil kesimpulan sebagai berikut:

1. Dari sejumlah data awal 136 data setelah mengalami proses awal (*pre-processing*) data menjadi 94 data yang dijadikan sebagai data dalam membandingkan metode.
2. Hasil dari perbandingan metode didapat bahwa metode *Fuzzy C-Means* dapat melakukan prediksi secara meningkat seiring dengan bertambahnya data latih, sedangkan metode *K-Nearest Neighbor* mengalami pengurangan nilai akurasi seiring dengan bertambahnya data latih. Namun, metode *K-Nearest Neighbor* tetap lebih baik akurasi yaitu diatas 90%.

7. SARAN

Pengembangan dapat dilakukan terhadap penelitian ini antara lain dalam saran sebagai berikut :

1. Perbandingan metode dapat dilakukan dengan metode lainnya ataupun dengan menggunakan data yang lainnya sehingga akan didapatkan hasil yang nantinya dapat dijadikan sebagai referensi untuk perbandingan pula.
2. Aplikasi dapat dikembangkan dengan menambahkan jumlah data tidak hanya sebatas pada satu Program Studi saja sehingga jumlah data yang digunakan akan lebih banyak lagi untuk mengetahui keakuratan aplikasi dan metode walaupun menggunakan metode yang sama.
3. Metode lainnya dapat digabungkan dengan metode pada penelitian ini dan kemudian dapat dijadikan sebagai bahan perbandingan.

8. DAFTAR PUSTAKA

- Kendall, K.E. dan Julie, E.K., 2011, *Systems Analysis and Design, Eighth Edition* Prentice Hall, New Jersey
- Khamis, H.S., Kriptu, W.C., dan Stephen, K., 2014, Application of k-Nearst Neighbour Classification in Medical Data Mining, *International Journal of Information and Communication Technology Research*, IV(4), Hal 121-128
- Prasetyo, E., 2012, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, Yogyakarta : Andi
- Putri, R.E., Suparti, dan Rita R., 2014, Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012, *Jurnal Gaussian*, Vol.3 No.4, Hal 831- 838
- Rosmiati dan Rusdiana L., 2016, Aplikasi Berbasis *Fuzzy C-Means* Dalam Penentuan Predikat Kelulusan Mahasiswa, *Jurnal Ilmu Komputer*, II (2), Hal 1-9
- Rusdiana L. dan Sam'ani, 2016, Pemodelan K-Means Untuk Menentukan Predikat Kelulusan Mahasiswa STMIK Palangkaraya, *Jurnal SAINTEKOM*, Vol.6 No.1, Hal 1-15
- Tari, L., Chitta B., dan Seungchan K., 2009, Fuzzy c-means Clustering With Prior Biological Knowledge, *Journal of Biomedical Informatics*, Hal.74-81
- Widodo, P.P., Rahmadya T.H., dan Herlawati, 2013, *Penerapan Data Mining Dengan Matlab*, Bandung: Rekayasa Sains