

TESIS

**Perbandingan Metode K-Means Dengan Fuzzy C-Means
Untuk Analisa Karakteristik Mahasiswa Berdasarkan
Kunjungan Ke Perpustakaan**

**(Studi Kasus Sekolah Tinggi Agama Hindu Negeri
Tampung Penyang Palangka Raya)**



Oleh :

Ni Putu Eka Merliana

No. Mhs. : 135302010/PS/MTF

**MAGISTER TEKNIK INFORMATIKA
PROGRAM PASCASARJANA
UNIVERSITAS ATMA JAYA YOGYAKARTA
2015**



UNIVERSITAS ATMA JAYA YOGYAKARTA

PROGRAM PASCASARJANA

PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA

PENGESAHAN TESIS

Nama : Ni Putu Eka Merliana
Nomor Mahasiswa : 135302010/PS/MTF
Konsentrasi : Enterprise Information System
Judul Tesis : **Perbandingan Metode K-Means Dengan Fuzzy C-Means Untuk Analisa Karakteristik Mahasiswa Berdasarkan Kunjungan Ke Perpustakaan**

Nama Pembimbing	Tanggal	Tanda tangan
Dra. Ernawati, MT	23-7-2015	
Dr. Ir. Alb. Joko Santoso, MT	23-7-2015	
Prof. Dr. Suyoto, Msc., Phd.	23-7-2015	

Ketua Program Studi

Prof. Dr. Suyoto, Msc., Phd.
PROGRAM
PASCASARJANA

HALAMAN PERNYATAAN

Saya yang bertanda tangan di bawah ini dengan sesungguhnya menyatakan bahwa tesis dengan judul :

**PERBANDINGAN METODE K-MEANS DENGAN FUZZY C-MEANS
UNTUK ANALISA KARAKTERISTIK MAHASISWA BERDASARKAN
KUNJUNGAN KE PERPUSTAKAAN**

Benar – benar hasil karya saya sendiri. Pernyataan, ide, maupun kutipan baik langsung maupun tidak langsung yang bersumber dari tulisan atau ide orang lain dinyatakan secara tertulis pada tesis ini dalam bentuk kutipan dan daftar pustaka sebagaimana layaknya karya ilmiah. Apabila di kemudian hari terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab saya.

Yogyakarta, 23 Juli 2015

Ni Putu Eka Merliana

INTISARI

Kurangnya minat membaca pada mahasiswa khususnya menjadi perhatian bagi lembaga pendidikan, karena membaca dapat menambah wawasan, pengetahuan dan informasi yang dimiliki sehingga dapat mempengaruhi prestasi akademik mahasiswa. Salah satu fasilitas yang mendukung hal tersebut adalah fasilitas perpustakaan. Penelitian ini akan menggunakan dua metode yaitu algoritma K-Means dan Fuzzy C-Means dimana dari hasil kedua algoritma tersebut akan dibandingkan untuk menentukan algoritma yang sesuai dalam menganalisa karakteristik mahasiswa berdasarkan kunjungan ke perpustakaan serta data akademik mahasiswa sehingga dapat menghasilkan informasi yang berhubungan dengan fungsi fasilitas perpustakaan dengan peningkatan kinerja akademik mahasiswa. Data yang diambil adalah data mahasiswa dari 2012-2014. Hasil penelitian dengan pengujian 885 data mahasiswa diperoleh hasil nilai cluster yang terbaik adalah metode K-Means dibandingkan Fuzzy C-Means, karena K-Means memperlihatkan adanya hubungan antara variabel baik itu dari data perpustakaan maupun data akademik mahasiswa.

Kata Kunci : Perpustakaan, Prestasi akademik, K-Means, Fuzzy C-Means

ABSTRACT

Lack of interest in reading in students especially a concern for educational institutions, because reading can add insight, knowledge and information held so as to affect students' academic achievement. One of the facilities that support it are the library facilities. This study will use two methods, K-Means algorithm and Fuzzy C-Means in which the results of the two algorithms will be compared to determine the appropriate algorithm for analyzing the characteristics of the students based on visits to the library and data student academic so as to produce information related to the function of the facility libraries with improved academic performance of students. The data is taken from the student data from 2012 to 2014. Results of the study by testing 885 student data obtained results cluster is best value K-Means method compared Fuzzy C-Means, as K-Means show an association between both variables from the data library and student academic data.

Keywords: Library, academic achievement, K-Means, Fuzzy C-Means

KATA PENGANTAR

Puji dan syukur kepada Tuhan Yang Maha Esa karena atas berkat dan rahmatNya yang melimpah sehingga penulis dapat menyelesaikan tesis ini dengan baik. Tesis disusun untuk memenuhi salah satu syarat kelulusan mahasiswa Magister Teknik Informatika Program Pascasarjana Universitas Atma Jaya Yogyakarta, tetapi juga menjadi sarana pembelajaran bagi penulis untuk melangsungkan penelitian kembali sesuai dengan kebutuhan pengguna. Penulis berharap tesis ini dapat dikembangkan lebih lanjut untuk membantu penelitian-penelitian berikutnya.

Penulis juga ingin menyampaikan terima kasih kepada semua pihak yang telah membantu dalam penyelesaian tesis secara langsung maupun tidak langsung yaitu kepada :

1. Ida Sang Hyang Widhi Wasa (Sri Krsna) yang selalu memberikan karuniaNya setiap saat kepada penulis
2. Bapak Prof. Dr. Suyoto, Msc., Phd., selaku Ketua Program Studi Magister Teknik Informatika Program Pascasarjana Universitas Atma Jaya Yogyakarta.
3. Ibu Dra. Ernawati, MT, selaku Dosen Pembimbing I yang telah banyak meluangkan waktu, tenaga dan pikiran dalam memberikan bimbingan dan bantuan yang berharga serta motivasi kepada penulis sehingga tesis ini dapat terselesaikan dengan baik.
4. Bapak Dr. Ir. Albertus Joko Santoso, M.T., selaku Dosen Pembimbing II yang telah banyak membantu dan membimbing dalam menyelesaikan tesis ini dengan baik
5. Suami dan anakku tercinta yang selalu memberikan semangat dan doa dalam penyelesaian tesis ini
6. Mama dan Papa yang selalu mendukung dan menyemangati penulis
7. Rekan – rekan Magister Teknik Informatika : Mbak Florent, Johan, Carlo, Diana, Andry dan teman-teman MTF angkatan September 2014 yang telah memberikan bantuan berupa dukungan dan semangat selama menyelesaikan tesis ini

8. Para penghuni Asrama Narayana Smriti yang telah mendukung secara spiritual
9. Citivitas Akademika STAHN Tampung Penyang Palangka Raya yang telah membantu baik itu dalam support doa dan semangat kepada penulis
10. Segenap dosen dan staf tata usaha Program Pascasarjana Universitas Atma Jaya Yogyakarta yang telah memberikan bimbingan dan bantuan demi kelancaran penulis selama menempuh pendidikan
11. Semua teman-teman dan kenalan yang tidak dapat penulis sebutkan satu persatu yang telah memberikan dorongan dan semangat yang sangat berarti.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna, karena itu penulis sangat berterima kasih atas segala kritik dan saran yang akan diberikan sehingga tesis ini bermanfaat bagi pembaca.

Akhir kata semoga laporan ini dapat berguna dan bermanfaat bagi semua pihak.

Yogyakarta , 22 Juli 2015

Penulis

DAFTAR ISI

BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Keaslian Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Tujuan Penelitian	5
1.7 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	7
2.1 Tinjauan Pustaka	7
2.2 Landasan Teori	11
2.2.1 Data Mining	11
2.2.2 Clustering	12
2.2.3 K-Means	13
2.2.4 Fuzzy C-Means	19
BAB III METODOLOGI PENELITIAN	25
3.1 Jenis Penelitian	25
3.2 Objek Penelitian	25
3.2.1 Objek Instansi	25
3.2.2 Objek Data	25
3.3 Jenis dan Sumber data	26
3.3.1 Jenis Data	26
3.3.2 Sumber Data	27
3.4 Contoh perhitungan manual pada K-Means dan Fuzzy C-Means	28
3.4.1 Metode K-Means	28
3.4.2 Metode Fuzzy C-Means	33
3.5 Langkah Penelitian	39
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	42
4.1 Analisa Sistem	42
4.1.1 Diagram Konteks	42
4.1.2 Data Flow Diagram (DFD Level 0)	43
4.1.3 DFD Level 1 Clustering	44
4.1.4 DFD Level 1 Mengelola Buku Tamu	44
4.1.5 Entity Relationship Diagram (ERD)	46
4.1.6 Rancangan Tabel	47
4.1.7 Flowchart Algoritma K-Means	48
4.1.8 Flowchart Algoritma Fuzzy C-Means	51
4.2 Analisa penentuan jumlah cluster terbaik	54

4.2.1 Uji Coba Elbow dengan K-Means	54
4.2.2 Uji Coba Elbow dengan Fuzzy C-Means	59
4.3 ANALISA HASIL CLUSTER	64
4.3.1 Analisa K-Means 885 data	64
4.3.2 Analisa Fuzzy C-Means 885 data	66
4.4 KARAKTERISTIK DATA PADA K-MEANS	69
4.4.1 Kluster I	69
4.4.2 Kluster II	72
4.4.3 Kluster III	75
BAB V KESIMPULAN DAN SARAN	79
5.1 Kesimpulan	79
5.2 Saran	80
DAFTAR PUSTAKA	81

DAFTAR TABEL

Tabel 3.1. Sampel data Clusterisasi	28
Tabel 3.2. Sampel data perhitungan manual dengan K-Means	28
Tabel 3.2. Sampel data perhitungan manual dengan K-Means	30
Tabel 3.4. Pengelompokan data berdasarkan nilai cluster terdekat	30
Tabel 3.5. Perhitungan jarak ke cluster baru	31
Tabel 3.6. Pengelompokan data berdasarkan nilai cluster terdekat	32
Tabel 3.7. Bilangan random awal	33
Tabel 3.8 Cluster I baru	34
Tabel 3.9 Cluster II baru	35
Tabel 3.10 Cluster III baru	35
Tabel 3.11 Tabel cluster baru	36
Tabel 3.12 Hasil perhitungan fungsi objektif masing-masing data	36
Tabel 3.13 Derajat keanggotaan yang baru	37
Tabel 3.14 Pengelompokan data pada Fuzzy C-Means	37
Tabel 4.1 Tabel clusterisasi	48
Tabel 4.2 Tabel buku tamu	49
Tabel 4.3. Hasil <i>Sum of Square Error</i> 54 Data dengan K-Means	55
Tabel 4.4. Hasil <i>Sum of Square Error</i> 139 Data dengan K-Means	56
Tabel 4.5. Hasil <i>Sum of Square Error</i> 261 Data dengan K-Means	58
Tabel 4.6. Hasil <i>Sum of Square Error</i> 885 Data dengan K-Means	59
Tabel 4.7 Hasil <i>Sum of Square Error</i> 54 Data dengan Fuzzy C-Means	60
Tabel 4.8. Hasil <i>Sum of Square Error</i> 139 Data dengan Fuzzy C-Means	61
Tabel 4.9. Hasil <i>Sum of Square Error</i> 261 Data dengan Fuzzy C-Means	62
Tabel 4.10. Hasil <i>Sum of Square Error</i> 885 Data dengan Fuzzy C-Means	63
Tabel 4.11 Perbandingan Hasil Cluster K-Means dan Fuzzy C-Means 885 data	67

DAFTAR GAMBAR

Gambar 3.1 Data Buku tamu unit perpustakaan STAHN-TP Palangka Raya.....	27
Gambar 3.2. Bagan Alir Penelitian	39
Gambar 4.1 Diagram konteks	43
Gambar 4.2 Diagram Level 0	43
Gambar 4.3 Diagram Level 1 Clustering	44
Gambar 4.4 Diagram Level 1 Mengelola Buku Tamu	45
Gambar 4.5 <i>Entity Relationship Diagram</i> (ERD)	47
Gambar 4.6 <i>Flowchart</i> algoritma K-Means	51
Gambar 4.7 <i>Flowchart</i> algoritma Fuzzy C-Means	54
Gambar 4.8 Grafik <i>Sum of Squared Error</i> 54 Data dengan K-Means	56
Gambar 4.9 Grafik <i>Sum of Squared Error</i> 139 Data dengan K-Means	57
Gambar 4.10 Grafik <i>Sum of Squared Error</i> 261 Data dengan K-Means	58
Gambar 4.11. Grafik <i>Sum of Squared Error</i> 885 Data dengan K-Means	59
Gambar 4.12. Grafik <i>Sum of Squared Error</i> 54 Data dengan Fuzzy C-Means	61
Gambar 4.13. Grafik <i>Sum of Squared Error</i> 139 Data dengan Fuzzy C-Means	62
Gambar 4.14. Grafik <i>Sum of Squared Error</i> 261 Data dengan Fuzzy C-Means	63
Gambar 4.15. Grafik <i>Sum of Squared Error</i> 885 Data dengan Fuzzy C-Means	64
Gambar 4.16. Hasil Cluster K-Means 885 data mahasiswa	65
Gambar 4.17. Hasil Cluster Fuzzy C-Means 885 data mahasiswa	66
Gambar 4.18. Grafik Hasil Cluster K-Means 885 data	68
Gambar 4.19. Grafik Hasil Cluster Fuzzy C-Means 885	68
Gambar 4.20. Hasil Cluster K-Means	70
Gambar 4.21. Grafik kluster I berdasarkan jenis kelamin	72
Gambar 4.22. Grafik kluster I berdasarkan program studi	72
Gambar 4.23. Grafik kluster I berdasarkan IPK	73
Gambar 4.24. Grafik kluster II berdasarkan IPK	74
Gambar 4.25. Grafik kluster II berdasarkan Program Studi	75
Gambar 4.26. Grafik kluster II berdasarkan IPK	75
Gambar 4.27. Grafik kluster III berdasarkan jenis kelamin	77
Gambar 4.28. Grafik kluster III berdasarkan program studi	77
Gambar 4.29. Grafik kluster III berdasarkan IPK	78

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kebutuhan membaca pada masyarakat Indonesia tergolong sangat rendah, ini dikarenakan masyarakat Indonesia belum menempatkan kebutuhan membaca sebagai hal yang penting. Dimana dengan membaca wawasan akan menjadi luas, pengetahuan yang diperoleh akan semakin banyak dan informasi yang didapat juga semakin banyak. Dalam hal ini mahasiswa yang sering melakukan aktifitas membaca akan dapat mempengaruhi prestasi akademik yang dimiliki karena berhubungan dengan pengetahuan dan wawasan yang dimiliki. Ada beberapa faktor yang mempengaruhi kinerja atau prestasi akademik seorang mahasiswa yaitu dilihat dari faktor *internal* dan faktor *eksternal*. Faktor *Internal* antara lain meliputi komunikasi mahasiswa di kelas, jumlah buku yang dimiliki, pengetahuan dan wawasan yang dimiliki, cara dosen mengajar di kelas, fasilitas dalam mengajar, lingkungan kelas, sistem ujian dan teknologi yang digunakan di kelas. Sedangkan faktor *eksternal* meliputi kegiatan ekstrakurikuler, masalah keluarga, pekerjaan dan keuangan, sosial dan masalah lainnya (Mushtaq & Khan, 2012). Secara umum *indikator* yang digunakan untuk melihat prestasi akademik seorang mahasiswa adalah Indeks Prestasi Kumulatif (IPK) yang diperoleh dari beberapa faktor yaitu dilihat dari nilai kuis, ujian tengah semester dan ujian akhir semester (O. J et al., 2010; Shovon & Haque, 2012).

Perpustakaan menjadi tempat yang digunakan untuk proses pembelajaran yang dapat mempengaruhi peningkatan prestasi akademik seseorang, ini disebabkan karena pengetahuan dan wawasan yang didapat dari membaca (Basheer & Razzaq, 2012). STAHN-TP Palangka Raya merupakan salah satu lembaga pendidikan berbasis agama Hindu yang memiliki beberapa fasilitas pendukung salah satunya adalah perpustakaan. Fasilitas tersebut belum berbasis *online*, sehingga dalam melakukan peminjaman buku, mahasiswa STAHN-TP langsung datang ke perpustakaan, tapi ada juga beberapa mahasiswa yang belum menggunakan fasilitas perpustakaan yang disiapkan oleh pihak kampus. Penelitian ini akan menganalisa karakteristik mahasiswa berdasarkan kunjungan ke perpustakaan serta data akademik mahasiswa. Analisa karakteristik tersebut akan mengelompokkan data mahasiswa berdasarkan prestasi akademiknya sehingga menghasilkan informasi yang berhubungan dengan fungsi fasilitas perpustakaan dengan peningkatan kinerja akademik mahasiswa.

Dalam memisahkan data yang memiliki karakteristik yang berbeda dan mengelompokkan data yang memiliki karakteristik yang sama dipergunakan metode *clustering*. Penelitian ini akan menggunakan dua metode *clustering* yaitu algoritma K-Means dan algoritma Fuzzy C-Means dimana hasil dari kedua algoritma tersebut akan dibandingkan untuk menentukan algoritma yang sesuai untuk menganalisa karakteristik mahasiswa berdasarkan kunjungan mahasiswa ke perpustakaan.

1.2 Rumusan Masalah

Dari uraian latar belakang diatas, maka rumusan masalah dalam penelitian ini adalah

- a. Bagaimana hubungan data akademik mahasiswa dengan data kunjungan ke perpustakaan ?
- b. Diantara dua metode *clustering* yaitu algoritma K-Means dengan Fuzzy C-Means, metode mana yang memberikan informasi yang terkait dengan hubungan antara prestasi akademik mahasiswa dengan data kunjungan ke perpustakaan ?
- c. Bagaimana cara mengembangkan aplikasi yang sudah dimiliki oleh pihak perpustakaan pada STAHN Tampung Penyang Palangka Raya untuk mengetahui karakteristik mahasiswa ?

1.3 Batasan Masalah

Permasalahan pada topik dalam penelitian ini akan dibatasi oleh hal-hal sebagai berikut :

- a. Penelitian ini membahas tentang analisa karakteristik mahasiswa dengan melihat data kunjungan ke perpustakaan dan data akademik mahasiswa dengan beberapa variable pendukung yaitu Indeks Prestasi Kumulatif (IPK), jumlah buku yang dipinjam, serta jumlah kunjungan mahasiswa ke perpustakaan.

- b. Untuk menghasilkan informasi yang diperlukan bagi STAHN Tampung Penyang maka penelitian ini akan mengambil sampel data dari tahun akademik 2011/2012, 2012/2013, dan 2013/2014.
- c. Penelitian ini menggunakan dua metode algoritma *clustering* yaitu algoritma Fuzzy C-Means dan algoritma K-Means untuk membandingkan metode mana yang lebih baik dalam memberikan hasil *cluster* sehingga dapat membantu dalam menganalisa karakteristik mahasiswa

1.4 Keaslian Penelitian

Keaslian penelitian dengan judul “Perbandingan Metode K-Means Dengan Metode Fuzzy C-Means Untuk Analisa Karakteristik Mahasiswa Berdasarkan Kunjungan Ke Perpustakaan” adalah benar adanya. Keaslian ini dapat dibuktikan dengan belum ditemukannya buku, artikel, atau jurnal ilmiah yang ditulis oleh penulis lain, walaupun pada beberapa bagian dari isi penelitian ini mengutip pada penelitian serupa yang pernah dilakukan. Tata cara pengutipan yang dilakukan telah mengikuti tata cara dan etika penulisan karya ilmiah yang lazim.

1.5 Manfaat Penelitian

Beberapa manfaat yang akan diperoleh dari penelitian ini adalah sebagai berikut:

- a. Dapat membantu dalam memberikan informasi yang digunakan untuk mengetahui hubungan antara prestasi akademik dengan kunjungan mahasiswa ke perpustakaan.
- b. Memberikan kemudahan kepada Sekolah Tinggi Agama Hindu Negeri Tampung Penyang dalam memberikan informasi yang berhubungan dengan manfaat dalam penggunaan perpustakaan
- c. Memberikan gambaran terhadap metode yang dapat memberikan hasil *cluster* yang baik untuk membantu dalam menganalisa karakteristik mahasiswa berdasarkan kunjungan mahasiswa ke perpustakaan dan data akademik mahasiswa

1.6 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

- a. Dapat mengetahui karakteristik mahasiswa dalam hubungan antara penggunaan fasilitas perpustakaan dengan pengaruh terhadap prestasi akademik mahasiswa
- b. Dapat mengetahui metode yang dapat memberikan nilai *cluster* yang baik sehingga dapat membantu dalam menganalisa karakteristik mahasiswa berdasarkan kunjungan mahasiswa ke perpustakaan dan data akademik mahasiswa
- c. Mendapatkan informasi yang berguna bagi pihak STAHN Tampung Penyang Palangka Raya untuk menganalisa karakteristik mahasiswa.

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam laporan penelitian ini terbagi menjadi lima bab yaitu :

a. BAB I PENDAHULUAN

Bab ini memuat tentang latar belakang, rumusan masalah, batasan masalah, keaslian penelitian, manfaat penelitian, tujuan penelitian dan sistematika penulisan.

b. BAB II TINJAUAN PUSTAKA

Bab ini menguraikan penelitian-penelitian terdahulu (tinjauan pustaka) dan teori-teori yang mendukung penelitian (landasan teori).

c. BAB III METODOLOGI PENELITIAN

Bab ini memuat tentang metode dan langkah-langkah yang akan digunakan pada penelitian tesis.

d. BAB IV PEMBAHASAN

Bab ini memuat tentang hasil serta pembahasan dari penelitian yang dilakukan. Pembahasan berisikan tentang analisa data yang diperoleh baik itu secara kualitatif, kuantitatif maupun normatif. Selain itu pada bab ini juga menjelaskan tentang analisa, perancangan serta implementasi dari aplikasi yang dibangun.

e. BAB V KESIMPULAN DAN SARAN

Bab ini memuat tentang kesimpulan dari hasil penelitian yang telah dilakukan serta saran dalam pengembangan penelitian lebih lanjut.