

PENGELOMPOKAN MINAT BACA MAHASISWA MENGUNAKAN METODE K-MEANS

Widya Safira Azis¹ dan Dedy Atmajaya²

¹safiraazis18@gmail.com dan ²dedy.atmajaya@umi.ac.id

^{1,2}Teknik Informatika Universitas Muslim Indonesia

Abstrak

Pengelompokan minat baca mahasiswa berdasarkan kriteria buku yang dibaca, buku yang dipinjam, dan juga mempertimbangkan jumlah stok buku yang tersedia dapat membantu dalam proses penambahan koleksi buku yang telah ada pada perpustakaan Utsman Bin Affan UMI. Salah satu cara untuk mengelola data tersebut yaitu menggunakan data mining dengan memanfaatkan metode K-Means. Data buku dikelompokkan menjadi 3 cluster yaitu prioritas, dipertimbangkan, dan tidak diprioritaskan dalam perencanaan penambahan koleksi buku. Hasil dari penelitian ini adalah cluster dengan nilai terbesar pada centroid akhir merupakan cluster yang direkomendasikan dalam perencanaan penambahan koleksi buku.

Kata kunci: *clustering*, *k-means*, data, perpustakaan

Copyright © 2016 -- Jurnal Ilmiah ILKOM -- All rights reserved.

1. Pendahuluan

Perpustakaan adalah suatu unit kerja yang berupa tempat menyimpan koleksi bahan pustaka yang diatur secara sistematis dengan cara tertentu untuk dipergunakan secara berkesinambungan oleh pemakainya sebagai sumber informasi [2].

Perpustakaan Utsman Bin Affan saat ini, memiliki koleksi buku ± 22.540 Judul 82.285 Exp, dalam bidang ilmu : karya umum, filsafat dan psikologi, agama, ilmu-ilmu sosial, bahasa, ilmu-ilmu murni (pasti/alam), ilmu-ilmu terapan, kesenian, hiburan, olahraga, kesusastraan, geografi dan sejarah umum. Agar koleksi buku di dalam perpustakaan dapat dimanfaatkan secara maksimal, harus diterapkan suatu sistem untuk mengatur koleksi buku tersebut. Saat ini, sistem yang diterapkan di perpustakaan Utsman Bin Affan Universitas Muslim Indonesia berjalan kurang efektif. Ini terlihat dari mulai tahap perencanaan yang meliputi pemilihan hingga penambahan koleksi buku yang tidak maksimal.

Dalam proses penambahan koleksi buku, pihak perpustakaan hanya menjadikan statistik peminjaman buku terlaris sebagai acuan dalam peningkatan jumlah koleksi buku yang telah ada, tanpa mempertimbangkan buku mana yang menjadi prioritas utama untuk ditingkatkan koleksinya. Padahal, dari tahap pemilihan koleksi, dapat dilihat kualitas perpustakaan dalam memberikan informasi kepada pengguna. Sehingga, secara tidak langsung, dapat menimbulkan pengaruh yang sangat besar, yaitu mulai berkurangnya minat baca dari pengguna perpustakaan dalam memanfaatkan koleksi yang ada.

Dalam memaksimalkan proses penambahan koleksi buku di perpustakaan Utsman Bin Affan UMI, maka dibutuhkan sebuah sistem yang digunakan untuk mengelola data tersebut secara komputerisasi. Salah satu cara untuk mengelola data tersebut yaitu dengan menggunakan *data mining*. Dengan memanfaatkan metode *k-means* pada *data mining*, setiap data buku akan dikelompokkan menjadi 3 *cluster*. yaitu prioritas tinggi, dipertimbangkan, dan bukan prioritas untuk perencanaan penambahan koleksi buku. Hasil yang didapatkan yaitu, untuk data buku yang berada pada *cluster* prioritas tinggi, inilah yang akan dijadikan sebagai bahan evaluasi bagi pihak perpustakaan dalam meningkatkan koleksi buku yang sudah ada.

Tujuan utama dari penelitian ini ialah untuk mendesain aplikasi pengelompokan minat baca pada Perpustakaan Utsman Bin Affan UMI, yang diharapkan dapat membantu pihak perpustakaan dalam memecahkan masalah penambahan koleksi buku dengan tepat dan cepat.

2. Landasan Teori

2.1. Algoritma K-Means Clustering

K-Means termasuk dalam *partitioning clustering* yaitu setiap data harus masuk dalam *cluster* tertentu dan memungkinkan bagi setiap data yang termasuk dalam *cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* yang lain. *K-Means* memisahkan data ke *k* daerah

bagian yang terpisah, dimana k adalah bilangan integer positif. Algoritma *K-Means* sangat terkenal karena kemudahan dan kemampuannya untuk mengkluster data besar dan *outlier* dengan sangat cepat.

Secara sederhana algoritma *K-Means* adalah sebagai berikut [5]:

1. Tentukan k sebagai jumlah *cluster* yang ingin dibentuk.
2. Bangkitkan k *centroid* (titik pusat *cluster*) awal secara *random* atau ditentukan secara *default*.
3. Hitung masing-masing jarak setiap data ke masing-masing *centroid*.
4. Setiap data memilih *centroid* yang terdekat.
5. Tentukan posisi *centroid* baru dengan cara menghitung nilai rata-rata dari data-data yang memilih pada *centroid* yang sama.
6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama [1].

2.2. Penerapan Algoritma K-Means

Algoritma *K-Means*:

1. Membaca data buku yang telah dikategorikan.
2. Menentukan jumlah atribut = 3, dan jumlah cluster = 3.
3. Menset nilai awal centroid secara default.
4. Melakukan klusterisasi dengan mengelompokkan data berdasarkan jarak terdekat.
5. Jika data buku pinjam, baca dan stok buku ditambahkan, maka akan terjadi perubahan data klusterisasi. Jika kondisi memenuhi maka sistem akan melakukan klusterisasi kembali dengan mengelompokkan data berdasarkan jarak terdekat.
6. Klusterisasi akan dilakukan sampai hasilnya konvergen dan tidak ada lagi perpindahan data.

2.3. Pengukuran Kemiripan dengan Jarak atau *Distance*

Langkah penting dalam *clustering* adalah menentukan perhitungan jarak / *distance measure*, yang sangat menentukan perhitungan bagaimana kemiripan antar dua elemen. Untuk menghitung jarak / *distance measure*, dapat menggunakan rumus *Euclidean Distance* yaitu [7] :

$$d(x, y) = \sqrt{(x_1 - s_1)^2 + (y_2 - t_2)^2} \dots \dots \dots (1)$$

Pada persamaan 1,

(x,y) : titik koordinat objek

(s,t) : titik koordinat *centroid*

$d(x,y)$: *Euclidean distance* yaitu jarak antara data pada titik x dan titik y menggunakan kalkulasi matematika (Bezdek, 1981). Misalnya : titik data bisa didefinisikan sebagai **Person** dalam *database*, sementara atributnya adalah **Age, Height, Weight, Income**.

3. Metode

Pada penelitian ini, dilakukan beberapa tahapan yang saling berkaitan antara satu tahap dengan tahap lainnya. *Flowchart* tahapan penelitian yang dilakukan dapat dilihat pada gambar 1 [3].



Gambar 1. *Flowchart* Tahapan Penelitian [3]

Berdasarkan Gambar 1. dapat dilihat bahwa tahap pertama yang dilakukan dalam penelitian ini adalah identifikasi masalah. Identifikasi masalah dilakukan untuk melihat dengan jelas tujuan dan sasaran dari penelitian yang akan dilakukan. Masalah penelitian dipilih berdasarkan beberapa pertimbangan, antara lain sisi waktu, biaya maupun kontribusi diberikan bagi pengembangan ilmu pengetahuan dan teknologi. Tahap selanjutnya adalah perumusan masalah. Perumusan masalah perlu dilakukan agar proses penelitian lebih terarah sehingga dapat memberikan petunjuk untuk pengumpulan data serta pencarian metode yang tepat untuk pemecahan masalah tersebut. Setelah merumuskan masalah, tahap berikutnya yang dilakukan adalah penelusuran pustaka. Pada tahap ini dilakukan pencarian pustaka dan pengumpulan informasi yang berhubungan dengan penelitian. Penelusuran pustaka berguna untuk menghindari duplikasi dari pelaksanaan penelitian.

Tahap berikutnya yaitu rancangan penelitian. Rancangan penelitian merupakan rencana menyeluruh dari penelitian mencakup hal-hal yang akan dilakukan dalam penelitian ini. Adapun rancangan penelitian dari penelitian ini adalah rentang waktu penelitian dilakukan, teknik pengumpulan data dan pengolahan data. Data yang dikumpulkan dalam tahap pengumpulan data ini merupakan data sekunder yang diperoleh melalui wawancara di Perpustakaan Utsman Bin Affan untuk mendapatkan data yang diperlukan dalam penelitian ini dan data primer yaitu data yang berasal dari pihak yang bersangkutan atau langsung dari pegawai yang bertugas. Data tersebut kemudian diolah yang berfungsi untuk sistem yang akan dibangun. Hasil dari pengolahan data yang dilakukan melalui proses pengujian akan digunakan dalam tahap penyimpulan hasil sebagai kesimpulan dari penelitian.

4. Hasil

Perhitungan dalam modul *k-means* menggunakan data yang terdiri dari 10 data dengan kategori buku dengan kriteria : buku yang dipinjam, buku yang dibaca dan stok buku. Percobaan dilakukan dengan menggunakan parameter-parameter berikut : Jumlah *cluster* : 3, Jumlah data : 10.

Tabel 1. Data Kategori Buku

No	No Kelas Buku	Kategori buku	Stok	Pinjam	Baca
1	0	Karya Umum	33	3	1
2	100	Filsafat dan Psikologi	22	1	1
3	200	Agama	146	17	9
4	300	Ilmu Sosial	102	12	1
5	400	Bahasa	18	5	2
6	500	Ilmu Murni	34	3	1
7	600	Ilmu Terapan	301	26	8
8	700	Kesenian, Hiburan, Olah Raga	11	2	2
9	800	Kesusastraan	9	1	2
10	900	Geografi dan Sejarah Umum	2	1	2

Tabel 1. Merupakan data yang digunakan untuk melakukan percobaan perhitungan manual.

Berikut adalah langkah-langkah algoritma dalam modul K-Means:

1. Pusat awal klaster didapatkan dari data sendiri bukan dengan menentukan titik baru. Untuk kasus ini terdapat 3 pusat data yaitu centroid 1, centroid 2, centroid 3. Centroid 1 = Rendah, Centroid 2 = Sedang, Centroid 3 = Tinggi.
2. Perhitungan jarak dengan pusat klaster Untuk mengukur jarak antara data dengan pusat klaster digunakan *Euclidian distance*. Algoritma perhitungan jarak data dengan pusat klaster
 - a. Ambil nilai data dan nilai pusat klaster
 - b. Hitung *Euclidian distance* data dengan tiap pusat klaster
3. Pengelompokkan data jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat klaster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat klaster terdekat. Algoritma pengelompokkan data :
 - a. Ambil nilai jarak tiap pusat klaster dengan data
 - b. Cari nilai jarak terkecil
 - c. Kelompokkan data dengan pusat klaster yang memiliki jarak terkecil.

4. Penentuan pusat kluster baru Untuk mendapatkan pusat kluster baru bisa dihitung dari rata-rata nilai anggota kluster dan pusat kluster. Pusat kluster yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi akan berhenti jika telah memenuhi maksimum iterasi yang dimasukan oleh user atau hasil yang dicapai sudah konvergen (pusat kluster baru sama dengan pusat kluster lama).

Dari algoritma di atas maka dapat dilakukan perhitungan sebagai berikut :

Penentuan Pusat Kluster Awal

1. Untuk pusat *cluster* buku pinjam:
Rendah : 2, sedang : 4, tinggi : 6.
2. Untuk pusat *cluster* buku baca :
Rendah : 2, sedang : 4, tinggi : 6.
3. Untuk pusat *cluster* stok :
Rendah : 100, sedang : 200, tinggi : 300.

Hasil Perhitungan Iterasi-1 :

Tabel 2. Iterasi - 1

No	No Kelas	Nama	Centroid 1			Centroid 2			Centroid 3			Centroid 1	Centroid 2	Centroid 3
			300.0	2.0	2.0	200.0	4.0	4.0	100.0	6.0	6.0			
1	0	Karya Umum	33			3			1					Ok
2	100	Filsafat dan Psikologi	22			1			1					Ok
3	200	Agama	146			17			9					Ok
4	300	Ilmu Sosial	102			12			1					Ok
5	400	Bahasa	18			5			2					Ok
6	500	Ilmu Murni	34			3			1					Ok
7	600	Ilmu Terapan	301			26			8		Ok			
8	700	Kesenian, Hiburan, Olah Raga	11			2			2					Ok
9	800	Kesusastraan	9			1			2					Ok
10	900	Geografi dan Sejarah Umum	2			1			2					Ok

Untuk menghitung jarak pusat kluster, menggunakan rumus *Euclidean Distance* [4].

$$d = |x - y| \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \dots \dots \dots (2)$$

- Ket :
- d = perhitungan jarak terhadap pusat cluster
 - x = titik koordinat objek (jumlah pinjam, baca, dan stok)
 - y = titik koordinat centroid (nilai centroid untuk pinjam, baca dan stok)
 - $\sum_{i=1}^n$ = n merupakan banyaknya data yang akan diukur jaraknya, sedangkan i-1 adalah proses klusterisasi dimulai dari iterasi pertama.
 - x_i = titik koordinat objek ke-i
 - y_i = titik koordinat centroid ke-i

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat kluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat kluster terdekat. Dari hasil perhitungan didapatkan jarak terdekat dengan pusat data yaitu 1 pada centroid 1, 0 pada centroid 2, dan 9 pada centroid 3 sehingga terbentuklah centroid baru.

Rumus yang digunakan untuk mencari centroid baru. yaitu [4]:

$$\frac{x_1+x_2+\dots+x_n, y_1+y_2+\dots+y_n}{n} \dots \dots \dots (3)$$

- Ket :
- $x_1 + x_2 + \dots + x_n$ adalah titik koordinat x pada data yang tergabung dalam suatu centroid.
 - $y_1 + y_2 + \dots + y_n$ adalah titik koordinat y pada data yang tergabung dalam suatu centroid.
 - n adalah banyaknya data yang tergabung dalam suatu pusat data (centroid).

Hasil Perhitungan Iterasi - 2 :

Tabel 2. Iterasi - 2

No	No Kelas	Nama	Centroid 1			Centroid 2			Centroid 3			Centroid 1	Centroid 2	Centroid 3
			301.0	26.0	8.0	0.0	0.0	0.0	41.9	5.0	2.3			
1	0	Karya Umum	269.1			33.2			9.2					Ok
2	100	Filsafat dan Psikologi	280.2			22.0			20.3					Ok
3	200	Agama	155.3			147.3			105.0					Ok
4	300	Ilmu Sosial	199.6			102.7			60.5					Ok
5	400	Bahasa	283.8			18.8			23.9			Ok		
6	500	Ilmu Murni	268.1			34.2			8.3					Ok
7	600	Ilmu Terapan	0.0			302.2			260.0		Ok			
8	700	Kesenian, Hiburan, Olah Raga	291.1			11.4			31.0			Ok		
9	800	Kesusastraan	293.1			9.3			33.1				Ok	
10	900	Geografi dan Sejarah Umum	300.1			3.0			40.1				Ok	

Hasil Perhitungan Iterasi - 3 :

Tabel 3. Iterasi - 3

No	No Kelas	Nama	Centroid 1			Centroid 2			Centroid 3			Centroid 1	Centroid 2	Centroid 3
			301.0	26.0	8.00	10.0	2.2	2.00	67.4	7.20	2.60			
1	0	Karya Umum	199.61			92.52			34.97					Ok
2	100	Filsafat dan Psikologi	155.26			136.98			79.47					Ok
3	200	Agama	280.21			12.11			45.85			Ok		
4	300	Ilmu Sosial	269.08			23.03			34.69			Ok		
5	400	Bahasa	283.84			8.46			49.45			Ok		
6	500	Ilmu Murni	268.08			24.03			33.70			Ok		
7	600	Ilmu Terapan	0.00			292.03			234.42		Ok			
8	700	Kesenian, Hiburan, Olah Raga	291.05			1.03			56.64			Ok		
9	800	Kesusastraan	293.13			1.60			58.73				Ok	
10	900	Geografi dan Sejarah Umum	300.10			8.10			65.70				Ok	

Hasil Perhitungan Iterasi - 4 :

Tabel 4. Iterasi - 4

No	No Kelas	Nama	Centroid 1			Centroid 2			Centroid 3			Centroid 1	Centroid 2	Centroid 3
			301.0	26.0	8.00	18.4	2.2	1.57	124.0	14.5	5.00			
1	0	Karya Umum	199.61			84.14			22.50					Ok
2	100	Filsafat dan Psikologi	155.26			128.63			22.50					Ok
3	200	Agama	280.21			3.84			102.97			Ok		
4	300	Ilmu Sosial	269.08			14.60			91.81			Ok		
5	400	Bahasa	283.84			2.78			106.47			Ok		
6	500	Ilmu Murni	268.08			15.60			90.82			Ok		
7	600	Ilmu Terapan	0.00			283.64			177.40		Ok			
8	700	Kesenian, Hiburan, Olah Raga	291.05			7.45			113.73			Ok		
9	800	Kesusastraan	293.13			9.53			115.83				Ok	
10	900	Geografi dan Sejarah Umum	300.10			16.48			122.78				Ok	

Pada iterasi-1 sampai iterasi-4, rumus yang digunakan sama. Iterasi akan terus berlangsung sampai hasilnya konvergen. Dimana jika jarak data antara titik pusat yang sama pada suatu iterasi. Hasil percobaan ini dapat terlihat pada iterasi-3 dan iterasi-4, dimana anggota cluster pada iterasi-3 dan iterasi-4 menunjukkan hasil yang konvergen.

5. Kesimpulan dan Saran

5.1 Simpulan

Hasil penelitian menunjukkan bahwa pada Perpustakaan Utsman Bin Affan UMI belum menerapkan sistem komputerisasi secara maksimal, hal ini terlihat dengan masih digunakannya perhitungan secara manual dalam penyusunan data statistik peminjaman. Dengan adanya aplikasi pengelompokan minat baca mahasiswa menggunakan metode klasterisasi pada Perpustakaan Utsman Bin Affan UMI, pengelolaan dan proses komunikasi dalam instansi menjadi lebih efektif, cepat dan tepat. Dimana aplikasi ini, *manage* data buku yang dipinjam dan data buku yang dibaca. Tentunya data ini yang akan menjadi data olahan untuk proses klasterisasi *k-means*. Hasil klasterisasi dijadikan acuan untuk penambahan koleksi buku di perpustakaan.

5.2 Saran

Untuk penelitian yang sejenis dalam menggunakan metode klasterisasi (*k-means*), dapat digunakan beberapa variasi lain dari *k-means*. Dapat juga membandingkan dengan metode klasterisasi lainnya, sehingga diperoleh hasil pemodelan yang maksimal. Sedangkan untuk pengembangan aplikasi disarankan untuk melakukan sinkronisasi dengan sistem yang ada secara *real-time* sehingga diperoleh data yang akurat.

Daftar Pustaka

- [1] Agusta, Yudi. 2007. "K-Means – Penerapan , Permasalahan dan Metode Terkait", Jurnal Sistem dan Informatika Vol.3. hal 47-60, Stikom Bali.
- [2] C. Larasati Milburga, et.all. 1986. "Membina Perpustakaan Sekolah", Yogyakarta: Kanisius, hlm. 17
- [3] Hasibuan, Zainal A, 2007, Metode Penelitian Pada Bidang Ilmu Komputer Dan Teknologi Informasi : Konsep, Teknik Dan Aplikasi, Jakarta : Fakultas
- [4] Ibrahim, Adhi Yusran. 2014. "Penerapan Klasterisasi K-Means Dalam Penentuan Penerima Beasiswa Bidikmisi Pada Universitas Muslim Indonesia". FTI, Teknik Informatika, Universitas Muslim Indonesia.
- [5] J. C. Bezdek. 1981. "Pattern Recognition with Fuzzy Objective Function Algorithms". New York : Plenum Press.