

Analisis Clustering Algoritma K-Means Sebagai Rekomendasi Penambahan Koleksi Buku Di Perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun

Taufik Firmansyah^{1,*}, Poningsih², Sundari Retno Andani³

¹STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

^{2,3}AMIK Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

Email: ¹taufikfirmansyah2021@gmail.com

Abstract

This study aims to determine the classification of reading interest based on the books read and the number of book stocks in the library of Madrasah Tsanawiyah Negeri 2 Simalungun. The data will be taken from the book borrowing archive in the library. To manage the data, the researcher uses data mining, Microsoft Excel and Rapidminer by utilizing the k-means algorithm method. The results of this study will be taken from the cluster with the largest value at the final centroid in order to recommend the number of additions to the stock of books that are in great demand by students in the library of Madrasah Tsanawiyah Negeri 2 Simalungun, and the results of cluster c1 are 5 books and c2 are 70 books.

Keywords: Data Mining, Library, k-means Algorithm, Clustering, Knowledge Discovery in Database

Abstrak

Penelitian ini bertujuan untuk menentukan klasifikasi minat baca berdasarkan buku yang dibaca dan jumlah stok buku yang ada di perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun. Data akan diambil dari arsip peminjaman buku di perpustakaan. Untuk mengelola data tersebut peneliti menggunakan data mining, Microsoft Excel dan Rapidminer dengan memanfaatkan metode algoritma k-means. Hasil dari penelitian ini akan diambil dari cluster dengan nilai terbesar pada centroid akhir guna untuk merekomendasikan jumlah penambahan stok buku yang banyak diminati siswa di perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun, dan diperoleh hasil cluster c1 sebanyak 5 buku dan c2 sebanyak 70 buku.

Kata Kunci: Data Mining, Perpustakaan, Algoritma K-Means, Clustering, Knowledge Discovery in Database

1. PENDAHULUAN

Perpustakaan merupakan sumber belajar untuk memelihara dan meningkatkan efisiensi dan efektifitas proses belajar mengajar. Perpustakaan yang terorganisir secara baik dan sistematis, secara langsung ataupun tidak langsung dapat memberikan kemudahan bagi proses belajar mengajar disekolah tempat perpustakaan tersebut berada. Hal ini terkait dengan kemajuan bidang pendidikan dan dengan adanya perbaikan metode belajar mengajar yang dirasakan tidak bisa dipisahkan dari masalah penyediaan fasilitas dan sarana pendidikan, dalam pelayanan perpustakaan harus dilengkapi fasilitas yang memadai agar tujuan dan fungsi perpustakaan dapat terpenuhi. Madrasah Tsanawiyah Negeri 2 Simalungun merupakan lembaga pendidikan yang sederajat dengan sekolah lanjutan menengah pertamayang beralamatkan di jalan asahan km.13 senio kabupaten simalungun yang memiliki ciri islam yang dikelola dan dikembangkan dibawah naungan Kementrian Agama Republik Indonesia, Dalam menentukan jumlah stok buku yang akan ditambah dari arsip data pinjaman buku oleh perangkat perpustakaan MTsN 2 Simalungun masih menggunakan sistem pengarsipan secara manual. Dalam proses pengarsipan data perpustakaan MTsN 2 Simalungun masih memiliki kendala dalam menentukan jumlah buku yang akan ditambah, Kendala yang sering terjadi yaitu waktu yang dibutuhkan pada saat pendataan, Karena pengumpulan data masih menggunakan metode pengarsipan berkas yang berbentuk buku dalam menentukan jumlah stok buku yang akan ditambah tidak terurai secara sistematis dan terkadang masih tidak tepat sasaran.

Pada penelitian ini, penulis akan membuat solusi dengan menggunakan metode algoritma *k-means*. [1] Menggunakan metode k-means dalam Pengelompokan Minat Baca Mahasiswa Menggunakan Metode *K-means*. [2] Juga menggunakan metode kmeans dalam Pengelompokkan Keaktifan Peminjaman Buku di Perpustakaan STMIK Widya Pratama, dimana Algoritma *K-means* dapat digunakan untuk mengelompokkan keaktifan peminjaman buku dengan variabel IPK, dan jumlah pinjam. Data yang digunakan diperoleh dari arsip pinjaman buku tahun ajaran 2018. Tujuan dari penelitian ini adalah membantu pihak sekolah khususnya pihak perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun dalam menganalisis jumlah pembaca buku sehingga dapat dilakukan rencana rekomendasi penambahan buku di perpustakaan. Adapun yang menjadi kontribusi dari hasil penelitian ini yaitu untuk mengkluster jumlah buku yang ada di perpustakaan menggunakan metode algoritma *k-means* dan dibantu dengan software *Rapidminer*, yang diharapkan dapat mempermudah pustakawan(petugas perpustakaan) dalam menentukan jumlah stok buku yang harus ditambah dengan tepat sasaran.

2. METODOLOGI PENELITIAN

2.1. Data Mining

Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan bermanfaat yang tersimpan didalam database besar. Data mining adalah bagian dari proses KDD (*Knowledge Discovery in Database*) yang terdiri dari beberapa tahapan seperti pemilihan data, prapengolahan, transformasi, data mining, dan evaluasi hasil [3].

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan [4].

2.2. Algoritma K-Means

K-means merupakan salah satu algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/*clustering* suatu data. Ada banyak pendekatan untuk membuat *cluster*, diantaranya adalah membuat aturan yang mendikte keanggotaan dalam group yang sama berdasarkan tingkat persamaan diantara anggota-anggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah *clustering*. Metode *k-means* adalah metode yang termasuk dalam algoritma *clustering* berbasis jarak yang membagi data ke dalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numerik [5].

Berikut langkah-langkah melakukan *clustering* dengan metode algoritma *k-means* [6] adalah sebagai berikut :

- Pilih jumlah *cluster* k .
- Inisialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara, Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka random.
- Alokasi semua data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut :

$$D(ij) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Dimana :

$D(Ij)$ = Jarak data ke I ke pusat *cluster* j

X_{ki} = Data ke i atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

- Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data /objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.

$$R_K = \frac{1}{N_K} (x_{1k} + x_{1k} + \dots + x_{nk}) \quad (2)$$

Dimana :

R_K = Rata-rata baru

N_K = Jumlah training pattern pada *cluster* (k)

X_{nk} = Pola ke (n) yang menjadi bagian *cluster* (k)

Tugaskan lagi tiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau kembali ke langkah nomor 3.

3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan di Madrasah Tsanawiyah Negeri 2 Simalungun, Dengan menggunakan arsip pinjaman buku dipergustakaan untuk mendapatkan data sebagai rekomendasi penambahan buku. Pada penelitian ini penulis memperoleh data sebanyak 75 buku pada tahun 2018. Berikut ini merupakan data yang diperoleh di perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun pada tahun 2018.

Tabel 1. Data Pinjaman Buku Tahun 2018

Jumlah Pembaca Buku Perpustakaan Tahun 2019 Di Madrasah Tsanawiyah Negeri 2 Simalungun														
No	Nama Buku	01	02	03	04	05	06	07	08	09	10	11	12	Jlh
1	Semesta Sebelum Dunia	2	1	0	1	0	0	0	0	0	0	0	0	4
2	Ilmu Pengetahuan Alam	6	6	1	7	0	0	0	2	0	2	7	0	31
3	Buah Kesabaran Dalam Iman	1	0	0	0	0	0	0	0	0	0	0	0	1
4	Al-Qur'an	8	1	0	5	0	0	8	4	5	0	0	0	31
5	Mc & Pidato	1	0	0	0	0	0	0	0	0	0	0	0	1

Jumlah Pembaca Buku Perpustakaan Tahun 2019 Di Madrasah Tsanawiyah Negeri 2 Simalungun														
No	Nama Buku	01	02	03	04	05	06	07	08	09	10	11	12	Jlh
6	Kisah Teladan Wanita-wanita mulia	1	0	0	0	0	0	0	0	0	0	0	0	1
7	Sejarah Kebudayaan Islam	1	0	0	0	0	0	0	0	0	0	0	0	1
8	Winna Efendi	2	0	0	0	0	1	0	0	0	0	0	0	3
9	Bahasa Inggris	1	0	0	0	0	0	0	0	0	0	0	0	1
10	Putri Salju	2	0	1	0	0	0	0	0	0	0	0	0	3
11	Danau Toba	2	0	1	0	0	0	0	0	0	0	0	0	3
12	27 Sedunia	2	0	0	0	0	0	0	0	0	0	0	0	2
13	Malin Kundang	3	1	0	0	0	0	0	0	0	0	0	0	4
14	Ada Apa dengan Vita	1	0	0	0	0	0	0	0	0	0	0	0	1
15	Abu Nawas	2	1	0	0	0	0	0	0	0	0	0	0	3
16	Perjalanan Gulliver	2	0	0	0	0	0	0	0	0	0	0	0	2
17	Semoga Bunda Disayang Allah	1	0	0	0	0	0	0	0	0	0	1	0	2
18	30 Dongeng Sebelum Tidur	1	1	0	0	0	0	0	0	0	0	0	0	2
19	Ramadhan Rasulullah Tentang Kiamat	1	0	0	0	0	0	0	0	0	0	0	0	1
20	Abu Dzar Al Ghifari	1	0	0	0	0	0	0	0	0	0	0	0	1
21	Bakti Seorang Pelajar	1	0	0	0	0	0	0	0	0	1	0	0	2
22	Saad Bin Abi	1	1	0	0	0	0	0	0	0	0	0	0	2
23	Pendidikan Jasmani	2	1	1	7	0	0	0	0	0	1	0	0	12
24	Mirror	1	2	1	0	0	0	0	0	0	0	0	0	4
25	Harta Halal	1	0	0	0	0	0	0	0	0	0	0	0	1
26	Sains Dan Teknologi	1	0	0	0	0	0	0	0	0	0	0	0	1
27	Ilmu pengetahuan Sosial	1	1	0	19	0	0	9	4	0	0	0	0	34
28	Akidah Akhlak	1	0	0	1	0	0	0	0	0	2	2	0	6
29	Fikih	7	0	0	0	0	0	0	0	0	0	0	0	7
30	Matematika	2	2	0	6	1	0	1	0	0	0	1	0	13
31	Qur'an Hadist	1	32	20	1	0	0	1	0	0	0	2	0	57
32	Bahasa Indonesia	1	0	1	2	0	0	0	0	0	0	0	0	4
33	Biologi	2	1	0	0	0	0	0	0	0	0	0	0	3
34	Bahasa Inggris	1	3	1	0	0	0	0	0	0	0	0	0	5
35	Pendidikan Kewarganegaraan	0	6	1	13	0	0	0	0	1	1	2	0	24
...
...
73	Kisah Kehidupan Nabi Muhammad Saw	0	0	0	0	0	0	0	0	0	0	1	0	1
74	Putri Cinderella	0	0	0	0	0	0	0	0	0	0	1	0	1
75	Istri-istri calon penghuni surga	0	0	0	0	0	0	0	0	0	0	1	0	1
Total		64	69	45	66	2	2	19	15	18	19	26	0	345

Proses *cluster* dengan mengambil jarak terdekat dari setiap data yang diolah. Dari data pinjaman buku tahun 2018 di perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun. *Cluster* buku yang paling banyak di minati (C1) yakni 5 jenis buku, *cluster* buku yang sedikit diminati (C2) yakni 70 jenis buku.

Penentuan pusat awal *cluster* (*centroid*) ditentukan secara manual atau *random* yang diambil dari data yang ada dalam *range*. Nilai *cluster* 1 diambil dari data rata-rata tertinggi dan nilai *cluster* 2 diambil dari data yang terendah.

Tabel 2. Perhitungan Jarak Pusat *Cluster*

<i>Centroid</i>	01	02	03	04	05	06	07	08	09	10	11	12
Diambil data ke-31 sebagai pusat <i>cluster</i> ke-1	1	32	20	1	0	0	1	0	0	0	2	0
Diambil data ke-3 sebagai pusat <i>cluster</i> ke-2	1	0	0	0	0	0	0	0	0	0	0	0

Maka penulis menampilkan 10 contoh perhitungan untuk jarak dari *centroid* ke-1 adalah sebagai berikut :

$$D_{Semestasebelumalam,C1} = \sqrt{(0,33333 - 4,75)^2} = 4,41667$$

$$D_{Ilmupengetahuanalam,C1} = \sqrt{(2,58333 - 4,75)^2} = 2,16667$$

$$D_{Buahkesabarandalamiman,C1} = \sqrt{(0,083333 - 4,75)^2} = 4,66667$$

$$D_{Al-quran,C1} = \sqrt{(2,583333 - 4,75)^2} = 2,16667$$

$$D_{Mc\&pidato,C1} = \sqrt{(0,083333 - 4,75)^2} = 4,66667$$

$$D_{KisahTeladanwanitamulia,C1} = \sqrt{(0,083333 - 4,75)^2} = 4,66667$$

$$D_{Sejarahkebudayaanislam,C1} = \sqrt{(0,083333 - 4,75)^2} = 4,66667$$

$$D_{Winnaefendi,C1} = \sqrt{(0,25 - 4,75)^2} = 4,5$$

$$D_{Bahasainggris,C1} = \sqrt{(0,083333 - 4,75)^2} = 4,66667$$

$$D_{Putrisalju,C1} = \sqrt{(0,25 - 4,75)^2} = 4,5$$

Dan seterusnya sampai dengan $D_{Istricalon penghunisurga,C1}$.

Hasil akhirnya jika *centroid* baru berbeda dengan *centroid* sebelumnya, maka proses dilanjutkan ke langkah berikutnya. Namun jika *centroid* yang baru dihitung sama dengan *centroid* sebelumnya, maka proses *clustering* selesai. Selanjutnya perhitungan untuk jarak dari *centroid* ke-2 menggunakan cara yang sama. Sehingga didapat tabel jarak dari *centroid* dan mencari nilai minimal dari kedua *centroid*. Dalam menentukan *cluster* dengan mencari nilai *cluster* berdasarkan nilai minimal dari nilai *cluster* dan diletakkan pada *cluster* yang sesuai dengan nilai minimal. Berikut adalah tabel *cluster*.

Tabel 3. Jarak *Centroid*

Nama Buku	C1	C2	Jarak <i>Centroid</i>	C1	C2
Semesta Sebelum Dunia	4.416667	0.25	C2		1
Ilmu Pengetahuan Alam	2.166667	2.5	C1	1	
Buah Kesabaran Dalam Iman	4.666667	0	C2		1
Al-Qur'an	2.166667	2.5	C1	1	
Mc & Pidato	4.666667	0	C2		1
Kisah Teladan Wanita-wanita mulia	4.666667	0	C2		1
Sejarah Kebudayaan Islam	4.666667	0	C2		1
Winna Efendi	4.5	0.166667	C2		1
Bahasa Inggris	4.666667	0	C2		1
Putri Salju	4.5	0.166667	C2		1
Danau Toba	4.5	0.166667	C2		1
27 Sedunia	4.583333	0.083333	C2		1
Malin Kundang	4.416667	0.25	C2		1
Ada Apa dengan Vita	4.666667	0	C2		1
Abu Nawas	4.5	0.166667	C2		1
Perjalanan Gulliver	4.583333	0.083333	C2		1
Semoga Bunda Disayang Allah	4.583333	0.083333	C2		1
30 Dongeng Sebelum Tidur	4.583333	0.083333	C2		1
Ramadhan Rasulullah Tentang Kiamat	4.666667	0	C2		1
Abu Dzar Al Ghifari	4.666667	0	C2		1
Bakti Seorang Pelajar	4.583333	0.083333	C2		1
Saad Bin Abi	4.583333	0.083333	C2		1
Pendidikan Jasmani	3.75	0.916667	C2		1
Mirror	4.416667	0.25	C2		1
Harta Halal	4.666667	0	C2		1
Sains Dan Teknologi	4.666667	0	C2		1
Ilmu pengetahuan Sosial	1.916667	2.75	C1	1	
Akidah Akhlak	4.25	0.416667	C2		1
Fikih	4.166667	0.5	C2		1
Matematika	3.666667	1	C2		1
Qur'an Hadist	0	4.666667	C1	1	
Bahasa Indonesia	4.416667	0.25	C2		1
Biologi	4.5	0.166667	C2		1
Bahasa Inggris	4.333333	0.333333	C2		1
Pendidikan Kewarganegaraan	2.75	1.916667	C1	1	
...
...
Kisah Kehidupan Nabi Muhammad Saw	4.666667	0	C2		1
Putri Cinderella	4.666667	0	C2		1
Istri-istri calon penghuni surga	4.666667	0	C2		1

Untuk mencari nilai *centroid* selanjutnya dengan menggunakan *centroid* baru dengan menjumlahkan nilai sesuai yang tertera pada *cluster* tabel diatas. Adapun *centroid* baru untuk mencari *cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *cluster* tersebut kemudian membagikannya sebanyak jumlah nilai sebagai berikut :

$$Rata_{Semestasebelumndunia} = \frac{2+1+0+1+0+0+0+0+0+0+0+0}{12} = 0,033333$$

$$Rata_{Ilmupengetahuanalam} = \frac{6+6+1+7+0+0+0+2+0+2+7+0}{12} = 2,583333$$

$$Rata_{Buahkesabarandalamiman} = \frac{1+0+0+0+0+0+0+0+0+0+0+0}{12} = 0,083333$$

$$Rata_{Al-quran} = \frac{8+1+0+5+0+0+8+4+5+0+0+0}{12} = 2,583333$$

$$Rata_{Mc&pidato} = \frac{1+0+0+0+0+0+0+0+0+0+0+0}{12} = 0,083333$$

$$Rata_{KisahTeladanwanitamulia} = \frac{1+0+0+0+0+0+0+0+0+0+0+0}{12} = 0,083333$$

$$Rata_{Sejarahkebudayaanislam} = \frac{1+0+0+0+0+0+0+0+0+0+0+0}{12} = 0,083333$$

$$Rata_{Winmaefendi} = \frac{2+0+0+0+0+1+0+0+0+0+0+0}{12} = 0,25$$

$$Rata_{Bahasainggris} = \frac{1+0+0+0+0+0+0+0+0+0+0+0}{12} = 0,083333$$

$$Rata_{Putrisalju} = \frac{2+0+1+0+0+0+0+0+0+0+0+0}{12} = 0,25$$

Dan seterusnya sampai $Rata_{Istricalonpenghunisurga}$.

Maka perhitungan dihentikan dan hasil yang diperoleh yaitu :

- Cluster* 1 (C1) memiliki 5 data yang diartikan bahwa kelompok pertama adalah kategori pinjaman buku terbanyak pada tahun 2018.
- Cluster* 2 (C2) memiliki 70 data yang diartikan bahwa kelompok kedua adalah kategori pinjaman buku sedikit pada tahun 2018.

Dari hasil pembahasan pada penelitian ini dilakukan proses perhitungan *cluster* dengan bantuan aplikasi *Rapidminer*, diperoleh hasil data pinjaman buku yang termasuk dalam *cluster* tertinggi (C1) sebanyak 5 buku dan *cluster* rendah (C2) sebanyak 70 buku. Hasil yang didapat dari perhitungan manual dengan di aplikasi *Rapidminer* sesuai atau sama (balance). Sedangkan hasil yang diperoleh dari perhitungan pada Microsoft Excel tidak ada perbedaan.

4. KESIMPULAN

Berdasarkan penelitian dalam menentukan jumlah buku diperpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun dengan Algoritma K-means dapat disimpulkan penerapan Datamining dengan menggunakan algoritma K-means dalam pengelompokan nama buku pada cluster 1 yaitu cluster tertinggi sebanyak 5 buku, dan cluster 2 yaitu cluster terendah sebanyak 70 buku. Pengujian data pada *Rapidminer* 5.3 menggunakan K-means dapat menampilkan dua kelas dari hasil klasifikasi. Dengan memperoleh hasil yang sama antara perhitungan dengan microsoft excel dan pengujian menggunakan software *Rapidminer*, maka dapat disimpulkan bahwa penelitian ini berhasil.

REFERENCES

- [1] W. S. Azis And D. Atmajaya, "Pengelompokan Minat Baca Mahasiswa Menggunakan Metode K-Means," *Ilk. J. Ilm.*, Vol. 8, No. 2, Pp. 89–94, 2016, Doi: 10.33096/Ilkom.V8i2.51.89-94.
- [2] N. I. Royanti And B. Ismanto, "Pengelompokan Keaktifan Peminjaman Buku Di Perpustakaan Stmik Widya Pratama Dengan Metode K-Means," *Ic-Tech*, Vol. 15, No. 2, Pp. 1–23, 2016.
- [3] A. E. Wicaksono, "Implementasi Data Mining Dalam Pengelompokan Peserta Didik Di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means," *J. Teknol. Rekayasa*, Vol. 21, No. 3, Pp. 206–216, 2016, [Online]. Available: [Http://www.sman16bekasi.sch.id](http://www.sman16bekasi.sch.id).
- [4] P. Soleh, A. Tholib, And M. N. F. Hidayat, "Penerapan Data Mining Untuk Analisa Pola Pembelian Produk Menggunakan Algoritma Frequent Pattern – Growth," *Rekayasa*, Vol. 14, No. 3, Pp. 456–460, 2022, Doi: 10.21107/Rekayasa.V14i3.11365.
- [5] W. M. P. Dhuhita, "Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita," *J. Inform. Darmajaya*, Vol. 15, No. 2, Pp. 160–174, 2015.
- [6] K. Fatmawati And A. P. Windarto, "Data Mining : Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," Vol. 3, No. 2, Pp. 173–178, 2018.)