

## Penerapan Algoritme K-Means untuk Pengelompokan Buku Berdasarkan Tingkat Minat Pembaca

Andri Juliyanto  
Universitas Buana Perjuangan  
Karawang, Indonesia  
if15.andrijuliyanto@mhs.ubpkarawang.ac.id

Amril Mutoi Siregar  
Universitas Buana Perjuangan  
Karawang, Indonesia  
amrilmutoi@ubpkarawang.ac.id

Dwi Sulistya Kusumaningrum  
Universitas Buana Perjuangan  
Karawang, Indonesia  
dwi.sulistya@ubpkarawang.ac.id

### **Abstract—**

Perpustakaan menyediakan layanan baca dan peminjaman buku untuk menunjang sarana proses belajar mengajar. Dinas Perpustakaan Kabupaten Karawang mencatat setiap transaksi peminjaman buku yang dipinjam oleh anggota perpustakaan. Selama ini anggota perpustakaan kesulitan menemukan rak-rak buku yang sesuai dengan tingkat minat pembaca. Maka dalam mengatasi permasalahan penelitian ini dilakukan dengan menerapkan algoritme K-Means. Penerapan algoritme K-Means dalam penelitian ini dilakukan untuk menentukan pengelompokan buku, sehingga hasil yang didapat mampu membantu petugas perpustakaan menyusun rak-rak buku. Berdasarkan hasil perhitungan dalam penelitian ini yang menggunakan data latih jenis buku sebanyak 38 jenis buku dan 13 atribut data menghasilkan tiga kluster. Kluster 1 adalah kluster paling diminati yang menghasilkan 31 buku, kluster 2 yaitu kluster cukup diminati menghasilkan tiga buku dan kluster 3 merupakan kluster kurang diminati menghasilkan empat buku.

**Kata kunci** — K-Means, Pengelompokan, Perpustakaan

### **I. PENDAHULUAN**

Perpustakaan adalah tempat atau gedung yang menyediakan koleksi buku dan bahan kepustakaan yang dapat meningkatkan suatu mutu pendidikan [1,2]. Selama ini anggota perpustakaan pada Dinas Perpustakaan Kabupaten Karawang mengalami kesulitan menemukan buku yang dicari. Salah satu cara untuk mengatasi permasalahan pada Dinas Perpustakaan Kabupaten Karawang adalah dengan cara pengelompokan menggunakan algoritme K-Means. Algoritme K-means terbukti mampu menyelesaikan beberapa masalah pada penelitian sebelumnya [3,4,5,6].

Penelitian sebelumnya melakukan penelitian tentang pengelompokan beasiswa bidik misi dengan menggunakan algoritme K-Means oleh Jaroji, dkk [3]. Penelitian tersebut dilakukan dengan cara mengolah data pelamar beasiswa berdasarkan delapan atribut yaitu nisn, nama, penghasilan, status rumah, kondisi rumah, tanggungan, status orang tua dan akademik. Hasil penelitian tersebut diperuntukan sebagai masukan bagi Ristekdikti untuk memberikan beasiswa bagi mahasiswa yang kondisi dari keluarga yang tidak mampu. Hasil penelitian ini terdapat tiga kluster yaitu kluster sangat layak dengan 24 orang yang direkomendasikan, kluster layak dengan 32 orang yang direkomendasikan dan kluster kurang layak sebanyak 56 orang. Lalu pada penelitian berikutnya tentang menentukan pengelompokan nilai gizi balita oleh Irfiani dan Rani [4] mampu menghasilkan lima kluster dengan metode K-Means. Cara untuk menentukan nilai gizi pada balita adalah dengan mengukur berat badan dan tinggi badan berdasarkan umur. Hasil kluster yang didapatkan terdiri dari kluster balita obesitas sebanyak 30%, kluster balita gizi lebih sebanyak 24%, kluster gizi baik sebanyak 35%, kluster gizi kurang sebanyak 4% dan kluster gizi buruk sebanyak 7%. Kemudian penelitian tentang penjualan produk dengan menerapkan metode K-Means oleh Metisen dan Sari [5] menghasilkan dua kluster. Penelitian tersebut dilakukan dengan cara mengelompokkan produk laris dan tidak laris terjual. Kluster yang didapatkan adalah kluster penjualan kurang menggolongkan empat produk dan kluster penjualan tinggi menggolongkan enam produk. Di samping itu, Ghofar dan Kurniawan [6] juga melakukan penelitian tentang pengelompokan pelanggan dengan metode K-Means. Penelitian tersebut dilakukan dengan cara mengolah data transaksi, penukaran voucher dan pembelian tunai dengan hasil pengelompokan sebanyak tiga kluster. Kluster dalam penelitian tersebut menghasilkan kluster satu berjumlah 310 data, kluster dua berjumlah 37 data dan kluster tiga berjumlah sepuluh data.

Berdasarkan masalah yang telah diselesaikan pada penelitian sebelumnya [3,4,5,6], bahwa dengan menggunakan algoritme K-Means berhasil membuktikan pengelompokan dari permasalahan yang ada. Maka dalam penelitian ini menerapkan algoritme K-Means dalam menentukan pengelompokan buku berdasarkan tingkat minat pembaca untuk mengetahui kluster paling diminati, cukup diminati dan kurang diminati. Hal ini dilakukan untuk membantu petugas Dinas Perpustakaan Kabupaten Karawang menyusun rak-rak buku berdasarkan tingkat minat pembaca.

Selebihnya penelitian ini akan dijelaskan pada bagian II tentang Algoritme K-Means, bagian III tentang Implementasi algoritme K-Means, dan bagian IV tentang hasil dan pembahasan serta bagian V tentang kesimpulan.

## II. ALGORITME K-MEANS

Algoritme K-Means ditemukan oleh Lloyd pada tahun 1957 dan diperkenalkan pada tahun 1982 [7]. K-Means merupakan salah algoritme yang membagi data ke dalam bentuk satu pengelompokan atau lebih [5]. Keunggulan algoritma K-Means adalah waktu komputasi yang cepat dan proses yang sederhana [8]. Tahapan untuk melakukan algoritme K-Means langkah pertama menentukan jumlah kluster yang akan dibentuk. Kemudian menentukan nilai pusat *centroid* awal kluster secara acak. Lalu hitung jarak dari setiap data ke masing-masing *centroid*. Setelah menentukan nilai pusat *centroid* awal, kemudian melakukan perhitungan jarak pada setiap data ke masing-masing *centroid* ke pusat kluster awal. Jika pusat kluster baru sama dengan pusat kluster lama maka proses berhenti.

Selanjutnya cara untuk menentukan jarak pada suatu data ke kluster dilakukan perhitungan jarak menggunakan teori *Euclidean* yang ditunjukkan pada persamaan 1.

$$R(i, j) = \sqrt{(x_{1j} - x_{1k})^2 + (x_{2j} - x_{2k})^2 + \dots + (x_{nj} - x_{nk})^2} \quad (1)$$

Dengan ini persamaan 1 diuraikan sebagai berikut :

$R(j, k)$  = Jarak data ke  $j$  ke pusat cluster  $k$

$X_{nj}$  = Data ke  $j$  pada atribut data ke  $n$

Apabila setelah proses perhitungan persamaan 1 dilakukan kemudian selanjutnya menentukan titik *centroid* baru dengan cara menghitung nilai rata-rata dengan menggunakan persamaan 2.

$$q(i) = \frac{p_1 + p_2 + \dots + p_n}{\sum n} \quad (2)$$

Keterangan :

$q(i)$  = Centroid data

$p_1$  = Data ke  $i$  pada atribut data ke  $k$

$\sum n$  = Jumlah Data

## III. IMPLEMENTASI ALGORITMA K-MEANS TERHADAP PENGELOMPOKAN BUKU

### 1. Data Penelitian

Data yang digunakan dalam penelitian ini didapatkan dari Dinas Perpustakaan Kabupaten Karawang. Data tersebut adalah data transaksi peminjaman buku pada tahun 2017 sampai dengan tahun 2018. Data yang terdiri dari 14,855 transaksi peminjaman buku, 38 jenis buku dan 13 atribut. 13 atribut yaitu Jenis buku, Januari, Februari, Maret, April, Mei, Juni, Juli, Agustus, September, Oktober, November, dan Desember.

### 2. Penerapan Algoritme K-Means

Tahap penerapan algoritme K-Means untuk melakukan proses pengolahan data dilakukan dengan tahapan – tahapan algoritme K-means, Kemudian menentukan  $k$  kluster sebanyak tiga kluster secara acak. Hasil penerapan algoritme K-means dalam tahap ini akan menghasilkan  $k$  kluster paling diminati, kluster cukup diminati dan kluster kurang diminati. Selanjutnya melakukan tahap perhitungan dengan persamaan 1, agar menghasilkan nilai jarak pada setiap data buku ke *centroid*. Sebagai contoh data yang diambil untuk perhitungan *centroid* 1 adalah “Buku Sosiologi” pada jenis buku, lalu sebagai cara untuk menghitung jarak pada setiap data ke *centroid* menggunakan persamaan 1.

$$R(1,1) = \sqrt{\frac{((29 - 25)^2 + (17 - 40)^2 + (27 - 29)^2 + (13 - 22)^2 + (28 - 29)^2 + (15 - 21)^2 + (25 - 31)^2 + (30 - 18)^2 + (24 - 19)^2 + (32 - 16)^2 + (21 - 18)^2 + (20 - 32)^2)}{12}} = 35.79$$

selanjutnya hitung jarak pada setiap data “Buku Sosiologi” ke *centroid* 2 dengan menggunakan persamaan 1.

$$R(1,2) = \sqrt{\frac{((29-29)^2 + (17-46)^2 + (27-109)^2 + (13-75)^2 + (28-27)^2 + (15-32)^2 + (25-38)^2 + (30-10)^2 + (24-211)^2 + (32-34)^2 + (21-23)^2 + (20-22)^2)}{217.37}} =$$

Agar mengetahui *centroid* 3 pada data “Buku Sosiologi” maka selanjutnya melakukan perhitungan dengan menggunakan persamaan 1.

$$R(1,3) = \sqrt{\frac{((29-57)^2 + (17-54)^2 + (27-56)^2 + (13-68)^2 + (28-86)^2 + (15-54)^2 + (25-27)^2 + (30-65)^2 + (24-24)^2 + (32-87)^2 + (21-76)^2 + (20-46)^2)}{217.33}} = 137.33$$

Setelah selesai melakukan perhitungan jarak dengan menggunakan persamaan 1, maka selanjutnya menentukan pusat *centroid* baru 1 ( $q_1$ ) dari kluster dengan menggunakan persamaan 2, contoh untuk menghitung *centroid* baru digunakan pada “Bulan Januari”.

$$q_1 = \frac{25+28+23+28+17+19+14+30+27+29+45+23+54+13+31+21+25+23+29+10+23+18+19+16+15+24+15+30+35+18+36}{31} = 24.613$$

Setelah melakukan perhitungan *centroid* 1 ( $q_1$ ) kemudian melakukan perhitungan *centroid* baru 2 dengan persamaan 2, contoh yang digunakan *centroid* adalah pada “Bulan Januari”.

$$q_2 = \frac{29+40+22}{3} = 30.333$$

Setelah melakukan perhitungan *centroid* 2 baru dilakukan perhitungan *centroid* baru 3 dengan persamaan 2, contoh yang digunakan *centroid* adalah pada “Bulan Januari”.

$$q_3 = \frac{57+34+80+78}{4} = 62.250$$

Berdasarkan persamaan 1 dan persamaan 2, maka cara sederhana algoritma K-Means ditunjukkan pada Algoritme 1 [9].

#### Algoritme 1

##### Input 1:

$R = \{X_1 + X_2 + \dots + X_n\}$  //kumpulan data

$K$  // menentukan jumlah kluster

##### Output:

$K$  // kumpulan kluster

##### Algoritme K-Means:

Menentukan pusat kluster awal  $q_1, q_2, \dots, q_n$

##### repeat

Menentukan data ke kluster yang memiliki rata-rata;

Menghitung rata-rata pada setiap kluster untuk menentukan *centroid* baru;

**until** pengelompokan berhasil

### 3. Evaluasi

Setelah mengetahui hasil kluster buku berdasarkan tingkat minat pembaca dengan menghitung secara manual dengan menggunakan persamaan 1 dan persamaan 2 dengan algoritme K-Means. Kemudian dalam tahap evaluasi dengan menggunakan teknik komputasi.

## IV. HASIL DAN PEMBAHASAN

Hasil pengelompokan dengan perhitungan manual dapat menghasilkan perhitungan pengelompokan mengalami tiga kali pengulangan. Hasil perhitungan ditunjukkan pada Tabel 1. Pada jenis buku Jurnalis termasuk kluster satu sebab antara  $q_1, q_2$  dan  $q_3$  nilai terkecil terdapat pada  $q_1$ . Kemudian buku Psikologi pada kluster 2 karena perbandingan antara  $q_1, q_2$  dan  $q_3$  nilai terkecil terdapat pada  $q_2$ . Buku agama termasuk dalam kluster 3 karena perbandingan antara  $q_1, q_2$  dan  $q_3$  nilai terkecil terdapat pada  $q_3$ .

Tabel 1 Hasil Perhitungan Manual Pada Data Transaksi Peminjaman Buku.

Jenis Buku	$q_1$	$q_2$	$q_3$
Jurnalis	22.78	172.83	180.69
Majalah	30.449	177.97	186.63

Jenis Buku	q1	q2	q3
Media massa	71.677	147.33	132.05
Kutipan	82.645	149.7	141.02
Metafisika	32.351	184.59	207.23
Epistimologi	26.225	173.42	192.58
Psikologi	204.675	74.63	208.31
Filosofis Logis	41.913	192.09	214.62
Etik	49.413	171.65	173.97
Pemikiran Filosofis	33.427	187.62	206.07
Agama	135.57	182.61	89.86
-			
-			
-			
Biologi	32.262	173.87	189.65

Setelah mengetahui hasil pengelompokan pada Tabel 1, kemudian mengurutkan nilai *centroid* terkecil untuk menentukan kluster paling, kluster cukup dan kluster kurang. Sehingga mendapatkan hasil yang ditunjukkan pada Tabel 2.

Tabel 2 Hasil Perhitungan Manual dan Evaluasi Secara Komputasi

Kluster	Jenis Buku	Jumlah
Paling	Jurnalis, Majalah, Media massa, Kutipan, Metafisika, Epistimologi, Filosofis Logis, Etik, Pemikiran Filosofis, Sosiologi, Statistik, Ilmu politik, Ekonomi, Hukum, Pendidikan, Norma, Farmasi, Pertanian, Manufaktur, Konstruksi, Seni dan Rekreasi, Arsitektur, Olahraga, Kesenian, Musik, Literatur, Retorika, Sejarah, Geografi, Biografi, Biologi	31 Jenis buku
Cukup	Psikologi, Teknik, Teknik Kimia	3 Jenis Buku
Kurang	Agama, Bahasa, Teknologi, Sastra	4 Jenis Buku

Hasil pengelompokan buku berdasarkan tingkat minat pembaca menggunakan algoritme K-Means dengan melakukan perhitungan persamaan 1 dan persamaan 2 menghasilkan tiga kluster. Kluster paling diminati menghasilkan 31 jenis buku. Tiga buku tertinggi pada kluster paling diminati yaitu buku sosiologi, statistik, ilmu politik. Kluster cukup diminati menghasilkan tiga jenis buku yaitu buku psikologi, teknik dan teknik kimia. Kluster kurang diminati menghasilkan empat jenis buku dengan tiga buku tertinggi yaitu Agama, Bahasa, Teknologi.

## V. KESIMPULAN DAN SARANA

Pada hasil penelitian yang ditunjukkan pada bagian IV, maka dapat disimpulkan bahwa algoritme K-Means mampu memberikan hasil pengelompokan buku berdasarkan tingkat minat pembaca. Kluster yang diperoleh sebanyak tiga kluster, kluster terdiri dari kluster buku yang memiliki tingkat minat pembaca yang paling diminati sebanyak 31 buku dengan nilai tiga buku tertinggi yaitu buku sosiologi, statistik, ilmu politik. Lalu kluster cukup diminati sebanyak tiga buku yaitu buku psikologi, teknik dan teknik kimia. Kemudian kluster kurang diminati sebanyak empat buku dengan nilai tiga buku tertinggi yaitu buku Agama, Bahasa, Teknologi. Penelitian ini berhasil untuk menentukan pengelompokan buku berdasarkan tingkat minat pembaca. Sehingga saran untuk penelitian selanjutnya, sebaiknya dikembangkan aplikasi perpustakaan untuk merekomendasikan buku kepada anggota yang bertujuan untuk menyarankan buku yang sesuai dengan buku yang telah di baca sebelumnya.

## PENGAKUAN

Naskah ilmiah ini adalah sebagian penelitian Tugas Akhir milik Andri Juliyanto dengan judul Pengelompokan buku berdasarkan tingkat minat pembaca dengan algoritme K-Means (Studi kasus Dinas perpustakaan Kabupaten Karawang) yang dibimbing oleh Amril Mutoi Siregar dan Dwi Sulistya Kusumaningrum.

## DAFTAR PUSTAKA

- [1] Deputi II Perpustnas RI, Peningkatan Layanan Perpustakaan Sekolah, Seminar Ilmiah Nasional, 2010.

- [2] Badan Pengembangan dan Pembinaan Bahasa, Kemendikbud RI. 2016: KBBI daring, <http://kbbi.kemdikbud.go.id/entri/perpustakaan>, diakses 25 September 2019.
- [3] Jaroji, J., Danuri, D., & Putra, F. P., K-Means Untuk Menentukan Calon Penerima Beasiswa Bidik Misi Di Polbeng. *INOVTEK Polbeng - Seri Informatika*, 1(1), 87. <http://doi.org/10.35314/isi.v1i1.129>, 2016.
- [4] Irfiani, E., & Rani, S. S., Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita, 6(4), 161–168, 2018.
- [5] Metisen, B. M., & Sari, H. L., Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila, 11(2), 110–118, 2015.
- [6] Ghofar, M. A., & Kurniawan, Y. I., Aplikasi Pengelompokan Pelanggan Pada UMS Store Menggunakan Algoritma K-Means, *Program Studi Informatika, Fakultas Komunikasi dan Informatika*, 4(1), 2018.
- [7] Morissette, L., & Chartier, S., The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, 9(1), 15–24. <https://doi.org/10.20982/tqmp.09.1.p015>, 2018.
- [8] A. Yasid, "Implementasi Automatic Clustering Menggunakan Differential Evolution dan CS Measure untuk Analisis Data Kemahasiswaan," *Jurnal Ilmiah NERO*, vol. 1, p. 47, 2014.
- [9] Dunham, M.: *Data mining: Introductory and Advance Topics*. N.J. Prentice Hall, 2003.