

PENERAPAN DATA MINING CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKAN BUKU PERPUSTAKAAN POLITEKNIK NEGERI BALIKPAPAN

Maria Ulfah¹⁾, Andi Sri Irtawaty²⁾

^{1,2} Jurusan Teknik Elektro, Politeknik Negeri Balikpapan
E-mail: maria.ulfah@poltekba.ac.id

Abstract

Politeknik Negeri Balikpapan's Library, annually adds to its collection of reading books. In order to procure the book according to the needs of users, information on the collection of books needed or in demand is needed. To answer this problem, it is necessary to have a clustering system for existing books in the library by looking at the aspect of borrowing frequency. The Clustering system is made using the K-Means method with the selection of 3 Clusters, namely, very attractive, desirable and less desirable. From the results of data processing through the Rapid Miner Application with K=3, the results obtained are cluster_0 (low) consisting of 82 book titles with the frequency of borrowing books in the rare category in other words less desirable to borrow, cluster_1 (medium) consists of 23 book titles which are books with moderate borrowing frequency, in other words, are categorized as desirable for borrowing, cluster_2 (high) consisting of 2 book titles, is the most popular group of books consisting of 2 titles, namely Teknologi Beton and buku Teori dan Praktik hotel Front Office. The results of grouping the data of these books can be used as input for library managers in procuring book collections based on the frequency of borrowing books.

Keywords: *Clustering, K-Means, Rapid Miner, Library, Bouldin*

PENDAHULUAN

Perpustakaan merupakan tempat dimana terdapat beranekaragaman koleksi pustaka yang dapat memperkaya pengetahuan pengunjungnya (Indhitya, 2022). Perpustakaan Politeknik Negeri Balikpapan bertujuan mendukung kebutuhan informasi, kondisi minat baca perpustakaan dapat dilihat dari persentase jumlah pengunjung, peminjam dan buku yang dipinjam. Data peminjaman buku dijadikan sebagai laporan. Pihak Perpustakaan setiap tahunnya melakukan pengadaan koleksi buku baru untuk diperbanyak sehingga harus mengetahui jenis koleksi buku prioritas untuk diperbanyak. Sehingga diperlukan adanya analisis pengolahan data transaksi peminjaman buku yang menjelaskan informasi sebaran intensitas peminjaman buku, informasi pengolahan tersebut membantu penambahan koleksi.

Untuk mengetahui Buku apa yang paling diminati, dilakukan teknik clustering menggunakan metode K-Means. Algoritma K-Means merupakan algoritma

pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah K Cluster yang sudah ditetapkan di awal (Iin, 2018). Oleh karena itu penulis dalam penelitian ini mengambil judul “Penerapan Data Mining Clustering Menggunakan Metode K-Means Dalam Pengelompokan Buku Perpustakaan Politeknik Negeri Balikpapan” untuk membantu mengelompokkan data buku pinjaman yang diminati dalam mendukung Pengelola Perpustakaan Politeknik Negeri Balikpapan dalam pengadaan koleksi buku.

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakut dari berbagai database besar (Kusrini dkk, 2009). Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual (Pramudiono, 2006). Data mining sering juga disebut Knowledge Discovery in Database (KDD), yaitu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Knowledge Discovery in Database (KDD) suatu teknik pembentukan pola atau rule dalam informasi. Informasi yang dihasilkan didapatkan dari suatu data yang besar atau dikenal dengan tambang data yang disimpan dalam basis data yang awalnya belum diketahui dan menghasilkan suatu data yang potensial bermanfaat. Iterasi dalam data mining disebut proses KDD (Januardi, 2020)

. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu hierarchical clustering dan non hierarchical clustering (Budi, 2007). Metode non-hierarchical clustering dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan K-Means Clustering (Sulaiman, 2020).

K-Means adalah sebuah algoritma untuk mengklasifikasikan atau mengelompokkan objek objek (dalam hal ini data) berdasarkan paramater tertentu kedalam sejumlah group, sehingga berjalan lebih cepat dari pada hierartical clustering (jika kecil) dengan jumlah variabel yang besar dan menghasilkan cluster yang lebih rapat (Sulaiman,2020). K-Means memiliki properti: selalu ada K cluster, paling tidak memiliki satu data dalam tiap cluster, cluster ini merupakan non-hierarki dan tidak akan terjadi Overlap, dan

setiap member dari sebuah cluster berdekatan di cluster terhadap cluster lainnya karena kedekatan tidak selalu melibatkan pusat dari cluster itu.

METODE PENELITIAN

Berikut metode pelaksanaan penelitian:



Gambar 1. Metode Penelitian

1. Tahap Perencanaan

Langkah awal dalam penelitian ini adalah melakukan perencanaan penelitian. Ada empat kegiatan dalam perencanaan yaitu penentuan tujuan penelitian, identifikasi masalah, penentuan batasan masalah dan studi pustaka.

2. Tahap Preprocessing

Dalam tahap preprocessing kegiatan yang dilakukan adalah menyetik kembali data yang didapatkan kedalam microsoft excel untuk mencatat semua transaksi peminjaman buku. Setelah data di salin pada tahap ini juga akan dilakukan cleaning data, yaitu penghapusan data yang tidak jelas penulisannya atau data yang tidak bisa dibaca. Proses cleaning ini dilakukan agar mendapatkan hasil perhitungan yang benar.

3. Tahap Pengolahan Data

Mengolah data dengan software rapidminer dan juga mengolah data berdasarkan identifikasi masalah pada penelitian. Menggunakan metode K-Means, dan melalui penerapan algoritma k-means clustering diharapkan dalam mengolah data ini mendapatkan hasil yang baik untuk mengelompokkan buku berdasarkan frekuensi peminjaman. Hasil belajar dan menghasilkan suatu pengetahuan baru. Mengolah data dengan software Rapidminer

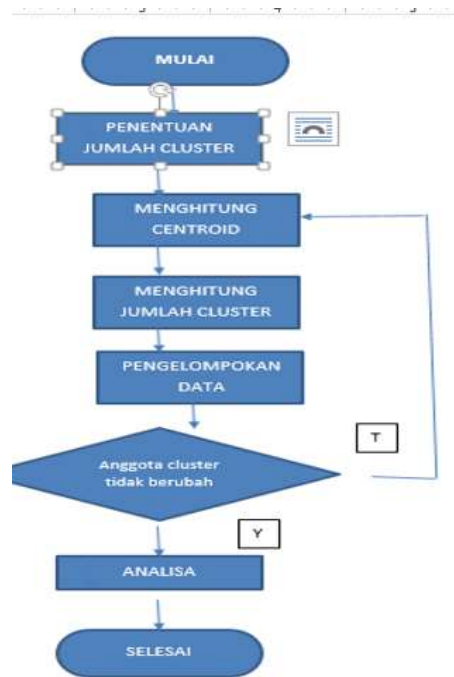
4. Tahap Analisa

Setelah semua data terkumpul selanjutnya dilakukan tahap analisa.

5. Tahap Dokumentasi

Proses dokumentasi dilakukan dari awal penelitian hingga akhir penelitian. Hasil akhir dari proses dokumentasi berupa laporan kemajuan dan Laporan Akhir

Berikut diagram alir penelitian:



Gambar 2. Diagram Alir

HASIL DAN PEMBAHASAN

A. Data Set

Dataset yang digunakan adalah data peninjaman buku pada Perpustakaan Politeknik Negeri Balikpapan dari kurun waktu 2019- Juni 2022. Banyaknya jenis buku yang terdata sejumlah 107 judul dengan total frekuensi peminjaman 231.

B. Preprocessing Data

Pada tahapan ini, menggunakan tools RapidMiner dengan metode K-Means. Tampilan New Process ialah untuk mempersiapkan lembar kerja pada RapidMiner. Mengimport data yang akan diuji berisi format .xls atau .xlsx. selanjutnya ialah model untuk mengimport file Microsoft Excel. Untuk memasukkan data yang akan dijalankan perlu melalui klik kanan -> Insert Operator -> Data Access -> File -> Read Excel

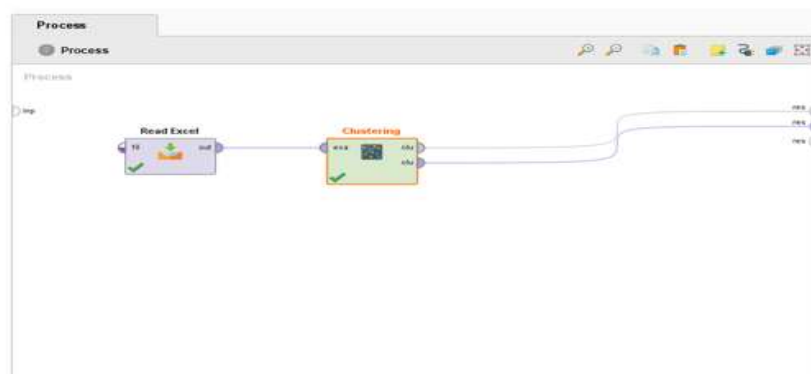
C. Modelling dan Evaluasi

Pada tahapan ini, menambahkan operator K-Means. melalui klik kanan -> Insert Operator -> Modeling -> Segmentation -> K-Means.

Berikutnya mengerjakan pengaturan pada menu Parameter Clustering K-Means, dilakukan pengaturan nilai k, dimana k menjadi nilai yang akan digunakan untuk penentuan jumlah cluster yang akan dibuat. Disini jumlah cluster yang akan dibuat yakni sebanyak 3 cluster (rendah, sedang, tinggi)

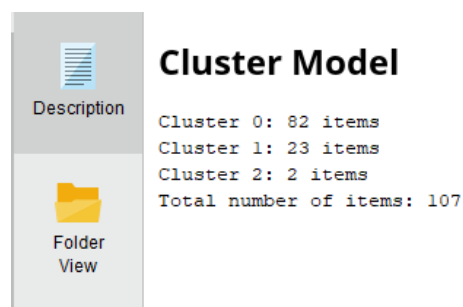
Berikutnya mengerjakan pengaturan pada menu Parameter Clustering K-Means, dilakukan pengaturan nilai k, dimana k menjadi nilai yang akan digunakan untuk penentuan jumlah cluster yang akan dibuat. Disini jumlah cluster yang akan dibuat yakni sebanyak 3 cluster (rendah, sedang, tinggi)

Bentuk pemodelan menjadi seperti berikut:



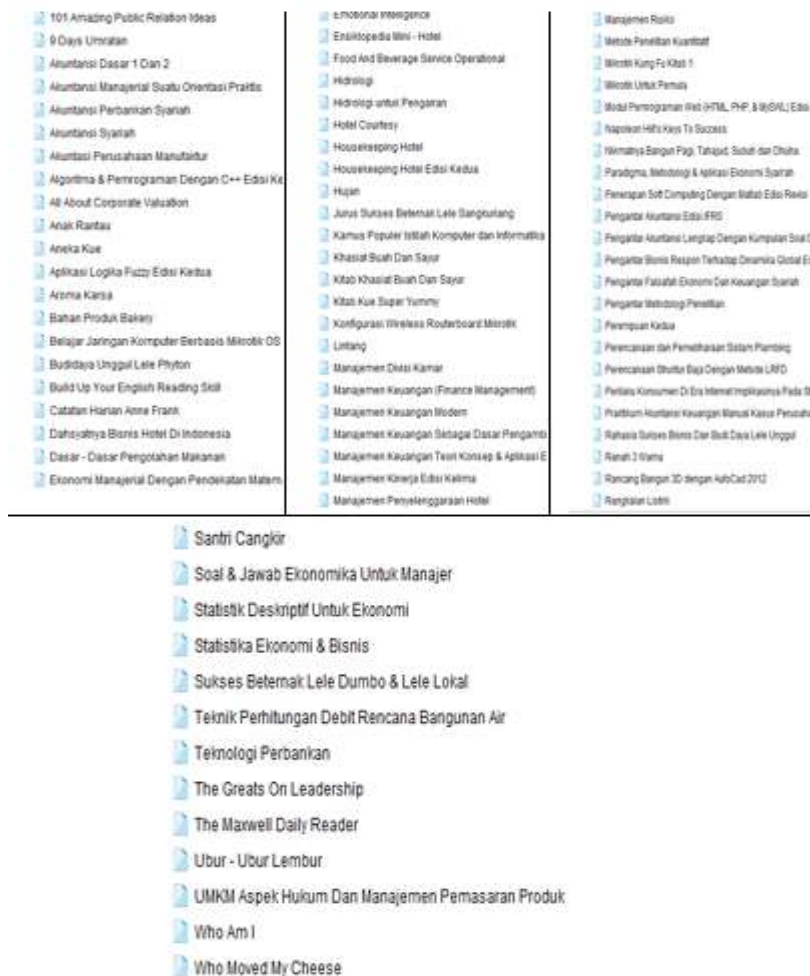
Gambar 3. Pemodelan Clustering dengan operator K-Means

Pada Gambar 4 hasil data terurai menjadi 3 cluster yaitu cluster 0, cluster 1, cluster 2 dengan setiap cluster mengantongi hasil pengelompokkan data cluster_0 terdiri 82 judul buku, cluster_1 terdiri dari 23 judul buku, cluster_2 terdiri dari 2 judul buku.



Gambar 4. Cluster Model

Berikut anggota dari Cluster 0:



Gambar 5. Anggota Cluster 0

Berikut anggota dari Cluster 1:



Gambar 6. Anggota Cluster 1

Berikut anggota dari Cluster 2:



Gambar 7. Anggota Cluster 2

Dari hasil clusterisasi dengan $K=3$, terlihat bahwa pada Cluster 2 (tinggi) merupakan kelompok buku yang paling diminati yang terdiri dari 2 judul buku yakni Teknologi Beton dan buku Teori dan Praktik hotel Front Office judul-judul buku yang paling banyak diminati untuk dipinjam selama kurun waktu 2019 hingga Juni 2022 sebanyak 14 dan 13 kali frekuensi peminjaman.

Cluster 1 (sedang) terdiri dari 23 judul buku yang merupakan buku-buku dengan frekuensi peminjaman sedang dengan kata lain dikategorikan diminati untuk dipinjam. Sedangkan Cluster 0 (rendah) terdiri dari 82 judul buku dengan frekuensi peminjaman buku nya dalam kategori jarang dengan kata lain kurang diminati untuk dipinjam.

Dari hasil perfomansi didapatkan nilai index Davies Bouldin sebesar 0,407 seperti gambar 8.



Gambar 8. Nilai Index Davies Bouldin

SIMPULAN

Dari hasil pengolahan data melalui Aplikasi Rapid Miner dengan $K=3$, didapatkan hasil cluster_0 (rendah) terdiri 82 judul buku dengan frekuensi peminjaman buku nya dalam kategori jarang dengan kata lain kurang diminati untuk

dipinjam, cluster_1 (sedang) terdiri dari 23 judul buku yang merupakan buku-buku dengan frekuensi peminjaman sedang dengan kata lain dikategorikan diminati untuk dipinjam, cluster_2 (tinggi) terdiri dari 2 judul buku merupakan kelompok buku yang paling diminati yang terdiri dari 2 judul buku yakni Teknologi Beton dan buku Teori dan Praktik hotel Front Office. Hasil performansi didapatkan nilai index Davies Bouldin sebesar 0,407. Pengelompokan data buku-buku ini dapat menjadi masukan bagi pengelola perpustakaan dalam pengadaan koleksi buku berdasarkan frekuensi peminjaman buku.

DAFTAR PUSTAKA

- Budi, S. (2007). Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.
- Iin, P. (2018). Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center. CESS (Journal of Computer Engineering System and Science), 3(1), 87–93.
- Indhitya,R. (2022). Penerapan Clustering K-Means Untuk Mendukung Pengelolaan Koleksi Pada Perpustakaan Fakultas Teknik Universitas Negeri Gorontalo. Jurnal Teknik
- Januardi, N. (2020). Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means
- Kusrini dan Lutfi, E.T. 2009. Alogaritma Data Mining. Yogyakarta: Andi Offset
- Pramudiono, (2006), Apa Itu Data Mining?, [online], diakses tanggal 15 Maret 2022
- Sulaiman. (2020). Analisis pola belanja konsumen menggunakan algoritma k-means dan apriori pada haura swalayan. UIN Sultan Syarif Kasim Riau