

Analisa dan Perbandingan Metode Klasterisasi untuk Mengelompokkan Koleksi Buku Perpustakaan

Nia Adila^{*1}, Yesi Novaria Kunang², Izman Herdiansyah³, Edi Surya Negara⁴

^{1,2,3,4}Magister Teknik Informatika, Universitas Bina Darma, Indonesia
Email: 1niaadila04@gmail.com, 2yesinovariakunang@binadarma.ac.id,
3m.herdiansyah@binadarma.ac.id, 4e.s.negara@binadarma.ac.id

Abstrak

Perpustakaan merupakan salah satu sumber informasi yang dapat dimanfaatkan untuk memenuhi suatu kebutuhan informasi ilmiah oleh mahasiswa sehingga disetiap perguruan tinggi diwajibkan memiliki perpustakaan salah satunya Universitas Bina Darma Palembang. Perpustakaan Universitas Bina Darma Palembang memiliki tingkat pengunjung dan anggota yang cukup tinggi sehingga terdapat koleksi buku yang sudah lengkap dan telah terintegrasi dengan baik secara terkomputerisasi. Dari adanya kumpulan koleksi data buku yang sering di inputkan, perpustakaan harus *intens update* data buku sehingga akan membutuhkan suatu pengelompokan buku untuk mengetahui informasi data setiap kelompok buku dan membantu mengelompokkan buku secara otomatis berdasarkan subjek yang di butuhkan perpustakaan dan akan dianalisis menjadi suatu data pengambilan sebuah keputusan. Pada penelitian ini akan berfokus pada data buku perpustakaan Universitas Bina Darma. Data tersebut disimpan berupa catalog buku dan hanya diupdate pada setiap pelaporannya pertahun. Dari banyak data catalog buku yang ada pihak perpustakaan kurang memanfaatkan data tersebut sehingga peneliti tertarik menganalisa data buku pada perpustakaan Universitas Bina Darma, adanya proses Analisa pada data akan menggunakan perbandingan dua metode pada perpustakaan Universitas Bina Darma kemudian akan menghasilkan sebuah perbandingan pengelompokan informasi data buku dan kemudian akan dibandingkan menggunakan dua metode pengelompokan klasterisasi pada Perpustakaan Universitas Bina Darma yaitu metode *K-Means* dan *K-Medoids*. Penelitian akan menghasilkan perbandingan antara setiap variable dari dataset yang sudah ada yaitu pertama proses general untuk mengeksplorasi variabel tentang buku-buku yang ada dalam data ini secara umum dan menghasilkan perbandingan *cluster*. Kemudian akan dilanjutkan pada beberapa proses untuk hasil implementasi menggunakan dua metode *K-Means* dan metode *K-Medoids*. Kemudian data di *preprocessing* dengan memanfaatkan *software microsoft excel* dan *google colab*. Maka penelitian ini dapat menjadi sebuah informasi pengambil keputusan dan perbandingan *cluster* ari data Perpustakaan Universitas Bina Darma.

Kata kunci: *Clustering, Google Colab, K-Means, K-Medoids, Perpustakaan.*

Analysis and Comparison of Clustering Methods to Group Library Book Collections

Abstract

The library is one of the sources of information that can be utilized to fulfill the need for scientific information by students so that every university is required to have a library, one of which is Bina Darma University, Palembang. The Bina Darma University Library, Palembang has a fairly high level of visitors and members so that there is a complete collection of books that has been well integrated in a computerized manner. From the collection of book data collections that are often inputted, the library must intensively update book data so that it will require a grouping of books to find out data information for each group of books and help group books automatically based on the subjects needed by the library and will be analyzed into data for decision making. This study will focus on the book data of the Bina Darma University library. The data is stored in the form of a book catalog and is only updated in each annual report. Of the many book catalog data available, the library does not utilize the data so that researchers are interested in analyzing book data at the Bina Darma University library, the analysis process on the data will use a comparison of two methods at the Bina Darma University library, then it will produce a comparison of book data information grouping and then it will be compared using two clustering methods at the Bina Darma University Library, namely the K-Means and K-Medoids methods. The study will produce a comparison between each variable from the existing dataset, namely the first general process to explore variables about the books in this data in general and produce a cluster comparison. Then it will be continued in several processes for the implementation results using two methods K-Means and the K-Medoids method. Then the data

is preprocessed using Microsoft Excel and Google Colab software. So this study can be an information decision maker and cluster comparison of Bina Darma University Library data.

Keywords: Clustering, Google Colab, K-Means, K-Medoids

1. PENDAHULUAN

Perpustakaan merupakan salah satu sumber informasi yang dapat dimanfaatkan untuk memenuhi suatu kebutuhan informasi ilmiah oleh mahasiswa sehingga disetiap perguruan tinggi diwajibkan memiliki perpustakaan salah satunya Universitas Bina Darma Palembang. Pengelompokan data buku akan di bagi menjadi beberapa klasterisasi sehingga menggunakan perbandingan dari dua metode yang berbeda untuk mendapatkan informasi yang lebih akurat mutlak dan sesuai kebutuhan. Pada penelitian ini akan menggunakan perbandingan metode *K-Means* dan *K-Medoids* karena kedua metode ini memiliki perbedaan pada proses implementasinya. *K-Means* merupakan algoritma non hirarki yang berasal dari metode dan *clustering*[1], Sedangkan metode *K-Medoids* merupakan salah satu algoritma *clustering* yang memiliki kelebihan yaitu dapat mengatasi sensitive terhadap outlier, *K-Medoids* menentukan *centroid* secara acak, menghitung jarak dari setiap data terhadap masing-masing *centroid*. Dan dapat dilihat sudah ada dan cukup banyak para peneliti terdahulu menggunakan dua metode tersebut, berikut beberapa penelitian terdahulu yang mengimplementasikan metode *K-Means* dan *K-Medoids* sebagai klasterisasi data serta alat pengambil Keputusan, dari penelitian tersebut menggunakan dua metode, analisis cluster yang digunakan pada penelitian ini adalah *K-Means* dan *K-Medoids* dengan validitas Silhouette. Berdasarkan hasil dari penelitian analisis diketahui bahwa nilai Silhouette *k-means* (0.8018) lebih tinggi dibandingkan dengan nilai Silhouette *k-medoids* (0.7281). dalam hal ini dapat disimpulkan bahwa metode *k-means* lebih baik dibandingkan dengan *k-medoids*. Hasil analisis cluster menggunakan *K-Means* adalah 1) PTN dengan produktivitas publikasi ilmiah tinggi yaitu ITB, ITS, UGM, dan UI. Keempat PTN tersebut adalah PTN Badan Hukum (PTN-BH) yang berlokasi di Pulau Jawa, 2) PTN dengan produktivitas publikasi ilmiah sedang terdiri dari 16 PTN yang didominasi oleh PTN-BH dan PTN Badan Layanan Umum (PTN-BLU) dengan jumlah terbesar lokasi di Pulau Jawa, dan 3) PTN dengan produktivitas publikasi ilmiah rendah terdiri dari 102 PTN yang didominasi oleh PTN sebagai pengelola keuangan negara umum (PTN-Satker) dengan lokasi terbanyak di luar Pulau Jawa [2]. Dari penelitian terdahulu di atas dapat diketahui bahwa perbandingan dua metode antara *K-Means* dan *K-Medoids* dapat diimplementasikan dengan baik menggunakan data klasterisasi yang ada secara terintegrasi sesuai kebutuhan setiap instansi, dan dapat diproses dari beberapa tahapan dari metode *K-Means* maupun *K-Medoids*, kedua metode ini juga memiliki beberapa kelebihan dan kekurangan yang bisa dipertimbangkan sebagai alat pengambil keputusan yang efektif sesuai fungsi dan kebutuhan objek/intansi [3].

Penerapan perbandingan dari dua metode tersebut akan memberikan dan menghasilkan metode mana yang cocok dan sesuai fungsi kebutuhan sehingga digunakan untuk pengelompokan klasterisasi buku. Oleh karena itu pada penelitian ini akan di implementasikan metode tersebut dan dicari hasil akurasi dengan metode *K-Means* dan *K-Medoids* dan hasil pengelompokan dari tingkat akurasi klasterisasi tersebut, sehingga dapat menggali informasi mengenai hasil data mining pada pengelompokan buku perpustakaan universitas Bina Darma Palembang.

2. METODE PENELITIAN

Data mining sering juga disebut dengan *Knowledge Discovery in Database* atau disingkat KDD, adalah kegiatan yang meliputi pengumpulan, pemakaian data histori untuk menelusuri data yang ada untuk membangun sebuah model agar dapat mengenali pola data yang lain berukuran besar dan suatu teknik dalam pengambilan keputusan suatu informasi dengan pola yang terdapat pada sekumpulan data yang ada di database dengan cakupan jumlah yang cukup besar [4]. Data mining dapat didefinisikan sebagai proses data yang tidak diketahui dapat diverifikasi, dan mungkin bermanfaat untuk informasi [5]. Beberapa pandangan mengenai data mining juga dapat mengekstraksi (mengambil intisari) pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basis data dan manajemen data, prapemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pasca pemrosesan terhadap struktur yang ditemukan, visualisasi, dan online updating data[6].

Data mining merupakan suatu kegiatan yang didalamnya terdapat aktifitas seperti proses penggalian data, pengumpulan data, pencarian pengetahuan serta pemakaian data sebelumnya dengan tujuan yaitu menghasilkan suatu informasi, pengetahuan, pola, keteraturan atau hubungan antar data dengan jumlah yang sangat besar [7]. Metode *K-Means* merupakan suatu cluster analysis non hirarki [8] yang yang mengelompokkan suatu data berdasarkan karakteristiknya kedalam satu atau lebih cluster, data yang memiliki kesamaan variabel dikelompokkan dalam suatu cluster yang sama dan sebaliknya [9].

K-Medoids merupakan pengelompokan partisi menggunakan sekumpulan objek yang dikelompokkan sebanyak jumlah *cluster*, dimana setiap objek mewakili *cluster* tersebut. Objek yang mewakilinya disebut dengan medoids[5]. Medoid merupakan perwakilan *cluster* yang dijadikan sebagai pusat perhitungan jarak [10]. Titik acuan yang digunakan oleh K- Medoids tidak diambil dari perhitungan nilai mean dari objek dalam cluster seperti pada metode K-Means. Algoritma K-Medoids mengambil titik acuan berdasarkan objek representatif, dengan model partisi meminimalkan jumlah dari ketidaksamaan antara setiap objek dan titik acuan yang sesuai [11]. Adapun tahapan algoritma *K-Means* adalah sebagai berikut [12]:

1. Tentukan k sejumlah cluster yang akan dicari
2. Pilih titik centroid secara berurutan atau acak dari data awal sebanyak jumlah k.
3. Hitung jarak masing-masing data menggunakan titik centroid sebanyak jumlah k.
4. Setiap pusat cluster dihitung ulang berdasarkan dari nilai rata-rata dalam cluster yang didapatkan.
5. Langkah 3 dan 4 diulangi sampai kelompok jarak terkecil tidak mengalami perubahan pola cluster.

Metode *K-Medoids* merupakan pengelompokan partisi menggunakan sekumpulan objek yang dikelompokkan sebanyak jumlah *cluster*, dimana setiap objek mewakili *cluster* tersebut. Objek yang mewakilinya disebut dengan medoids [7].

Tahapan Algoritma *K-Medoids*:

1. Menentukan jumlah cluster yang akan dibentuk (nilai k).
2. Menginisialisasi pusat cluster awal (medoids) dari masing-masing cluster secara acak.
3. Menghitung setiap data observasi (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance.
4. Menghitung secara acak objek pada masing-masing cluster sebagai kandidat medoid baru.
5. Menghitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoid baru.
6. Menghitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoid.
7. Mengulangi langkah 4 sampai 6 hingga tidak terjadi perubahan medoid, sehingga didapatkan cluster beserta anggota cluster masing-masing.

Pada tahapan klasterisasi penelitian ini menggunakan dua perbandingan metode yaitu K-Means dan K-Medoids dimana setiap metode memiliki perbedaan setiap tahapan proses klasterisasi data, dan berikut tahapan algoritma metode K-Means dan K-Medoids.

2.1 Langkah-langkah algoritma K-Means

Berikut ini merupakan langkah-langkah algoritma *K-Means* [13]:

1. Mentukan jumlah Cluster (k) pada data set.
2. Tentukan titik pusat (Centroid) secara random.

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_1} X_{kj} \quad (1)$$

Keterangan:

- V_{ij} = Centroid rata-rata cluster ke-i untuk variable ke-j
 - N_i = Jumlah anggota cluster ke-i
 - i, k = Indeks dari cluster
 - j = Indeks dari variable
 - X_{kj} = Nilai data ke-k variable ke-j untuk cluster tersebut
3. Untuk menghitung jarak terdekat dengan centroid, dapat menggunakan rumus di bawah ini :

$$d = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2)$$

Keterangan:

- d = *Euclidean Distance*
 - i = banyak objek
 - x, y = Titik koordinat objek
 - s, t = Titik koordinat centroid
- a. Golongkan objek menurut jarak ke centroid terdekat.
 - b. Lakukan kembali langkah ke-3 dan langkah ke-4 dan iterasi mencapai centroid bernilai optimal.

2.2 Langkah-langkah algoritma K-Medoids

Berikut ini merupakan langkah-langka algoritma *K-Medoids* [14]:

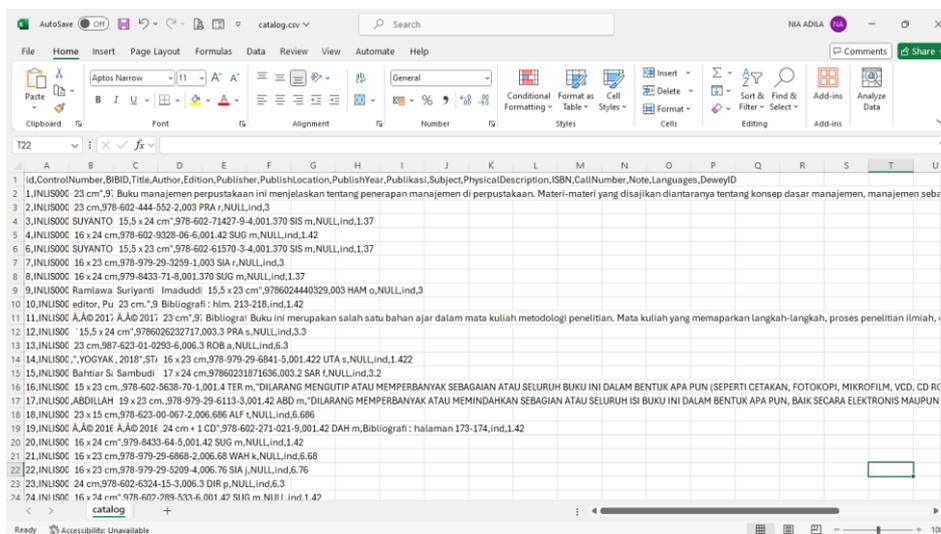
1. Inisialisasi pusat cluster sebanyak k (jumlah cluster)
2. Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan persamaan:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - s_i)^2} \quad (3)$$

3. Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoids baru.
4. Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoids baru.
5. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoid.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoids, sehingga didapatkan cluster beserta anggota clustermasing-masing.

Adapun fase dalam penelitian ini terdiri dari beberapa tahapan hingga mendapatkan hasil yang akurat. Penelitian ini menggunakan metode *K-Means* dan *K-Medoids* sebagai pengembangan model baru terhadap pengelompokkan klasterisasi data buku. *Tools* yang digunakan adalah *Google colab* untuk mengolah data-data yang ada. Pada proses analisa data penelitian ini menggunakan *Knowledge Discovery in Databases* atau disingkat KDD, berikut analisa data pada Perpustakaan Universitas Bina Darma ada 6 tahapan yaitu [15].

1. Pembersihan data merupakan proses menghilangkan noise pada data buku perpustakaan yang tidak konsisten atau data tidak relevan. Berikut hasil proses data *cleaning* data buku perpustakaan Univistersitas Bina Darma:



Gambar 1. Proses *Cleaning* Data Buku Perpustakaan

2. Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Dari data buku Perpustakaan Universitas Bina Darma yang di dapat yaitu berupa catalogs data buku beberapa tahun yang lalu, dengan jumlah data keseluruhan 17186 data. Kemudian sata di integrasikan dengan membuat database baru yaitu *catalog.csv*, dan hanya mengambil 13 atribut yaitu id, Control Number, BIBID, Title, Author, Edition, Publisher, Publish Year, Subject, Physical Description, Note, Languages, DeweyID.
3. Seleksi Data (*Data Selection*), data yang ada pada *catalog.csv* tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.
4. Transformasi data (*Data Transformation*) Kemudian data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining, metode data mining pada penelitian ini membutuhkan format csv sebelum bisa diaplikasikan pada *tools* program *python*.
5. Proses mining, Pada proses mining menggunakan metode *K-Means* dan *K-Medoids* yang diterapkan untuk menemukan informasi dari data *catalog.csv*.
6. Evaluasi pola (*pattern evaluation*) Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai[7].

Setelah data tersebut melewati fase pengolahan data maka didapatlah data yang akan diolah menjadi pengelompokkan klasterisasi data buku, maka data yang akan digunakan pada penelitian ini adalah 16783 data serta 13 atribut yaitu id, Control Number, BIBID, Title, Author, Edition, Publisher, Publish Year, Subject, Physical

Description, Note, Languages, DeweyID. Adapun tabel dari 13 atribut tersebut yang akan di proses kedalam data mining adalah sebagai berikut ini:

Tabel.1 Tabel Data Atribut

No	Nama Atribut	Keterangan
1	Id	Integer (Id)
2	Control Number	Number (Id)
3	Bibid	Integer(Id)
4	Title	Text (Atribut)
5	Author	Text (Atribut)
6	Edition	Text (Atribut)
7	Publisher	Text (Atribut)
8	Publish Year	Text (Atribut)
9	Subject	Text (Atribut)
10	Physical Description	Text (Atribut)
11	Note	Text (Atribut)
12	Languages	Text(Atribut)
13	DeweyID	Integer (Id)

Berdasarkan tabel atribut di atas, data buku selanjutnya akan diolah pada proses klasterisasi data dengan *read dataset*, *problem statement* dan perbandingan metode klasterisasi K-Means dan K-Medoids.

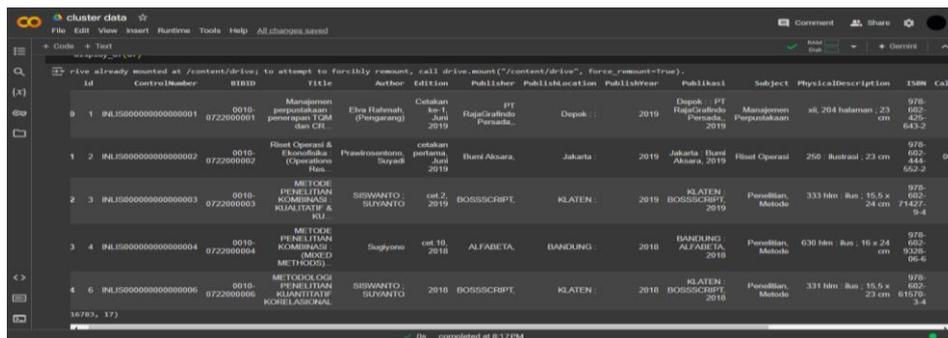
3. HASIL DAN PEMBAHASAN

Setelah melakukan analisis data pada bab sebelumnya, maka pada bab ini merupakan hasil dan pembahasan dari penelitian. Dengan menggunakan data buku perpustakaan Universitas Bina Darma maka akan mendapat beberapa informasi yang dapat diolah untuk beberapa proses implementasi *cluster*.

3.1. Proses Klasterisasi Data

3.1.1. Read Dataset

Pada proses ini akan diinputkan data buku perpustakaan yang diberi nama file *catalog.csv* kemudian dibaca dan menghasilkan berupa tabel seperti berikut ini:

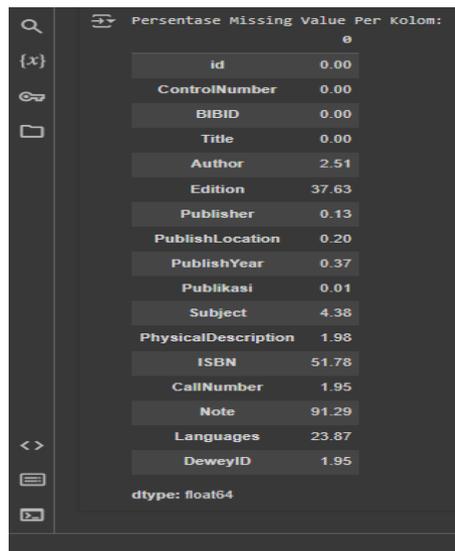


Gambar 2. Hasil read dataset

Dari gambar 2 hasil read dataset data dibaca dan menghasilkan sejumlah 16783 data dengan 17 kolom atribut. Dari read dataset selanjutnya masih diperlukan penanganan bagian missing values dan tipe data.

3.1.2. Problem Statement

Berdasarkan data buku (dan optionally, kolom-kolom lain), kelompokkan (clustering) buku-buku yang ada hingga dapat diatur dengan sesuai atribut yang akan digunakan. Karena datanya dominan berbentuk teks.



Gambar 3 Hasil Persentase Missing Value Per Kolom

Dapat dilihat dari gambar 3 menunjukkan terdapat 17 atribut dan jumlah missing values yaitu id (0.00), Control Number (0.00), BIBID (0.00), Title (0.00), Author (2.51), Edition (37.63), Publisher (0.13), Publish Location (0.20), Publish Year (0.37), Publikasi (0.01), Subject (4.38), Physical Description (1.98), ISBN (51.78), CallNumber (1.95), Note (91.29), Languages (23.87), DeweyID (1.95). dari jumlah missing atribut yang memiliki jumlah missing values terbanyak adalah Note. kemudian akan dilanjutkan pada proses general.

3.1.3. General

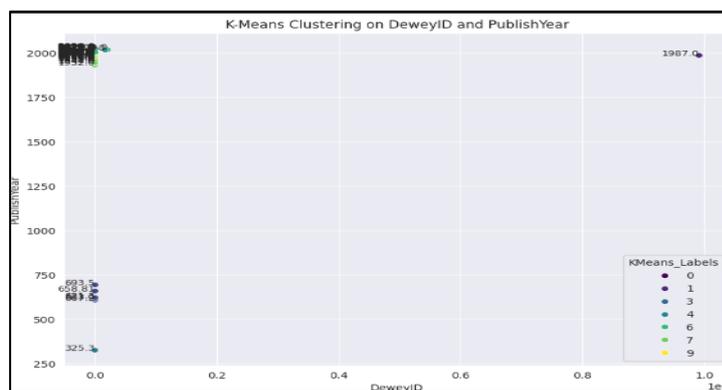
Mengeksplorasi variabel tentang buku-buku yang ada dalam data ini secara umum seperti berikut ini:

1. Ada berapa *unique title* di dalam data ini?
2. Apakah ini adalah buku yang sama atau berbeda? Apakah ada satu buku yang memiliki lebih dari satu ID?
3. Apakah ada judul yang memiliki duplikat hingga 4 kali?
4. Berapa banyak buku yang berada di satu subject umumnya?
5. Berapa banyak subject yang dimiliki satu buku pada umumnya?
6. Berapa banyak buku yang dimiliki satu *author* umumnya?
7. Berapa jumlah bahasa pada data buku?
8. Berapa banyak kategori deweyid pada data buku?

3.2. Proses Klasterisasi *K-Means* dan *K-Medoids*

Setelah proses general peneliti mendapatkan informasi atau topik terbaru yang dapat di implementasikan pada proses klasterisasi *K-Means* dan *K-Medoids*.

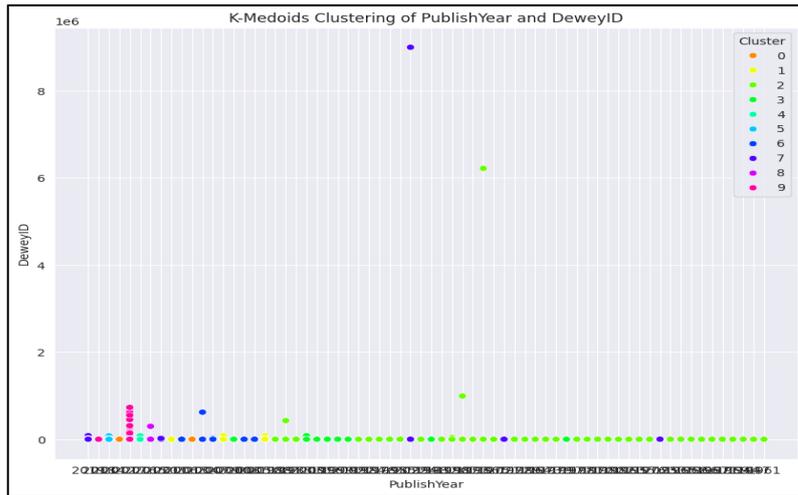
3.3. Klasterisasi *K-Means* DeweyID dengan Publishyear



Gambar 4. Hasil klasterisasi *K-Means* DeweyID dengan Publishyear

Dari gambar 4 dapat dilihat untuk mengetahui *cluster* pada data yaitu atribut DeweyID dan Publishyear dan terdapat 10 *cluster* yang dihasilkan. Dari hasil *cluster* yang ada diketahui dapat kita lihat perbandingan setiap *cluster* seperti pada tahun 2000 memiliki jumlah 1987 deweyid di *cluster* 0 dan kemudian *cluster* 1 sampai *cluster* 3 memiliki jumlah 693 deweyid dan paling rendah dengan jumlah 325 deweyid pada dibawah tahun 2000. pada tahun 2000 di *cluster* 4 sampai 9 memiliki jumlah deweyid terbanyak dan diketahui dari jumlah *cluster* yang terdapat pada gambar bahwa *cluster* 0 merupakan K-Means karena diketahui bahwa jumlah dan posisi *cluster* terletak di tengah diantara jumlah *cluster* tertinggi dan terendah.

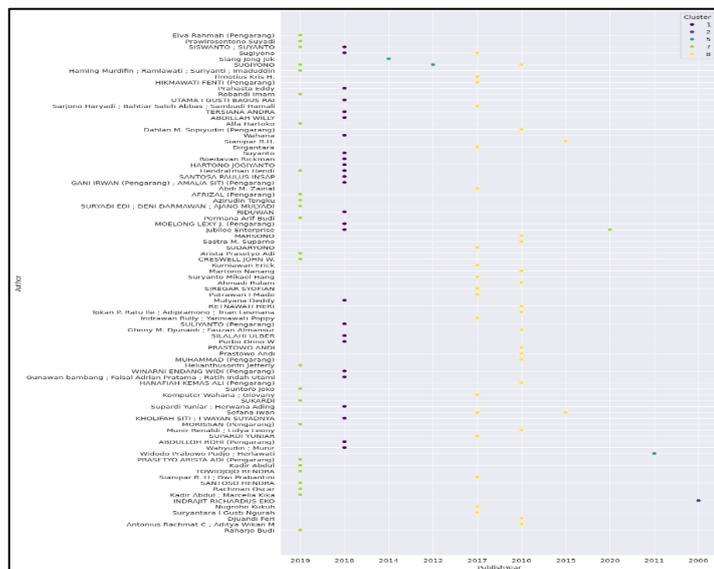
3.4. Klasterisasi K-Medoids DeweyID dengan Publishyear



Gambar 5. Hasil klasterisasi K-Medoids

Dari gambar 5 dapat dilihat untuk mengetahui K-Medoids *cluster* pada data yaitu atribut DeweyID dan Publishyear dan terdapat 10 *cluster* yang dihasilkan. Dari hasil *cluster* yang ada diketahui *cluster* diketahui bahwa pada tahun 2012 memiliki jumlah 1.370 deweyid yaitu di *cluster* 0, pada tahun 2014 memiliki jumlah 3.000 deweyid yaitu di *cluster* 5, dan pada tahun 2017 dengan jumlah 1.420 deweyid yaitu di *cluster* 9, untuk pada tahun 2018 dan 2019 terletak di satu *cluster* yang sama dengan jumlah deweyid tertinggi yaitu 25000 yaitu di *cluster* 7. Oleh karena itu dapat diketahui dari jumlah *cluster* yang terdapat pada gambar bahwa *cluster* 0 dan 5 merupakan titik pusat K-Medoids karena diketahui dari jumlah dan posisi *cluster* terletak pada diantara *cluster* 7 sampai 9.

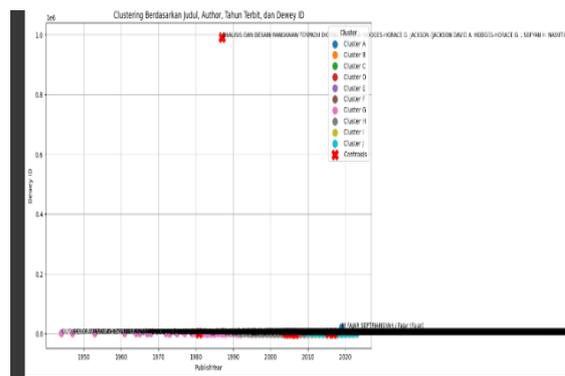
3.5. Klasterisasi K-Means Publishyear dengan Author



Gambar 6. Hasil Klasterisasi K-Means

Dari gambar 6 dapat dilihat untuk mengetahui K-Means cluster pada data yaitu atribut PublishYear dan Author dan terdapat 10 cluster yang dihasilkan. Dari hasil cluster yang ada diketahui cluster diketahui bahwa pada tahun 2019 dan 2020 sama-sama terletak di satu cluster tapi memiliki perbedaan jumlah author 2019 memiliki jumlah 25 author dan tahun 2020 memiliki jumlah 1 author yaitu di cluster 7, dan tahun 2018 memiliki jumlah 26 author yaitu di cluster 1, kemudian tahun 2014, 2012 dan 2011 dengan jumlah masing-masing 1 author terletak di cluster 5. Untuk tahun 2017,2016, dan 2015 juga terletak di satu cluster dengan jumlah author yang berbeda pada tahun 2017 dengan jumlah 18 author, 2016 yaitu 16 author sedangkan tahun 2015 hanya 2 author jadi total keseluruhan 3 tahun tersebut 36 author yaitu di cluster 8. Maka diketahui dari jumlah cluster yang terdapat pada gambar bahwa cluster 5 merupakan K-Means karena dilihat dari jumlah dan posisi cluster terletak ditengah diantara cluster 7 sampai 8.

3.6. Klasterisasi K-Means dengan Objek Variabel



Gambar 7. Hasil Klasterisasi K-Means Dengan Objek Variabel

Berikut penjelasan dari hasil tabel K-Means dengan objek Title, Author, PublishYear, dan DeweyID menggunakan 10 cluster. Type data Title (String), Author (String), PublishYear (Number), DeweyID (Float), Title dan author tidak bisa langsung di proses jadi di lakukan proses "Data Encoding" yang dapat mengubah data kategorikal menjadi data numerik sehingga data dapat diproses algoritma menggunakan K-Means hasil tabel merupakan hasil dari coding proses sederhana algoritma K-Means, pertama algoritma K-Means membantu menghitung jarak (dengan euclidean distance) antara data ke pusat cluster (centroid). Kedua data yang lebih dekat ke pusat cluster tertentu akan dimasukkan ke cluster tersebut penyebab bergabungnya data dalam 1 cluster, meski Title 1 dengan Title lainnya adalah title yang berbeda, representasi numerik mereka mungkin cukup dekat, sehingga keduanya dianggap "mirip" oleh algoritma, kolom PublishYear dan DeweyID dapat memberikan kontribusi signifikan terhadap jarak total, sehingga fitur seperti Title dan Author menjadi kurang dominan.

3.7. Evaluasi Perbandingan Klasterisasi Data

Dari proses kalasterisasi dua metode K-Means dan K-Medoids dapat dilihat perbandingan dari hasil cluster, yaitu membuktikan jika metode K-Means memiliki 10 cluster penuh dan memiliki titik tengah atau rata-rata yang dapat disimpulkan dan dikelompokan untuk sebuah alat pengambil keputusan berupa informasi dari jumlah setiap atribut pada data. Sedangkan metode K-Medoids hanya memiliki 10 cluster penuh dan 1 medoid yaitu yang mewakili setiap titik cluster seperti cluster 1 menunjukkan pusat cluster yang paling representatif dari setiap kelompok.

4. KESIMPULAN

Dapat dilihat dari hasil beberapa perbandingan cluster diatas dengan menggunakan dua metode yaitu metode K-Means dan K-Medoids mendapatkan suatu informasi yaitu perbandingan cluster dengan mengeksplorasi variabel tentang buku-buku yang ada dalam data ini secara umum, dan kemudian diproses pada klasterisasi seperti Klasterisasi K-Means DeweyID dengan Publishyear, Klasterisasi K-Medoids DeweyID dengan Publishyear dan Klasterisasi K-Means PublishYear dan Author dan Klasterisasi K-Means dengan objek variable membuktikan bahwa dari data buku banyak informasi yang bisa didapatkan dan akan mempermudah proses analisa pelaporan pada perpustakaan Universitas Bina Darma. Sehingga dapat disimpulkan bahwa metode K-Means dan K-Medoids bisa digunakan untuk metode Analisa data buku pada perpustakaan tapi dari penelitian ini dihasilkan lebih dominan menggunakan metode K-Means yang banyak menghasilkan perbandingan cluster pada data buku perpustakaan Bina Darma. Dan perbandingan setiap cluster akan digunakan sebagai pengambilan keputusan bagi

Perpustakaan Bina Darma. Diharapkan bagi pihak Perpustakaan Universitas Bina Darma untuk lebih memanfaatkan data buku yang ada, tidak hanya di *entry* saja berbentuk catalog dan disimpan dalam bentuk dokumen tetapi lebih di *update* untuk informasi data buku terbaru dalam bentuk *dashboard* informasi. Untuk penelitian selanjutnya dapat menambahkan proses klusterisasi dengan metode yang *datamining* yang lain seperti DBSCAN atau Agglomerative Clustering.

DAFTAR PUSTAKA

- [1] H. D. Tampubolon, D. Gultom, L. Y. Hutabarat, F. I. R.H Zer, and D. Hartama, "Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 146–151, 2020, doi: 10.36294/jurti.v4i1.1263.
- [2] E. Ermawati, I. Sriliana, and R. Sriningsih, "Clustering of State Universities in Indonesia Based on Productivity of Scientific Publications Using K-Means and K-Medoids," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 3, pp. 1617–1630, 2023, doi: 10.30598/barekengvol17iss3pp1617-1630.
- [3] N. A. Privandhani and Sulastri, "Clustering Pop Songs Based On Spotify Data Using K-Means And K-Medoids Algorithm," *J. Mantik*, vol. 6, no. 2, pp. 1542–1550, 2022.
- [4] A. D. Andini and T. Arifin, "Implementasi Algoritma K-Medoids Untuk Klusterisasi Data Penyakit Pasien Di Rsud Kota Bandung," *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 128–138, 2020, doi: 10.51977/jti.v2i2.247.
- [5] H. Zhang, L. Song, and S. Zhang, "Parallel Clustering Optimization Algorithm Based on MapReduce in Big Data Mining," *IAENG Int. J. Appl. Math.*, vol. 53, no. 1, 2023.
- [6] C. Wang, S. J. Zhao, Z. Q. Ren, and Q. Long, "Place-Centered Bus Accessibility Time Series Classification with Floating Car Data: An Actual Isochrone and Dynamic Time Warping Distance-Based k-Medoids Method," *ISPRS Int. J. Geo-Information*, vol. 12, no. 7, 2023, doi: 10.3390/ijgi12070285.
- [7] R. K. Purba and E. Bu'ulolo, "Implementasi Algoritma K-Medoids dalam Pengelompokan Mahasiswa yang Layak Mendapat Bantuan Uang Kuliah Tunggal," *INSOLOGI J. Sains dan Teknol.*, vol. 1, no. 2, pp. 79–86, 2022, doi: 10.55123/insologi.v1i2.195.
- [8] G. Wang, "Time Series Symbolization Method for the Data Mining K-Means Algorithm," *Discret. Dyn. Nat. Soc.*, vol. 2023, 2023, doi: 10.1155/2023/5365673.
- [9] A. Jauhari, D. R. Anamisa, and F. A. Mufarroha, "Analysis of Clusters Number Effect Based on K-Means Method for Tourist Attractions Segmentation," *J. Phys. Conf. Ser.*, vol. 2406, no. 1, pp. 1–8, 2022, doi: 10.1088/1742-6596/2406/1/012024.
- [10] N. Mirantika, T. S. Syamfithriani, and R. Trisudarmo, "Implementasi Algoritma K-Medoids Clustering Untuk Menentukan Segmentasi Pelanggan," *J. Nuansa Inform.*, vol. 17, no. 1, pp. 2614–5405, 2023, [Online]. Available: <https://journal.uniku.ac.id/index.php/ilkom>.
- [11] A. R. Rhamadani, N. N., Fauzi, A., Nurlaelasar, E., & Pratama, "Implementasi algoritma k-means dan k-medoids dalam pengelompokan nilai ujian nasional tingkat smk," ... *Innov.*, no. Ciastech, pp. 717–726, 2020.
- [12] Y. Kustiyahningsih, E. Rahmanita, E. M. S. Rochman, Imamah, A. Amalina, and M. H. A. Sobri, "Decision Support System of Salt Points Grouping Using K-Means Method," *J. Phys. Conf. Ser.*, vol. 2406, no. 1, pp. 1–9, 2022, doi: 10.1088/1742-6596/2406/1/012022.
- [13] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022, doi: 10.35957/jatisi.v9i1.1516.
- [14] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [15] H. Jiawei and K. Micheline, *Data mining: concepts and techniques second edition*. 2006.