

Pengelompokan Kategori Buku Berdasarkan Judul Menggunakan Algoritma *Agglomerative Hierarchical Clustering* Dan *K-Medoids*

Muhammad Ainur Rofiq¹, Anita Qoiriah²,

¹ Jurusan Teknik Informatika/Teknik Informatika, Universitas Negeri Surabaya

² Jurusan Teknik Informatika/Teknik Informatika, Universitas Negeri Surabaya

¹muhammadrofiq@mhs.unesa.ac.id

²anitaqoiriah@unesa.ac.id

Abstrak—Perpustakaan merupakan tempat yang menyediakan berbagai bahan pustaka yang dibutuhkan semua orang khususnya pelajar. Namun koleksi buku dalam jumlah banyak juga akan menyulitkan pustakawan untuk melakukan pengelolaan letak buku sehingga ada buku yang tidak terbaca dan jarang dipinjam. Untuk penyelesaian masalah tersebut akan digunakan teknik clustering untuk melakukan pengelompokan buku berdasarkan kategorinya. Ada beberapa algoritma clustering salah satunya *K-Medoids*. Selain menggunakan algoritma *K-Medoids*, pada penelitian ini juga menggunakan algoritma *Agglomerative Hierarchical Clustering* (AHC). Teknik clustering akan mengelompokkan judul buku sesuai kategorinya, dan judul buku yang ada dalam satu cluster akan digunakan sebagai bahan untuk analisis dalam mempermudah pustakawan untuk pengambilan keputusan dan melakukan pengelolaan penataan buku yang diminati dan merancang strategi untuk meningkatkan minat baca khususnya pada pelajar.

Kata Kunci— Perpustakaan, *clustering*, *K-Medoids*, *Agglomerative Hierarchical Clustering*.

I. PENDAHULUAN

Ruang baca atau perpustakaan sangat diperlukan disebuah lembaga pendidikan, baik sekolah maupun perguruan tinggi. Seiring berkembangannya ilmu pengetahuan menuntut perpustakaan untuk menyediakan koleksi buku dalam jumlah besar. Dengan koleksi buku yang banyak, salah satu masalah yang ditimbulkan adalah sulitnya pengelola ruang baca untuk melakukan penataan tata letak buku, sehingga buku yang tersedia tidak terbaca dan jarang dipinjam. pengelola ruang harus melakukan penataan buku berdasarkan kategori buku tersebut, agar buku yang tersedia mudah dicari dan dapat meningkatkan minat baca dari pengunjung di ruang baca.

Dalam menangani permasalahan diatas memerlukan sebuah solusi yang dapat mempermudah proses pengelompokan koleksi buku yang tersedia dapat dimanfaatkan secara maksimal. Penentuan kategori dapat berdasarkan judul buku yang tersedia. Karena judul buku dapat mewakili seluruh isi buku.

Clustering adalah metode yang ada pada data mining yang digunakan untuk memetakan data ke dalam beberapa kelompok yang lebih kecil dengan berdasarkan kesamaan karakteristik yang dimiliki. Beberapa teknik yang digunakan dalam melakukan clustering, seperti *K-Means*, *K-Medoids* dan lainnya. Teknik *K-Means* diperkenalkan pada tahun 1967 oleh MacQueen, sedangkan *K-Medoids* diperkenalkan pada tahun 1990 oleh Kaufman dan Rousseeuw [1]. Algoritma *K-Medoids* merupakan salah satu algoritma clustering, kelebihan dari metode *K-Medoids* yaitu dapat mengatasi kelemahan metode *K-Means* yang sensitive terhadap *outlier*. Algoritma *K-Medoids* dimulai dengan menentukan centroid secara acak. Kemudian hitung jarak dari setiap data terhadap masing – masing *centroid*. Dalam hal ini perhitungan jarak menggunakan rumus *Euclidian Distance*. Berdasarkan penelitian yang telah dilakukan oleh (Aishwaya) mendapat kesimpulan bahwa algoritma *K-Means* dinilai lebih efisien untuk melakukan clustering dengan dataset lebih kecil sedangkan *K-Medoids* memiliki performa yang lebih baik untuk dataset yang besar [2].

Algoritma *K-Medoids* memiliki tingkat akurasi sebesar 0.584 dengan jumlah *cluster* sebanyak lima *cluster*. Dalam penelitian ini juga ditambahkan metode *cosine similarity* yang dapat meningkatkan akurasi menjadi 0.55 yang semula yang memiliki nilai rata – rata 0.44 [3]. Pada penelitian yang lain disimpulkan beberapa hal, yaitu program aplikasi otomatisasi klasifikasi buku perpustakaan yang menggunakan bahasa Indonesia dengan menggunakan metode KNN nilai rata-rata akurasinya sebesar 72% dengan menggunakan jumlah data uji 50 buah dan rata-rata

untuk proses klasifikasi memerlukan waktu selama 2 menit 48 detik, apabila menggunakan metode KNN digabungkan dengan *K-Medoids* rata-rata akurasiya meningkat yaitu sebesar 84% dengan menggunakan 50 data uji dan proses klasifikasi memerlukan waktu selama 39,4 detik. Sedangkan klasifikasi dengan menggunakan metode KNN yang dikombinasikan dengan *K-Medoids* menghasilkan akurasi yang lebih tinggi serta memerlukan waktu lebih singkat jika dibandingkan dengan klasifikasi yang hanya menggunakan metode KNN [4].

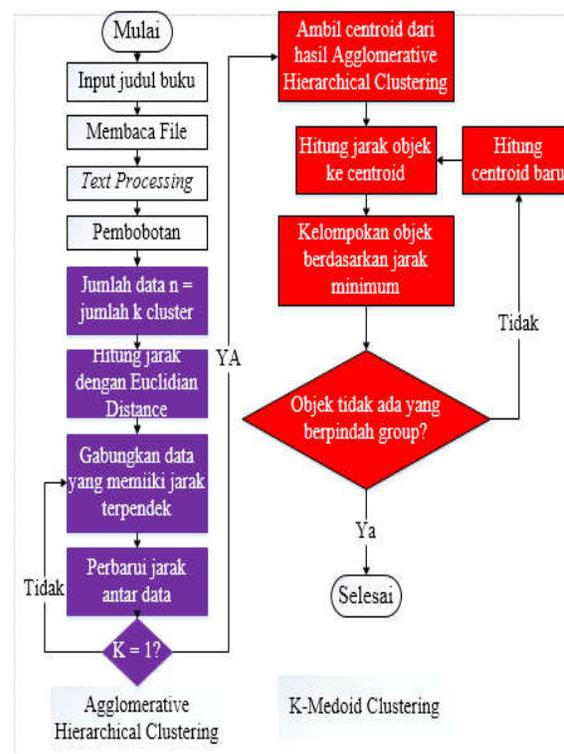
Selain algoritma *K-Means* dan *K-Medoids* ada juga algoritma *Agglomerative Hierarchical Clustering*. *Agglomerative Hierarchical Clustering (AHC)* adalah metode *clustering* yang memiliki sifat *bottom-up* yaitu menggabungkan n buah *cluster* menjadi satu *cluster* tunggal. Penelitian terdahulu tentang kombinasi algoritma *K-Means* dan *Agglomerative Hierarchical Clustering*. Dalam penelitian ini menyimpulkan bahwa kombinasi algoritma *AHC* dan *K-Means* menghasilkan pengelompokan data yang lebih baik jika dibandingkan dengan *K-Means* dalam semua pengujian [5].

Karena algoritma *K-Medoids* memiliki kekurangan, yaitu didalam algoritma ini terdapat *centroid* awal yang diambil secara acak, hal tersebut membuat hasil dari *clustering* kurang maksimal. Untuk itu dalam penelitian ini *centroid* awal akan dihitung menggunakan algoritma *Agglomerative Hierarchical Clustering*. Pada penelitian ini mengkombinasikan kelebihan dari *Agglomerative (Bottom-Up)* dan *K-Medoids (Top-Down)*. Dengan menggabungkan dua algoritma tersebut penelitian ini memiliki kelebihan dari penelitian sebelumnya yang hanya menggunakan satu algoritma dan nilai *centroid* awal yang ditentukan secara acak akan mempengaruhi hasil dari *clustering* yang akan dilakukan. Algoritma *bottom-up* sangat baik dalam mengidentifikasi kelompok kecil sedangkan algoritma *top-down* baik dalam mengidentifikasi kelompok besar [6].

II. METODE PENELITIAN

Pada penelitian kali ini bertujuan melakukan pengelompokan judul buku dengan menerapkan metode *clustering* dengan menggunakan algoritma *K-Medoids* yang digabung dengan *Agglomerative Hierarchical Clustering* yang digunakan untuk menghitung *centroid* awal dan *K-Medoids* untuk proses *clustering* kategori buku. Pada tahap *text processing*, koleksi dokumen yang berupa *string* karakter akan direpresentasikan dalam bentuk vektor kata. Dimana koleksi dokumen akan berupa matriks yang jumlah baris sesuai dengan jumlah kata dalam koleksi dokumen. Vektor kata akan disimpan dan nantinya akan diproses lebih lanjut ke tahap pembobotan kata. Pada tahap ini akan dilakukan

pemberian bobot pada setiap token dengan algoritma TF-IDF. Maka setiap token nantinya akan diwakili bobot yang akan digunakan untuk proses *clustering*. Matriks bobot yang dihasilkan pada proses pembobotan TF-IDF akan diolah menggunakan metode AHC. Hasil dari proses AHC untuk mengetahui *centroid* awal yang nantinya akan digunakan pada proses *K-Medoids*. Proses perhitungan algoritma *K-Medoids* ini menghasilkan pengelompokan terbaik untuk menentukan *cluster* kategori buku. Gambaran dari alur *clustering* dapat dilihat pada Gbr 1.



Gbr 1. Alur *Clustering*

1. Input dataset

Dataset yang digunakan yaitu dataset buku yang disimpan dengan format .xlsx. dengan jumlah data sebanyak 20.000 judul buku yang berbahasa indonesia dan juga bahasa asing.

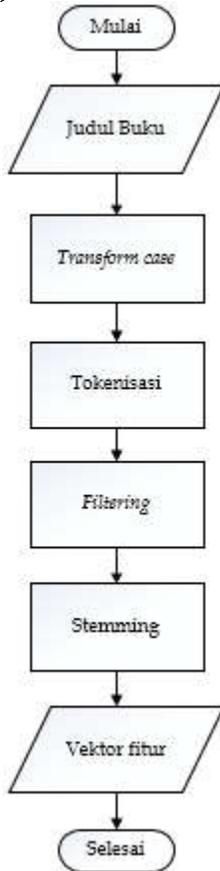
2. Membaca file dataset

Setelah dataset disimpan dalam file excel dengan format .xlsx langkah selanjutnya yakni melakukan pembacaan file .xlsx untuk mengambil isi dari dataset tersebut yang berupa judul buku dimana hasil dari ini adalah berupa fitur.

3. Text Processing

Setelah data berhasil terbaca oleh program, selanjutnya sistem akan melakukan *text processing*.

Tahapan ini merupakan langkah awal dalam pengelompokan dimana mengubah dokumen yang berupa *string* karakter menjadi representasi vektor kata yang lebih mudah diolah. Alur *text processing* dapat dilihat pada Gbr 2.



Gbr 2. Alur *Text Processing*

- a. Data yang diinputkan merupakan judul dari buku
- b. *Transform case* : Mengubah seluruh kata dalam *dataset* dari huruf kapital menjadi huruf kecil.
- c. *Tokenisasi* : Pemotongan kalimat menjadi sebuah kata yang berdiri sendiri.
- d. *Filtering* : Pengambilan kata penting dari hasil token. Pada tahap ini dapat menggunakan algoritma *stopword* yaitu menghilangkan atau membuang kata kurang penting atau *wordlist* yaitu menyimpan kata yang penting.
- e. *Stemming*
Merupakan proses untuk membentuk kata dasar dari suatu kata.

4. Pembobotan Ekstraksi Fitur

Pembobotan yaitu suatu proses untuk melakukan pemberian bobot kepada suatu kata yang ada pada setiap dokumen, hal ini didasarkan pada jumlah kemunculan suatu kata didalam dokumen. Metode TF-IDF merupakan metode yang sering dipakai

dalam *text mining* dan melakukan penelusuran informasi [7]. TF adalah jumlah kemunculan sebuah kata atau *term* (dianggap *t*) dalam suatu dokumen (dianggap *d*). Berikut ini adalah persamaan untuk menghitung TF :

$$TF(t, d) = f(t, d) \dots\dots\dots(1)$$

Dimana :
 - $f(t,d)$ = kemunculan kata *t* dalam dokumen *d*

DF atau document frequency adalah banyaknya dokumen yang dimiliki oleh term *t*. Berikut adalah persamaan untuk menghitung invers DF atau IDF :

$$IDF_{(t)} = \log \left(\frac{N}{nt} \right) \dots\dots\dots(2)$$

Dimana :
 - *N* = jumlah dokumen.
 - *nt* = jumlah dokumen yang mengandung kata *t*.

Sehingga pada TF-IDF ini akan memberikan bobot pada kata yang diambil berdasarkan nilai inverse DF dan nilai TF. Berikut adalah persamaan untuk mencari bobot menggunakan TF-IDF :

$$TF - IDF = f_{(t,d)} * \log \left(\frac{N}{nt} \right) \dots\dots\dots(3)$$

5. Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering (AHC) adalah suatu metode hierarchical clustering yang memiliki sifat bottom-up yaitu menggabungkan *n* ke dalam satu kelompok tunggal [8]. Dalam algoritma AHC terdapat tiga teknik pengelompokan :

1. *Single linkage*

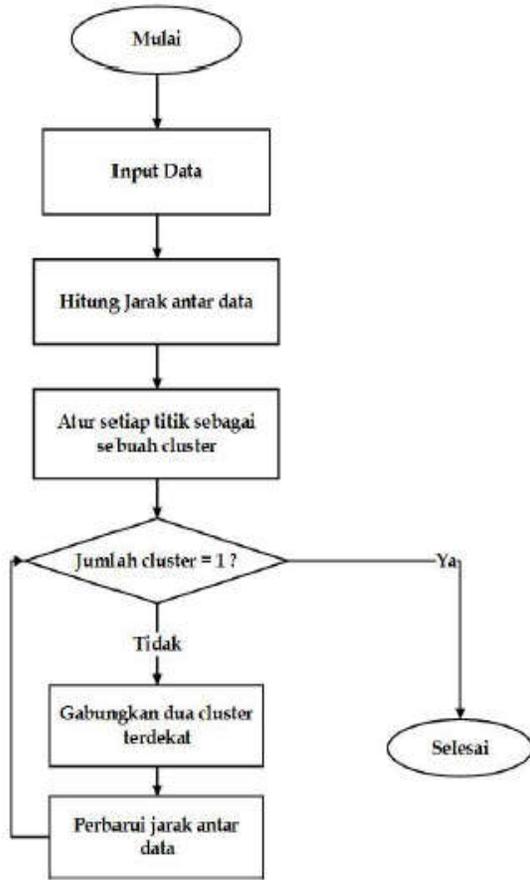
Teknik menggabungkan obyek yang didasarkan pada jarak terkecil.

2. *Complete linkage*

Teknik menggabungkan obyek yang didasarkan pada jarak terjauh.

3. *Average linkage*

Teknik menggabungkan obyek yang didasarkan pada jarak rata-rata pasangan anggota yang ada dalam himpunan antara dua buah kelompok. Flowchart dari algoritma AHC dapat dilihat pada Gbr 3.



Gbr 3. Flowchart AHC

Langkah – langkah perhitungan *Agglomerative Hierarchical Clustering* [9] adalah sebagai berikut :

1. Pada penelitian ini perhitungan jarak menggunakan *Euclidean Distance*. Seperti pada rumus dibawah ini :

$$x_{rs} = \sqrt{\sum_{k=1}^p (y_{rk} - y_{sk})^2} \dots (4)$$

Ketrangan :

- x_{rs} = Jarak antara objek r dengan s.
- y_{rk} = Nilai objek r pada variabel ke k.
- y_{sk} = Nilai objek s pada variabel ke k.
- p = Banyaknya variabel yang diamati.

2. Penggabungan dua objek terdekat
 Apabila nilai jarak dari objek a dan b yang terkecil dibanding jarak antar objek yang lain, maka gabungan dua objek yang pertama yaitu x_{ab} .

3. Perbarui matriks jarak
 Apabila x_{ab} merupakan jarak paling kecil dalam perhitungan jarak yang menggunakan *Euclidian Distance*, maka rumus untuk metode *agglomerative* sebagai berikut:
 - a. Rumus *single linkage*

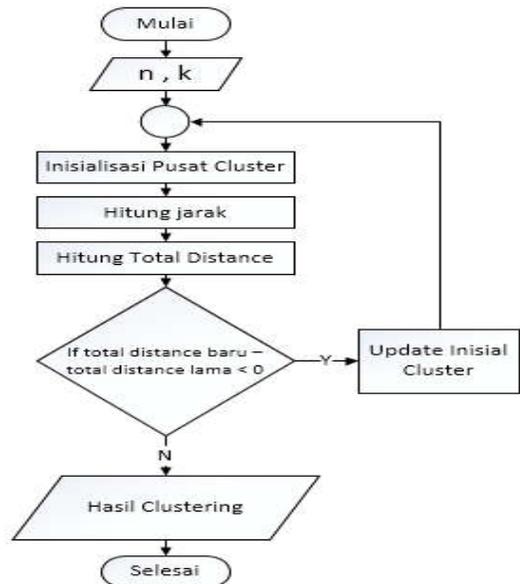
$$d_{(ab)c} = \min\{d_{a,c}; d_{b,c}\} \dots (5)$$
 - b. Rumus *average linkage*

$$d_{(ab)c} = \text{average}\{d_{a,c}; d_{b,c}\} \dots (6)$$
 - c. Rumus *Max linkage*

$$d_{(ab)c} = \max\{d_{a,c}; d_{b,c}\} \dots (7)$$
4. Selanjutnya ulang kembali langkah ke 2 dan langkah ke 3 sampai tersisa satu *cluster*.

6. Clustering menggunakan metode k-medoids

Algoritma *K-Medoids* serupa dengan algoritma *K-Means* yang sama – sama algoritma clustering namun tetap memiliki perbedaan yaitu sebagai pusat clusternya algoritma *K-Means* menggunakan nilai mean (rata – rata) sedangkan sebagai pusat clusternya algoritma *K-Medoids* menggunakan objek perwakilan (medoid) [10]. *K-Means* memiliki kelemahan dimana algoritma *K-Means* begitu sensitif pada *outlier* dikarenakan sebuah objek yang memiliki nilai besar secara substansial menyimpang dari distribusi data, dari kelemahan tersebut algoritma *K-Medoids* hadir untuk mengatasi kekurangan *K-means* [11]. Flowchart *K-Medoids* dapat dilihat pada Gbr 4.



Gbr 4. Flowchart *K-Medoids*

Berikut ini adalah tahapan algoritma *K-Medoids* :

1. Menentukan nilai jumlah kelompok (k).
2. Menentukan nilai centroid awal secara acak sebanyak jumlah kelompok yang telah ditentukan (k) dari n data.
3. Menghitung jarak dari obyek ke medoid sementara, setelah itu beri tanda pada nilai jarak terkecil obyek tersebut ke medoid kemudian hitung jumlahnya.
4. Langkah selanjutnya melakukan iterasi medoid.
5. Kemudian hitung jumlah simpangan (S).
6. Ulangi langkah ke 3 sampai 5 hingga anggota medoid tidak berubah.

III. HASIL DAN PEMBAHASAN

Penelitian pengelompokan kategori buku berdasarkan judul dengan menggunakan algoritma *K-Medoids* dan *Agglomerative Hierarchical Clustering*. Algoritma AHC digunakan untuk menentukan centroid awal. Aplikasi yang dibuat menggunakan bahasa pemrograman *Python* dan dijalankan menggunakan *google colab*.

A. Data Penelitian

Data yang digunakan pada penelitian ini yaitu data buku perpustakaan UNESA. Dataset yang digunakan terdiri dari 20.000 judul buku yang disimpan dalam bentuk file excel dengan format *.xlsx*. Dataset yang digunakan dalam penelitian ini disimpan berbentuk file *csv* bernama *dataset.csv*. Contoh dataset judul buku dalam penelitian ini, dapat dilihat pada Gbr 6.

337	English grammar in use: a self-study reference and practice book for intermediate students with answer
338	English grammar in use: a self-study reference and practice book for intermediate students with answer
339	English grammar in use: a self-study reference and practice book for intermediate students with answer
340	Teori dan contoh soal teknik elektro (menggunakan bahasa pemrograman MATLAB)
341	Teori dan contoh soal teknik elektro (menggunakan bahasa pemrograman MATLAB)
342	Teori dan contoh soal teknik elektro (menggunakan bahasa pemrograman MATLAB)
343	Teori dan contoh soal teknik elektro (menggunakan bahasa pemrograman MATLAB)
344	Teori dan contoh soal teknik elektro (menggunakan bahasa pemrograman MATLAB)
345	Perencanaan pendidikan: suatu pendekatan komprehensif
346	Perencanaan pendidikan: suatu pendekatan komprehensif
347	Perencanaan pendidikan: suatu pendekatan komprehensif
348	Perencanaan pendidikan: suatu pendekatan komprehensif
349	Perencanaan pendidikan: suatu pendekatan komprehensif
350	John Burroughs America
351	Cara memperoleh sertifikat tanah hak milik, tanah negara dan tanah pemda
352	psikologi Anak
353	Pertolongan pertama
354	Element of biology
355	Media pendidikan biologi avertebrata
356	Memahami sejarah bangsa-bangsa Asia selatan
357	Pending Emas
358	Akuntansi Biaya Pengumpulan Biaya dan Penentuan Harga Pokok buku I
359	The world around you

Gbr 6. Data Judul Buku

B. Baca Data

Proses yang pertama yaitu proses baca data buku oleh sistem. Pada baca data ini, apabila terdapat judul buku yang sama akan dihapus atau

dihilangkan. Sehingga proses ini akan didapatkan dataset judul buku yang *unique*. Pada penelitian ini dari total 20.000 data judul buku, judul buku yang sama akan dihapus sehingga yang berbeda dan dapat digunakan hanya 10.703 seperti yang terlihat pada Gbr 7. Proses ini dilakukan supaya data yang akan diproses selanjutnya menghasilkan kualitas yang lebih baik.

	judul
0	Basic physical chemistry
1	Bentuk evaluasi mata kuliah pada perguruan tin...
2	Concise english handbook
3	Techniques and principles in language teaching
4	Encounters
...	...
10699	Chemical engineering kinetics
10700	Tonggak antologi puisi Indonesia modern 2
10701	Alap-alapan drusilawati
10702	Beginilah menggunakan bahasa Indonesia
10703	Supervisi pengajaran teori dan praktik

10704 rows × 1 columns

Gbr 7. Hasil dari Pembacaan Dataset

C. Preprocessing

Langkah selanjutnya adalah *preprocessing*. *Preprocessing* merupakan proses penghilangan karakter – karakter yang tidak dibutuhkan dalam sistem. Karakter yang dihilangkan dalam preprocessing diantaranya tanda baca dan angka yang ada pada setiap judul buku. Selain itu, karakter huruf besar yang ada disetiap judul akan diubah menjadi huruf kecil. Seperti yang terlihat pada Tabel 1.

TABEL I
HASIL PREPROCESSING

No	Sebelum	Sesudah
1	Matematika I modul 6-9	matematika i modul
2	Kalkulus lanjutan 1	kalkulus lanjutan
3	Buku latihan lotus 123 versi 2.2	buku latihan lotus versi
4	Buku materi pokok 6 matematika IV modul 6-9	buku materi pokok matematika iv modul
5	Kimia 2 petunjuk guru SMU Kelas 2	kimia petunjuk guru smu kelas

D. TF-IDF

Setelah melalui tahap preprocessing, data judul buku akan dilakukan proses pembobotan TF-IDF. Proses pembobotan TF-IDF merupakan proses pemberian bobot pada data judul buku. Hasil dari proses ini berupa matriks bobot sejumlah 10.704 fitur.

Dari keseluruhan total fitur akan dilakukan proses ekstraksi fitur. Proses ekstraksi fitur merupakan proses pengambilan fitur penting dari total keseluruhan fitur. Hal ini dilakukan agar proses lebih cepat dengan hasil yang optimal. Proses ekstraksi fitur akan berpengaruh pada hasil clustering nantinya.

E. Agglomerative Hierarchical Clustering

Proses selanjutnya adalah menentukan nilai centroid menggunakan algoritma *Agglomerative Hierarchical Clustering*. Dan hasilnya yang berupa array akan disimpan dengan nama *init_centroid*. Yang nantinya akan digunakan dalam proses *clustering* menggunakan algoritma *K-Medoids*. Hasil dari perhitungan menggunakan algoritma AHC dapat dilihat pada Gbr 8.

Gbr 8. Hasil dari perhitungan AHC

F. K-Medoids

Setelah menentukan centroid awal dari proses AHC, proses selanjutnya adalah pengelompokan data dengan algoritma *K-Medoids*. Ada beberapa parameter yang diperlukan untuk proses clustering, diantaranya jumlah fitur dan jumlah cluster. Pada penelitian ini dilakukan percobaan sebanyak 24 kali untuk menentukan jumlah fitur maupun jumlah cluster. Dari 24 percobaan yang telah dilakukan, didapatkan jumlah fitur terbaik sebanyak 5000 fitur dengan 25 cluster. Hasil proses clustering akan disimpan ke dalam format file excel untuk mengetahui detailnya.

Preview dan hasil dari clustering yang telah dilakukan dapat dilihat pada Gbr 9 yang merupakan tampilan tabel dari program yang berupa *preview* dan Gbr 10 yang merupakan hasil clustering yang sudah disimpan ke dalam file excel. Detail dari hasil clustering terdapat pada tabel 2.

	cluster	judul
125	kelompok 14	matematika terapan bisnis ekonomi
188	kelompok 14	ilmu ekonomi aspek-aspek metodologinya
190	kelompok 14	pengantar matematika ekonomi
240	kelompok 14	indah ilmu ekonomi mementingkan rakyat
263	kelompok 14	seri sinopsis pengantar ilmu ekonomi ekonomi m...
...
9437	kelompok 14	sistem ekonomi demokrasi ekonomi
9467	kelompok 14	penduduk pembangunan ekonomi
9471	kelompok 14	pengantar ilmu ekonomi iii data ga lengkap
9475	kelompok 14	teori ekonomi mikro
9541	kelompok 14	ekonomi transportasi karakteristik teori kebij...

134 rows x 2 columns

Gbr 9. Tampilan setiap kelompok (Preview cluster 14)

Gbr 10. Hasil Clustering dalam file excel

TABEL III
DETAIL KESELURUHAN HASIL CLUSTERING

No	Cluster	Jumlah
1	Cluster 0	211
2	Cluster 1	2096
3	Cluster 2	94
4	Cluster 3	325
5	Cluster 4	4419
6	Cluster 5	262
7	Cluster 6	291
8	Cluster 7	74
9	Cluster 8	42
10	Cluster 9	211
11	Cluster 10	198
12	Cluster 11	193
13	Cluster 12	131
14	Cluster 13	29
15	Cluster 14	134
16	Cluster 15	32
17	Cluster 16	72
18	Cluster 17	84
19	Cluster 18	200
20	Cluster 19	82
21	Cluster 20	115
22	Cluster 21	161
23	Cluster 22	19

24	Cluster 23	39
25	Cluster 24	51

Berikut ini adalah contoh cluster yang terbentuk dan contoh judul buku yang terdapat dalam satu cluster :

- Cluster 0 dengan jumlah data 211 dengan data { fisika universitas optika fisika modern, penuntun praktikum biologi, kimia medisinal pendekatan biokimia, fisika universitas mekanika panas bunyi}.
- Cluster 1 dengan jumlah data 2096 dengan data { budaya masyarakat perbatasan hubungan sosial antargolongan etnik berbeda daerah sumatra barat, tata kelakuan lingkungan pergaulan keluarga masyarakat makassar}.
- Cluster 2 dengan jumlah data 94 dengan data { materi pokok pengantar matematika modern modul, matematika finansial modul, pengajaran matematika herry sukarman p manulu}.
- Cluster 3 dengan jumlah data 325 dengan data { pendekatan ilmu sosial metodologi sejarah, ilmu sosial dasar modul, pedoman guru ilmu pengetahuan sosial jilid sekolah dasar kelas iv}.
- Cluster 4 dengan jumlah data 4419 dengan data { strength materials textbook engineering students, laboratory manual general ecology, theory criticism plato present reader}.

Berdasarkan analisa dari hasil clustering yang dilakukan, didapatkan hasil sebagai berikut :

- Cluster 0 : Sebagian besar judul yang ada mengarah tentang Ilmu Pengetahuan Alam.
- Cluster 1 : Sebagian besar judul yang ada mengarah tentang budaya masyarakat.
- Cluster 2 : Sebagian besar judul yang ada mengarah tentang Matematika.
- Cluster 3 : Sebagian besar judul yang ada mengarah tentang Ilmu Pengetahuan Sosial.
- Cluster 4 : Sebagian besar judul yang ada mengarah tentang Buku Asing (Berbahasa Inggris).

Setelah melakukan pengelompokan judul buku akan dihitung nilai akurasi, pada perhitungan akurasi ini hanya diambil 100 judul buku, didapatkan hasil sebagai berikut :

TABEL IIIII
PERHITUNAN AKURASI

No	Judul	Hasil Sebenarnya	Hasil Clustering	Ket
1	Statistik teori dan aplikasi jilid 2	Statistik	Statistik	1
2	Pokok-pokok statistik 2 = statistik inferensif	Statistik	Statistik	1

3	Statistik untuk manajemen dan ekonomi jilid 1	Statistik	Ekonomi	0
4
5
98	Pengantar sejarah kebudayaan Indonesia 2	Indonesia - sejarah	sejarah	1
99	Masa akhir Majapahit: Girindrawardhana dan masalahnya	Indonesia-sejarah Majapahit	buku asing	0
100	Keterpurukan hukum di indonesia : penyebab dan solusinya	Hukum	Hukum	1

Perhitungan akurasi dari 100 judul yang dilakukan, judul buku yang sesuai (ditunjukkan dengan ket 1) sebanyak 76 judul. Sedangkan yang tidak sesuai sebanyak 26 judul. Dan nilai akurasi yang didapatkan sebesar 76%.

IV. KESIMPULAN

Dari hasil klasterisasi menggunakan algoritma *Agglomerative Hierarchical Clustering* dan *K-Medoids* ternyata judul buku yang berada dalam satu cluster memiliki kategori yang serupa. Dan dengan menganalisis hasil klasterisasi dapat digunakan sebagai bahan analisis dalam pengambilan keputusan yang bertujuan untuk mempermudah petugas perpustakaan dalam melakukan pengelolaan peletakan buku. Sehingga dapat memudahkan pengunjung perpustakaan untuk mencari buku yang dicari dan buku yang tersedia bisa terbaca dengan lebih optimal.

REFERENSI

- D. T. Velmurugan, "Efficiency of k-Means and K-Medoids Algorithms for Clustering Arbitrary Data Points," *IJCTA*, pp. 1-7, 2012.
- Y. H. D. A. G. Chrisnanto, "Penerapan Algoritma Partitioning Around Medoids(PAM) Clustering untuk Melihat Gambaran Umum Kemampuan Akademik Mahasiswa," *Seminar Nasional Teknologi Informasi dan Komunikasi*, pp. 1-5, 2015.
- A. Y. Rofiqi, "Clustering Berita Olahraga Berbahasa Indonesia Menggunakan Metode K-Medoid Bersyarat," *Jurnal Simantec*, vol. 6, pp. 25-32, 2017.
- N. N. E. Smrti, "Otomatisasi Klasifikasi Buku Perpustakaan dengan Menggabungkan Metode K-NN dengan K-Medoids," *Jurnal Ilmiah Teknologi Informasi*, vol. 4 No. 1, 2013.
- T. Alfina, B. Santosa and A. R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan

- Gabungan Keduanya dalam Cluster Data (Studi Kasus: 3Problem Kerja Praktek Teknik Industri ITS)," *Jurnal Teknik ITS*, vol. 1, 2012.
- [6] E. L. Agustina, M. B. Mitakda and Solimun, "Pemilihan Metode Pengelompokan Terbaik Kabupaten/Kota Berdasarkan Indikator Pendidikan Menggunakan Hybrid Melalui Mutual Cluster, Bottom-up dan Top-down," *Jurnal Mahasiswa Statistik*, vol. 1, pp. 205-208, 2013.
- [7] A. Sukma, B. Zaman and E. Purwanti, "Klasifikasi Dokumen Temu Kembali Informasi dengan K-Nearest Neighbour," *Record and Library Journal*, vol. 1 No. 2, pp. 129-138, 2015.
- [8] A. Bouguettaya, Q. Yu, X. Liu, X. Zhou and A. Song, "Efficient agglomerative hierarchical clustering," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2785-2797, 2015.
- [9] A. Fadliana and F. Rozi, "Penerapan Metode Agglomerative Hierarchical Clustering Untuk Klasifikasi Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana," *Chauchy Journal*, vol. 4 No. 1, pp. 35-40, 2015.
- [10] N. K. K. U. & S. D. D. Kaur, "K-Medoids Clustering Algorithm – A Review.," *International Journal of Computer Application and Technology (IJCAT)*, pp. 1-4, 2014.
- [11] D. D. C. Nugraha, M. Fahmi, Z. Naimah and N. Setiani, "Klasterisasi Judul Buku dengan Menggunakan Metode K-Means," *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, pp. 1-4, 2014.
- [12] D. B. S. Hery Februariyanti, "Hierarchical Agglomerative Clustering Untuk Pengelompokan Skripsi Mahasiswa," *Prosiding SINTAK*, pp. 1-4, 2017.