Analisis Klastering Buku sebagai Evaluasi untuk Peningkatan Minat Baca Perpusatakaan SMAN 1 Grogol

Diah Leni Karputri[1], Wiyli Yustanti[2]

[1,2] Jurusan Teknik Informatika/Sistem Informasi, Universitas Negeri Surabaya
[1]diah.18013@mhs.unesa.ac.id
[2]wiyliyustanti@unesa.ac.id

***Abstrak*— SMAN 1 Grogol merupakan sekolah menengah atas yang berlokasi di Kabupaten Kediri. Berdasarkan hasil wawancara dan observasi yang dilakukan kepada staf perpustakaan, saat ini kondisi minat baca di SMAN 1 Grogol tergolong cukup rendah, hal ini dilihat dari data rekap peminjaman buku di perpustakaan selama lima tahun terakhir yang fluktuatif. Pada perpustakaan penempatan buku juga belum optimal, sehingga perpustakaan SMAN 1 Grogol berkeinginan untuk meningkatkan pelayanannya. Peningkatan pelayanan memberikan kemudahan pada siswa dalam mencari buku sesuai dengan minat baca dan meningkatkan rasa tertarik terhadap buku lain karena setiap buku sudah dikelompokkan sesuai dengan kategori. Manfaat yang diperoleh pihak perpustakaan dapat membantu penentuan prioritas pengadaan buku selanjutnya. Untuk itu diperlukan analisis *clustering*. *Clustering* merupakan suatu metode untuk mengelompokkan data yang memiliki kemiripan karakteristik antara satu data dengan data yang lain. Pada penelitian ini akan menggunakan beberapa algoritma *clustering* yaitu *K-Means clustering*, *K-Medoids clustering*, dan *Fuzzy C-Means clustering*. Hasil percobaan dengan tiga algoritma *clustering* yang berbeda, mendapatkan *cluster* yang optimal berada pada *cluster* 8. Kemudian evaluasi *clustering* digunakan metode *silhouette coefficient*, apabila nilai *silhouette coefficient* mendekati 1 maka hasil *cluster* tersebut dapat dijadikan sebagai rekomendasi. Hasil evaluasi *clustering* mendapatkan *cluster* paling optimal adalah *cluster* 8 dengan nilai *slihouette* 0,323. Sedangkan algoritma dengan kinerja terbaik adalah *K-Medoids clustering* memperoleh nilai 0,21328.**

Kata Kunci— ***Clustering, Silhouette Coefficient, Analisis Clustering***

I. PENDAHULUAN

Perpustakaan merupakan sebuah lembaga yang mengelola koleksi buku, majalah, dan informasi dalam bentuk fisik, adanya perpustakaan memberikan peran penting pada perkembangan literasi di Indonesia. Perpustakaan memiliki peran terhadap tersedianya fasilitas yang memadai dan memberikan pelayanan terbaik dalam peningkatan literasi. Minat baca berasal dari dalam diri seseorang, sehingga apabila menghendaki peningkatan minat membaca butuh pemahaman kolektif [1].

Dari hasil survei pada tahun 2018 yang dilakukan oleh *Programme for International Student Assessment* (PISA), menunjukkan beberapa masalah pendidikan yang terjadi di Indonesia. Dalam kategori kemampuan membaca, sains, dan matematika, skor Indonesia tergolong rendah karena berada di urutan ke-74 dari 79 negara.

SMAN 1 Grogol merupakan sekolah menengah atas yang berlokasi di Kabupaten Kediri. Berdasarkan hasil wawancara dan observasi yang dilakukan kepada staf perpustakaan, saat ini kondisi minat baca di SMAN 1 Grogol tergolong cukup rendah, hal ini dapat dilihat dari data rekap peminjaman buku di perpustakaan selama lima tahun terakhir yang fluktuatif. Pada tahun pelajaran 2017/2018 mengalami penurunan sebanyak 1,3% dari tahun pelajaran 2016/2017, kemudian pada tahun pelajaran 2018/2019 mengalami kenaikan sebanyak 10,5% dari tahun sebelumnya, akan tetapi pada tahun pelajaran 2019/2020 kembali mengalami penurunan sebanyak 13,1% dari tahun sebelumnya dengan rata-rata per bulan 400 peminjam, dan pada tahun pelajaran 2020/2021 belum ada rekap peminjaman dikarenakan masih pada masa pandemi Covid-19. Rendahnya minat baca siswa diindikasikan bahwa kebanyakan siswa akan ke perpustakaan apabila mendapatkan tugas dari guru. Dari hasil *clustering* sementara yang dilakukan pada kategori kemampuan membaca sains dan matematika masih sedikit diminati. Buku yang banyak dipinjam adalah buku pengetahuan umum, bahasa, dan buku novel. Selain itu pembuatan rekap data peminjaman buku pada perpustakaan dilakukan dengan cara manual, hal ini berdampak pada proses rekap data peminjaman buku membutuhkan waktu yang cukup lama. Pada perpustakaan penempatan buku juga belum optimal, sehingga perpustakaan SMAN 1 Grogol berkeinginan untuk meningkatkan pelayanannya. Hal ini bertujuan memudahkan siswa dalam mencari buku sesuai dengan minat baca dan meningkatkan rasa tertarik terhadap buku lain karena setiap buku sudah dikelompokkan sesuai dengan kategori. Manfaat yang diperoleh pihak perpustakaan dapat membantu penentuan prioritas pengadaan buku selanjutnya. Untuk itu diperlukan teknik yang dapat mengelompokkan data.

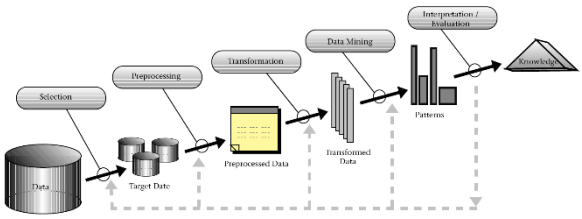
Penelitian tentang faktor-faktor yang mempengaruhi minat baca, evaluasi minat baca, dan pendekatan *data mining* untuk evaluasi minat baca sebelumnya sudah dilakukan. Penelitian sebelumnya diantaranya yaitu Penerapan *Data Mining* Algoritma *Naives Bayes* Dan *Part* Untuk Mengetahui Minat Baca Mahasiswa Di Perpustakaan STMIK Amikom Purwokerto [2], Penerapan *Data Mining Clustering* Dalam Mengelompokan Buku Dengan Metode *K-Means* [3], Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode *K-Means Clustering* [4], penelitian ini dilakukan dengan tujuan untuk mengoptimalisasikan penempatan buku dan penentuan prioritas pengadaan buku selanjutnya. Dalam pengujian ini digunakan data 40 sampel buku dan terdapat 166 pinjaman dalam waktu peminjaman 434 hari, perhitungan dilakukan sebayak 6 kali iterasi sehingga menghasilkan 3 *cluster* yaitu *cluster* 1 terdapat 4 buku yang

sangat diminati, *cluster* 2 terdapat 20 buku yang diminati, dan *cluster* 3 terdapat 16 buku yang kurang diminati.

Berdasarkan permasalahan yang ditemukan di lapangan, didukung penelitian terdahulu terkait dengan penentuan jumlah *cluster* yang memberikan solusi terhadap permasalahan yang terdapat pada organisasi, maka peneliti mengusulkan sebuah penelitian dengan mengambil topik penelitian yang berjudul "Analisis Klastering Buku Sebagai Evaluasi Untuk Peningkatan Minat Baca Perpusatakaan SMAN 1 Grogol". Penelitian ini bertujuan untuk mengetahui hasil *clustering* data peminjaman buku pada tahun pelajaran 2018/2019 - 2019/2020 dan mengetahui hasil terbaik dari metode yang digunakan. Pada percobaan menggunakan *software Rstudio*, atribut data yang digunakan yaitu judul buku dan frekuensi pinjam, percobaan ini menggunakan tiga algoritma, yaitu *K-Means clustering, K-Medoids clustering*, dan *Fuzzy C-Means clustering.* Evaluasi hasil *clustering* menggunakan nilai *Silhoutte Coefficient* sebagai acuan pengelompokan *cluster*, apabila hasil nilai *silhouette coefficient* mendekati 1 maka *cluster* yang terbentuk dikatakan representatif dan bisa digunakan sebagai rekomendasi.

II. METODE PENELITIAN

Dalam penelitian ini menggunakan teori *data mining*, merupakan suatu proses penambangan data, menganalisis dan mengolah menjadi sebuah informasi yang bermanfaat. *Data Mining* merupakan bagian dari proses *Knowledge Discovery In Databases* (KDD) yaitu salah satu metode yang dalam prosesnya bertujuan untuk mengidentifikasi dan menganalisis pola dalam data. Dalam penelitian [5] dijelaskan bahwa proses KDD memiliki beberapa langkah, yaitu *selection, preprocessing, transformation, data mining,* dan *interpretation/evaluation.* Tahapan *Knowledge Discovery in Databases (KDD)* dapat dilihat pada Gbr 1.



Gbr. 1 Tahapan *Knowledge Discovery in Databases* (KDD)

A. Data Selection

Pada tahap awal dimulai dari pengambilan data, penelitian ini menggunakan data peminjaman buku perpustakaan SMAN 1 Grogol pada tahun pelajaran 2018/2019 - 2019/2020. Sebelum dilakukan *clustering*, data yang digunakan akan melalui proses *data preprocessing* dan *data transformation* terlebih dahulu. Karena data yang didapatkan masih berupa data tekstual, maka harus diubah menjadi data yang siap digunakan untuk *clustering*.

B. Data Preprocessing

Preprocessing data merupakan sebuah proses sebelum dilakukannya proses *data mining*. Tujuan *preprocessing* adalah mengubah data mentah menjadi data yang bersih dan siap digunakan. Berikut adalah tahapan dari *preprocessing* data:

a. Case Folding

Tahap pertama yaitu *case folding*, pada tahap ini dilakukan untuk mengubah semua huruf yang ada pada data teks menjadi huruf kecil *(lowercase)*. Kemudian untuk karakter lain yang tidak termasuk huruf dan angka akan dianggap sebagai delimeter.

b. Tokenizing

Tokenizing adalah proses pemecahan kalimat menjadi kata atau frasa. *Tokenizing* juga dilakukan untuk menghapus angka, simbol, dan tanda baca yang tidak penting.

c. Filtering

Filtering merupakan tahapan yang dilakukan untuk membersihkan dan menghilangkan kata yang tidak memiliki makna atau tidak diperlukan dari hasil *tokenizing* sebelumnya.

d. Stemming

Tahap terakhir yaitu tahap *stemming*, pada tahap ini semua kata yang ada dalam dokumen akan diubah menjadi kata dasar. *Stemming* juga menghilangkan kata imbuhan pada kalimat dan kata.

C. Data Transformation

Transformation merupakan tahapan perubahan data, pada proses *data mining* data yang dapat diproses hanya data numerik, sedangkan data peminjaman buku masih berbentuk teks harus diubah menjadi data numerik dengan cara pembobotan kata. TF - IDF atau *Term Frequency - Inverse Document Frequency* merupakan proses pembobotan dalam *text mining*. Proses TF - IDF yaitu menghitung bobot dengan cara mengintegrasikan antara *term frequency (tf)* dan *inverse document frequency (idf)*, tujuan dilakukannya TF - IDF adalah untuk mengetahui seberapa penting satu kata dalam sebuah dokumen. Rumus dalam penggunakan TF adalah sebagai berikut Tokunaga & Iwayama (1994):

$$tf_{ij} = \frac{fd^{(i)}}{maxfd^{(i)}} \quad (1)$$

Dalam TF digambarkan bahwa (d) berarti dokumen dan (t) mewakili seberapa banyak kata yang muncul. Penggunaan rumus IDF dijelaskan sebagai berikut Tokunaga & Iwayama :

$$idf_t = \log \frac{N}{df_t} \quad (2)$$

N dilambangkan sebagai banyaknya kata dalam dokumen, sedangkan df mewakili jumlah dokumen yang mempunyai kata t. TF - IDF yang diintegrasikan secara bersamaan akan menjadi metode yang dapat meingkatkan performa. Rumus TF - IDF adalah sebagai berikut [6]:

$$w_{ij} = tf \times idf \quad (3)$$

$$idf = log\left(\frac{N}{df_t}\right) \quad (4)$$

dengan : i = 1,2,…,p (Jumlah variabel)

j =1,2,…,N (Jumlah data)

D. Data Mining

Tahapan *data mining* merupakan suatu tahapan diterapkannya algoritma proses *clustering*. Pada tahap *data mining* menggunakan 3 algoritma untuk dilakukan *clustering yaitu K-Means clustering, K-Medoids clustering,* dan *Fuzzy C-Means clustering*. Hasil dari pengolahan data setiap algoritma dapat membentuk kelompok *cluster* yang berbeda.

E. Evaluation

Eavulasi *clustering* merupakan tahap akhir dari KDD, proses evaluasi *clustering* menggunakan hasil *cluster* dari tiga algoritma *K-Means clustering, K-Medods Clustering,* dan *Fuzzy C-Means clustering.* Tujuan evaluasi *clustering* yaitu untuk mengetahui kualitas dari sebuah *cluster*. Proses evaluasi *clustering* menggunakan nilai *silhouette coefficient*, apabila hasil nilai *silhouette coefficient* mendekati 1 maka *cluster* yang terbentuk dikatakan representatif dan bisa digunakan sebagai rekomendasi.

III. HASIL DAN PEMBAHASAN

A. Data Selection

Pengambilan data pada penelitian ini dilakukan dengan melakukan permintaan permohonan data sekunder peminjaman buku tahun pelajaran 2018/2019 - 2019/2020 kepada perpustakaan SMAN 1 Grogol. Jumlah data yang digunakan sebanyak 8135 *records* peminjaman dengan 120 sampel buku dan 2 atribut data. Pada struktur awal data peminjaman buku terdapat beberapa struktur yang terdiri dari 6 atribut yaitu nama siswa, kelas, judul buku, frekuensi pinjam, tanggal pinjam, dan tanda tangan. Akan tetapi dalam penelitian ini, membutuhkan atribut judul buku dan frekuensi pinjam yang akan digunakan untuk *clustering*. Sebelum dilakukan proses *clustering* data yang digunakan harus melalui proses *data preprocessing* data *data transformation* terlebih dahulu.

TABEL I
REKAP DATA PEMINJAMAN BUKU

Data Peminjaman Buku Tahun Pelajaran 2018/2019					
No	Nama	Kelas	Judul Buku	Jumlah	Tanggal
1	Kerin	XII	Sastra Inggris	1	29/07/18
2	Berlian	X	Matematika	1	29/07/18
3	Roby	XI	Ekonomi	2	29/07/18
4	Ahmad	X	Last Memories	1	29/07/18
5	Hellena	XII	Rindu	1	29/07/18

B. Data Preprocessing

Data yang telah didapatkan dan melewati proses *data selection* akan dilakukan *preprocessing* terlebih dahulu. *Preprocessing* dilakukan untuk mengubah data mentah menjadi data yang bersih dan siap digunakan. Proses *data preprocessing* pada penelitian ini menggunakan *software Jupyter Notebook*. Tahapan dari *preprocessing* dibagi menjadi 4, yaitu *case folding, tokenizing, filtering*, dan *stemming*.

a. Case Folding

Tahap pertama dimulai dengan mengubah semua huruf yang berada pada dokumen menjadi huruf kecil *(lowercase).* Untuk melakukan *case folding* yaitu dimulai dengan memanggil fungsi *lower()* pada iterasi dokumen. Berikut adalah contoh hasil *preprocessing* tahap *case folding*.

TABEL II
CONTOH HASIL *CASE FOLDING*

No	Sebelum *Case Folding*	Sesudah *Case Folding*
1	Agama Islam	agama islam
2	SBMPTN Soshum 2018 Bahasa dan Sastra	sbmptn soshum 2018 bahasa dan sastra
3	Indonesia Bahasa dan Sastra	indonesia bahasa dan sastra
4	Inggris 2	inggris 2
5	Bahasa dan Sastra Jawa	bahasa dan sastra jawa

b. Tokenizing

Setelah proses *case folding* tahap selanjutnya *tokenizing* merupakan proses pemecahan kalimat menjadi kata atau frasa. *Tokenizing* juga dilakukan untuk menghapus angka, simbol, dan tanda baca yang tidak penting. Pada proses *tokenizing* dapat dilakukan dengan memanggil *library pandas*, *numpy*, *re*, dan *string*. Berikut adalah contoh hasil *preprocessing* tahap *tokenizing.*

TABEL III
CONTOH HASIL *TOKENIZING*

No	Sebelum *Tokenizing*	Sesudah *Tokenizing*
1	agama islam	agama islam
2	sbmptn soshum 2018	sbmptn soshum
3	bahasa dan sastra indonesia	bahasa dan sastra indonesia
4	bahasa dan sastra inggris 2	bahasa dan sastra inggris
5	bahasa dan sastra jawa	bahasa dan sastra jawa

c. Filtering

Pada tahap *filtering* akan dilakukan pembersihan dan menghilangkan kata yang tidak memiliki makna atau tidak diperlukan termasuk kata hubung seperti “dan”, “untuk”, “sebagai”, “tetapi”, dan “bagaikan”. Proses *filtering* dilakukan dengan memanggil *library nltk* dan mengunduh *stopwords* Bahasa Indonesia dari *nltk*. Berikut adalah contoh hasil *preprocessing* tahap *filtering*.

TABEL IV
CONTOH HASIL *FILTERING*

No	Sebelum *Filtering*	Sesudah *Filtering*
1	agama islam	agama islam
2	sbmptn soshum	sbmptn soshum
3	bahasa dan sastra indonesia	bahasa sastra indonesia
4	bahasa dan sastra inggris	bahasa sastra inggris
5	bahasa dan sastra jawa	bahasa sastra jawa

d. Stemming

Tahap terakhir dari *data preprocessing* adalah *stemming*, yaitu semua kata yang ada dalam dokumen akan diubah menjadi kata dasar. Dalam proses *stemming* diperlukan memanggil *library* Sastrawi, kata imbuhan seperti "-pen", "-me", "-kan", atau "-an" akan dihilangkan. Berikut adalah contoh hasil *preprocessing* tahap *stemming.*

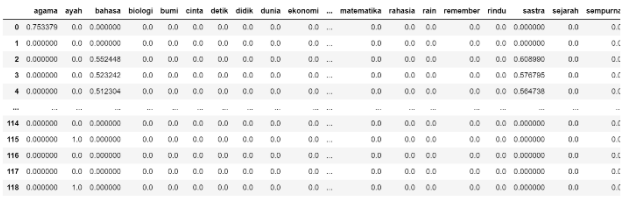
TABEL V
CONTOH HASIL *STEMMING*

No	Sebelum *Stemming*	Sesudah *Stemming*
1	agama islam	agama islam
2	pendidikan agama islam	didik agama islam
3	bahasa dan sastra indonesia	bahasa sastra indonesia
4	rahasia melepaskan	rahasia lepas
5	bahasa dan sastra jawa	bahasa sastra jawa

C. Data Transformation

Pada proses *data transformation* dilakukan dengan menggunakan metode TF-IDF, yaitu menghitung bobot atau tingkat pentingnya suatu kata dalam sebuah dokumen dengan cara mengintegrasikan antara *term frequency (tf)* dan *inverse document frequency (idf)*. Proses perhitungan TF-IDF dibagi menjadi 3 langkah, langkah pertama yaitu dengan menghitung *Term Frequency* (TF), kemudian menghitung *Inverse Document Frequency* (IDF), langkah terakhir menghitung hasil perkalian antara TF dan IDF. Proses *data transformation* dalam penelitian ini menggunakan bantuan dari *software Jupyter Notebook*.

Untuk menghitung *Term Frequency* (TF) digunakan modul *CountVectorizer* dari *library scikit-learn*, sedangkan untuk menghitung *Inverse Document Frequency* (IDF) dilakukan dengan memanggil modul *TfidfTransformer* dari *library scikit-learn*. Kemudian langkah terakhir yaitu menghitung hasil perkalian TF dengan IDF dilakukan dengan memanggil modul *Tfidfvectorizer* dari *library scikit-learn*. Pada objek *Tfidfvectorizer* dilakukan inisialisasi, memasukkan data pada *fungsi fit_transform().* Setelah itu dilanjutkan dengan membuat *DataFrame* yang berfungsi untuk menampilkan matriks hasil perhitungan TF-IDF. Berikut adalah matriks hasil perhitungan TF-IDF.



	agama	ayah	bahasa	biologi	bumi	cinta	detik	didik	dunia	ekonomi	...	matematika	rahasia	rain	remember	rindu	sastra	sejarah	sempurna
0	0.753379	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
1	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
2	0.000000	0.0	0.552448	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.608980	0.0	0.0
3	0.000000	0.0	0.523242	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.576795	0.0	0.0
4	0.000000	0.0	0.512304	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.564738	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
114	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
115	0.000000	1.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
116	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
117	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
118	0.000000	1.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0

Gbr. 2 Hasil Perhitungan TF-IDF

D. Data Mining

Tahap selanjutnya yaitu *clustering*, pada tahap ini data yang sudah dilakukan pembobotan *term* akan digunakan dalam proses *clustering* dengan menggunakan tiga algoritma *clustering* yaitu *K-Means clustering, K-Medoids clustering,* dan *Fuzzy C-Means clustering.* Proses *clustering* akan dilakukan tujuh kali percobaan pada K4, K5, K6, K7, K8, K9, dan K10. Untuk proses *clustering* dibantu aplikasi *Rstudio* dengan memanggil *library factoextra, library tidyverse, library cluster, library ppclust, library dplyr*, dan *library fclust.* Kemudian akan diketahui perubahan pada setiap klaster dengan melihat dari hasil *clustering* setiap algoritma.

TABEL VI
HASIL *CLUSTERING* K4

Cluster	**Algoritma**		
	K-Means	**K-Medoids**	**Fuzzy C-Means**
1	93	92	1
2	9	8	27
3	9	9	75
4	8	10	15

Pada tabel hasil *clustering* K4 dapat dilihat dengan jumlah *cluster* 4, hasil *cluster* yang terbentuk dari setiap algoritma berbeda. Pembentukan *cluster* K4 dengan algoritma *K-Medoids* menghasilkan nilai *silhoutte* tertinggi dibandingkan dua algoritma lain, yaitu dengan nilai 0,178. Hasil *clustering K-Medoids* K4 menunjukkan *cluster* 1 terdapat 92 buku pengetahuan umum, ekonomi, dan novel, *cluster* 2 terdapat 8 buku kimia, *cluster* 3 terdapat 9 buku bahasa dan sastra, dan *cluster* 4 terdapat 10 buku matematika.

TABEL VII
HASIL *CLUSTERING* K5

Cluster	**Algoritma**		
	K-Means	**K-Medoids**	**Fuzzy C-Means**
1	88	69	10
2	8	25	11
3	9	8	80
4	6	9	8
5	8	8	10

Pada tabel hasil *clustering* K5 dapat dilihat dengan jumlah *cluster* 5, hasil *cluster* yang terbentuk dari setiap algoritma berbeda. Pembentukan *cluster* K5 dengan algoritma *K-Medoids* menghasilkan nilai *silhoutte* tertinggi dibandingkan dua algoritma lain, yaitu dengan nilai 0,22. Hasil *clustering K-Medoids* K5 menunjukkan *cluster* 1 terdapat 69 buku pengetahuan umum dan novel, *cluster* 2 terdapat 25 buku

bahasa dan sastra, *cluster* 3 terdapat 8 buku ekonomi, *cluster* 4 terdapat 9 buku matematika, dan *cluster* 5 terdapat 8 buku kimia.

TABEL VIII
HASIL *CLUSTERING* K6

Cluster	Algoritma		
	K-Means	**K-Medoids**	**Fuzzy C-Means**
1	75	60	11
2	8	26	1
3	9	9	10
4	6	8	80
5	8	8	8
6	13	8	9

Pada tabel hasil *clustering* K6 dapat dilihat dengan jumlah *cluster* 6, hasil *cluster* yang terbentuk dari setiap algoritma berbeda. Pembentukan *cluster* K6 dengan algoritma *K-Medoids* menghasilkan nilai *silhoutte* tertinggi dibandingkan dua algoritma lain, yaitu dengan nilai 0,25. Hasil *clustering K-Medoids* K6 menunjukkan *cluster* 1 terdapat 60 buku pengetahuan umum dan novel, *cluster* 2 terdapat 26 buku bahasa dan sastra, *cluster* 3 terdapat 9 buku matematika, *cluster* 4 terdapat 8 buku ekonomi, *cluster* 5 terdapat 8 buku fisika, dan *cluster* 6 terdapat 8 buku kimia.

TABEL IX
HASIL *CLUSTERING* K7

Cluster	Algoritma		
	K-Means	**K-Medoids**	**Fuzzy C-Means**
1	47	53	8
2	8	25	81
3	9	9	9
4	6	8	7
5	8	8	0
6	13	8	13
7	28	7	0

Pada tabel hasil *clustering* K7 dapat dilihat dengan jumlah *cluster* 7, hasil *cluster* yang terbentuk dari setiap algoritma berbeda. Pembentukan *cluster* K7 dengan algoritma *K-Medoids* menghasilkan nilai *silhoutte* tertinggi dibandingkan dua algoritma lain, yaitu dengan nilai 0,288. Hasil *clustering K-Medoids* K7 menunjukkan *cluster* 1 terdapat 53 buku pengetahuan umum dan novel, *cluster* 2 terdapat 25 buku bahasa dan sastra, *cluster* 3 terdapat 9 buku matematika, *cluster* 4 terdapat 8 buku ekonomi, *cluster* 5 terdapat 8 buku fisika, *cluster* 6 terdapat 8 buku kimia, dan *cluster* 7 terdapat 7 buku geografi.

TABEL X
REKAPITULASI HASIL *CLUSTERING* K8

Cluster	Algoritma		
	K-Means	**K-Medoids**	**Fuzzy C-Means**
1	17	59	0
2	8	9	16
3	9	14	0
4	8	8	10
5	7	8	77
6	8	6	8
7	7	8	8
8	52	7	0

Pada tabel rekapitulasi hasil *clustering* K8 dapat dilihat pembentukan *cluster* K8 dengan algoritma *K-Medoids* menghasilkan nilai *silhoutte* tertinggi dibandingkan dua algoritma lain, yaitu dengan nilai 0,303. Hasil *clustering K-Medoids* K8 menunjukkan *cluster* 1 terdapat 59 buku pengetahuan umum dan novel, *cluster* 2 terdapat 9 buku matematika, *cluster* 3 terdapat 14 buku bahasa dan sastra, *cluster* 4 terdapat 8 buku ekonomi, *cluster* 5 terdapat 8 buku kimia, *cluster* 6 terdapat 6 buku biologi, *cluster* 7 terdapat 8 buku fisika, dan *cluster* 8 terdapat 7 buku geografi.

TABEL XI
REKAPITULASI HASIL *CLUSTERING* K9

Cluster	Algoritma		
	K-Means	**K-Medoids**	**Fuzzy C-Means**
1	8	53	0
2	32	11	0
3	10	7	10
4	8	10	0
5	7	8	0
6	6	8	110
7	9	8	0
8	29	7	0
9	8	5	0

Pada tabel rekapitulasi hasil *clustering* K9 dapat dilihat pembentukan *cluster* K9 dengan algoritma *K-Medoids* menghasilkan nilai *silhoutte* tertinggi dibandingkan dua algoritma lain, yaitu dengan nilai 0,34. Hasil *clustering K-Medoids* K9 menunjukkan *cluster* 1 terdapat 53 buku novel, *cluster* 2 terdapat 11 buku bahasa dan sastra, *cluster* 3 terdapat 7 buku biologi, *cluster* 4 terdapat 10 buku matematika, *cluster* 5 terdapat 8 buku ekonomi, *cluster* 6 terdapat 8 buku fisika, *cluster* 7 terdapat 8 buku kimia, *cluster* 8 terdapat 7 buku geografi, dan *cluster* 9 terdapat 5 buku sosiologi.

TABEL XII
REKAPITULASI HASIL *CLUSTERING* K10

Cluster	**Algoritma**		
	K-Means	**K-Medoids**	**Fuzzy C-Means**
1	41	48	0
2	10	11	0
3	8	7	10
4	6	10	0
5	6	8	110
6	4	8	0
7	15	8	0
8	8	7	0
9	5	6	0
10	11	5	0

Pada tabel rekapitulasi hasil *clustering* K10 dapat dilihat pembentukan *cluster* K10 dengan algoritma *K-Medoids* menghasilkan nilai *silhoutte* tertinggi dibandingkan dua algoritma lain, yaitu dengan nilai 0,355. Hasil *clustering K-Medoids* K10 menunjukkan *cluster* 1 terdapat 48 buku novel, *cluster* 2 terdapat 11 buku bahasa dan sastra, *cluster* 3 terdapat 7 buku biologi, *cluster* 4 terdapat 10 buku matematika, *cluster* 5 terdapat 8 buku ekonomi, *cluster* 6 terdapat 8 buku fisika, *cluster* 7 terdapat 8 buku kimia, *cluster* 8 terdapat 7 buku geografi, *cluster* 9 terdapat 5 buku sosiologi, dan *cluster* 10 terdapat 5 buku sejarah. Dari hasil *clustering* tiga algoritma dapat dilihat bahwa setiap algoritma menghasilkan *cluster* yang berbeda-beda. Perbedaan hasil setiap *cluster* dikarenakan rumus yang digunakan pada setiap algoritma *clustering* berbeda. Algoritma *K-Means clustering* dalam menentukan pusat *cluster* menggunakan nilai rata-rata, penentuan pusat *cluster* dari algoritma *K-Medoids clustering* berdasarkan nilai objek yang dipilih dari kumpulan objek (data), sedangkan untuk algoritma *Fuzzy C-Means clustering* dalam menentukan pusat *cluster*nya menggunakan nilai dari keanggotaan *fuzzy*.

E. Evaluation

Setelah proses *clustering* tahap terakhir dari KDD adalah proses evaluasi, pada proses evaluasi digunakan hasil *clustering* dari tiga algoritma yaitu *K-Means clustering, K-Medoids clustering,* dan *Fuzzy C-Means clustering* menggunakan metode *silhouette coefficient*. Evaluasi *clustering* ini bertujuan untuk mengetahui nilai pengelompokkan *cluster* yang paling optimal. Proses evaluasi *clustering* dapat dilakukan secara manual dan menggunakan bantuan aplikasi, akan tetapi pada penelitian ini akan menggunakan aplikasi *Rstudio*. Untuk mengukur pengelompokkan berdasarkan nilai *silhouette coefficient* terdapat beberapa kriteria subjektif sebagai berikut.

TABEL XIII
KETERANGAN KRITERIA SUBJEKTIF *SILHOUETTE COEFFICIENT*

Rentang Nilai Silhouette Coefficient	**Keterangan Silhouette Coefficient**
≤ 0,25	Klaster yang buruk
0,26 – 0,50	Klaster yang lemah
0,51 – 0,70	Klaster telah layak atau sesuai
0,71 – 1,00	Klaster kuat

Evaluasi *clustering* dilakukan dengan menggunakan data hasil *clustering* yang didapatkan dari tahap sebelumnya, pengujian ini dilakukan pada *cluster* 2 sampai *cluster* 10. Adapun hasil evaluasi *clustering* menggunakan metode *silhouette coefficient* dari tujuh kali percobaan *clustering*.

TABEL XIV
REKAPITULASI HASIL EVALUASI *CLUSTERING*

Cluster	**Algoritma**		
	K-Means	**K-Medoids**	**Fuzzy C-Means**
2	0,077	0,089	0,113
3	0,116	0,12	0,024
4	0,171	0,178	0,174
5	0,192	0,23	0,046
6	0,208	0,26	0,138
7	0,255	0,288	0,04
8	0,28	0,303	0,26
9	0,292	0,34	0,248
10	0,258	0,355	0,323

Pada tabel rekapitulasi hasil evaluasi *clustering* setelah dilakukan perhitungan sampai iterasi ke 10, dapat dilihat nilai *silhouette coefficient* sangat bervariasi dari tiga algoritma *clustering* yang digunakan. Besar nilai *silhouette coefficient* berada pada *range* -1 sampai 1 apabila hasil nilai *silhouette coefficient* mendekati 1 maka *cluster* yang terbentuk dikatakan representatif dan bisa digunakan sebagai rekomendasi. Sebaliknya jika mendekati angka -1 maka pengelompokan jumlah *cluster* semakin memburuk. Dari hasil evaluasi *clustering*, *cluster* terbaik algoritma *K-Means clustering* memperoleh nilai indeks 0,292 terletak pada *cluster* 9. Untuk algoritma *K-Medoids clustering*, *cluster* terbaik berada pada *cluster* 10 dengan nilai indeks sebesar 0,355. Kemudian untuk algoritma *Fuzzy C-Means clustering*, *cluster* terbaik diperoleh nilai indeks 0,323 dengan menempati *cluster* 10 sama seperti algoritma sebelumnya yaitu *K-Medoids*. Berdasarkan adanya perbedaan nilai indeks yang bervariasi pada beberapa algoritma, diperlukan perhitungan dalam menentukan algoritma terbaik. Hal ini bertujuan untuk mengetahui performa algoritma mana yang optimal dalam *clustering* data peminjaman buku. Penentuan algoritma terbaik dilakukan dengan menghitung rata-rata nilai *silhouette coefficient* setiap algoritma. Adapun rata-rata hasil evaluasi *clustering* menggunakan nilai *silhouette coefficient.*

TABEL XV
HASIL RATA-RATA *SILHOUETTE COEFFICIENT*

No	Algoritma	Rata-rata
1	K-Means	0,20544
2	K-Medoids	0,24033
3	Fuzzy C-Means	0,15365

Pada tabel rata-rata indeks *silhouette coefficient*, indeks rata-rata tertinggi berada pada algoritma *K-Medoids clustering*. Hal ini menunjukkan bahwa algoritma *K-Medoids clustering* memiliki performa lebih baik dibandingkan dengan algoritma *K-Means clustering* dan *Fuzzy C-Means clustering*. Berdasarkan indeks rata-rata *silhouette coefficient* dari *K-Medoids clustering* yaitu 0,24033, jumlah cluster yang terbentuk lebih representatif karena nilai indeks yang dihasilkan mendekati 1. Sehingga dalam penentuan algoritma terbaik untuk *clustering* data peminjaman buku adalah algoritma *K-Medoids clustering.* Kemudian berdasarkan keterangan nilai *silhouette coefficient K-Medoids clustering* dikatakan sebagai keterangan *slihouette coefficient cluster* yang buruk dengan nilai rata-rata *silhouette coefficient* ≤ 0,25.

F. Analisis Clustering Buku

Berdasarkan hasil *clustering* dan evaluasi *clustering* yang telah dilakukan pada tahap sebelumnya, algoritma K-Medoids menghasilkan jumlah *cluster* yang lebih representatif dan *cluster* terbaik berada pada *cluster* 8. Akan tetapi pada penelitian ini mengambil pengelompokkan *cluster* 4, karena untuk *cluster* 4 sudah bisa diinterpretasikan. Untuk mempermudah analisa hasil *clustering* akan dibuat *wordcloud* dari setiap *cluster*, proses pembuatan *wordcloud* menggunakan aplikasi *Rapidminer*. Adapun hasil *wordcloud* dari setiap *cluster* yang digunakan untuk analisa hasil *clustering* data peminjaman buku.

Gbr. 3 Hasil *Wordcloud Cluster* 1

Pada hasil *wordcloud cluster* 1 terdapat beberapa judul buku yang paling sering muncul, yaitu buku “ekonomi”, “geografi”, “klise”, “rain flowers”, “sejarah”, “biologi”, “fisika”, dan “sosiologi”. Dari judul buku tersebut dan melihat nilai rata-rata *descriptive statistic cluster* sebesar 69,19565, dapat disimpulkan bahwa pada *cluster* 1 terdapat 92 buku yang sering dipinjam dengan anggotanya adalah buku pengetahuan sosial, ekonomi, novel, biologi, fisika, dan sejarah.

Gbr. 4 Hasil *Wordcloud Cluster* 2

Pada hasil *wordcloud cluster* 2 terdapat beberapa judul buku yang paling sering muncul, yaitu buku “kimia”, “detik kimia”, “kimia umum”, dan “olimpiade kimia”. Dari judul buku tersebut dan melihat nilai rata-rata *descriptive statistic cluster* sebesar 35,25, dapat disimpulkan bahwa pada *cluster* 2 terdapat 8 buku yang cukup dipinjam dengan anggotanya adalah buku kimia.

Gbr. 5 Hasil *Wordcloud Cluster* 3

Pada hasil *wordcloud cluster* 3 terdapat beberapa judul buku yang paling sering muncul, yaitu buku “bahasa jawa”, “bahasa inggris”, “bahasa dan sastra inggris”, dan “bahasa dan sastra jawa”. Dari judul buku tersebut dan melihat nilai rata-rata *descriptive statistic cluster* sebesar 128,4444444, dapat disimpulkan bahwa pada *cluster* 3 terdapat 9 buku yang sangat sering dipinjam dengan anggotanya adalah buku bahasa dan sastra.

Gbr. 6 Hasil *Wordcloud Cluster* 4

Pada hasil *wordcloud cluster* 4 terdapat beberapa judul buku yang paling sering muncul, yaitu buku “matematika”, “un matematika”, “matematika umum”, dan “detik matematika”.

Dari judul buku tersebut dan melihat nilai rata-rata *descriptive statistic cluster* sebesar 33,55555556, dapat disimpulkan bahwa pada *cluster* 4 terdapat 9 buku yang jarang dipinjam dengan anggotanya adalah buku matematika.

Dari hasil *wordcloud* setiap *cluster* dapat diambil kesimpulan, pada saat ini minat baca siswa SMAN 1 Grogol untuk kemampuan membaca sains dan matematika masih sedikit diminati. Buku yang populer dipinjam berada pada *cluster* 1 dan 3 yaitu buku pengetahuan sosial, ekonomi, bahasa, dan novel. Sedangkan buku yang jarang dipinjam berada pada *cluster* 2 dan 4 yaitu buku kimia dan matematika. Dengan adanya hasil *clustering* diharapkan dapat membantu pihak perpustakaan untuk mengoptimalisasikan penempatan buku, sehingga memudahkan siswa untuk lebih cepat mencari buku sesuai dengan minat bacanya. Hal ini juga akan meningkatkan keinginan siswa untuk berkunjung ke perpustakaan karena buku sudah dikelompokkan dengan optimal, serta dapat membantu penentuan prioritas untuk pengadaan buku selanjutnya.

IV. KESIMPULAN

Dari penelitian yang telah dilakukan maka dapat diambil beberapa kesimpulan. Adapun kesimpulan yang didapatkan sebagai berikut:

1. Berdasarkan rata-rata indeks *silhouette coefficient* dari hasil *clustering* tiga algoritma yang digunakan. Algoritma pengelompokkan yang memiliki kinerja lebih baik dibandingkan dengan algoritma *K-Means clustering* dan *Fuzzy C-Means clustering* adalah algoritma *K-Medoids clustering* dengan nilai indeks sebesar 0,24033.
2. Dilihat dari indeks *silhouette coefficient*, jumlah *cluster* terbaik dari algoritma *K-Means clustering* berada pada *cluster* 9 memiliki nilai indeks sebesar 0,292, algoritma *K-Medoids clustering* memiliki nilai indeks sebesar 0,355 terletak pada *cluster* 10, dan algoritma *Fuzzy C-Means* memiliki nilai indeks sebesar 0,323 terletak pada *cluster* 10.
3. Berdasarkan hasil *wordcloud* setiap *cluste*r pada saat ini minat baca siswa SMAN 1 Grogol untuk kemampuan membaca sains dan matematika masih sedikit diminati. Buku yang populer dipinjam berada pada *cluster* 1 dan 3 yaitu buku pengetahuan sosial, ekonomi, bahasa, dan novel. Sedangkan buku yang jarang dipinjam berada pada *cluster* 2 dan 4 yaitu buku kimia dan matematika. Dengan adanya hasil *clustering* diharapkan dapat membantu pihak perpustakaan untuk mengoptimalisasikan penempatan buku, sehingga memudahkan siswa untuk lebih cepat mencari buku sesuai dengan minat bacanya. Hal ini juga akan meningkatkan keinginan siswa untuk berkunjung ke perpustakaan karena buku sudah dikelompokkan dengan optimal, serta dapat membantu penentuan prioritas untuk pengadaan buku selanjutnya.
4. Percobaan *clustering* dengan menggunakan tiga algoritma menghasilkan *cluster* dengan rata-rata interpretasi indeks *silhouette coefficient* yang lemah, hal ini dikarenakan terlalu banyak variasi buku yang digunakan pada proses *clustering*.

V. SARAN

Saran untuk penelitian yang akan dilakukan dengan melihat pada penelitian yang telah dilakukan adalah pada proses *clustering* dapat digunakan metode *clustering* lain seperti *WaveCluster, CLIQUE, K-Medians* atau metode lainnya. Sedangkan dalam proses evaluasi *clustering* dapat dilakukan dengan metode pengujian lain seperti metode *Davies-Bouldin Index* dan *Purity Score.*

REFERENSI

[1] Puji Hadayani, H. D. K., "Pengembangan Media Pembelajaran Buku Cerita Bergambar Untuk Meningkatkan Minat Membaca Siswa Sekolah Dasar", Jurnal Basicedu, vol. 4, hal 994-1003, April 2020.

[2] Imron, M., "Penerapan *Data Mining* Algoritma *Naives Bayes* dan *Part* Untuk Mengetahui Minat Baca Mahasiswa Di Perpustakaan STMIK Amikom Purwokerto", Jurnal Telematika, vol. 10, hal 121–135, Agustus 2017.

[3] Yulia, M. S., "Penerapan *Data Mining Clustering* Dalam Mengelompokan Buku Dengan Metode *K-Means*", *Indonesian Journal of Computer Science*, vol. 10, hal 62, April 2021.

[4] Fakhri, D. A., Defit, S., & Sumijan., "Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode *K-Means Clustering*", Jurnal Informasi Dan Teknologi, vol. 3, hal 160-166, Sep. 2021.

[5] Usama Fayyad, Gregory Piatetsky-Shapiro, P. S., "*Knowledge Discovery and Data Mining: Towards a Unifying Framework", Journal of Japan Society for Fuzzy Theory and Systems*, vol. 9, hal 851-860, 1996.

[6] Asiyah, S. N., & Fithriasari, K., "Klasifikasi Berita Online Menggunakan Metode *Support Vector Machine* dan *K-Nearest Neighbor*", Jurnal Sains Dan Seni ITS, vol. 5, hal 317-322, 2016.