Jurnal Teknologi Informasi dan Komunikasi

Vol: 13 No 1 2022 **E-ISSN: 2477-3255**

Diterima Redaksi: 05-02-2022 | Revisi: 23-04-2022 | Diterbitkan: 27-05-2022

Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis

Siti Ramadhani[1], Dini Azzahra,[2] Tomi Z[3]

[1,2]Program Studi Teknik Informatika Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim

[3]Pusat Teknologi dan Pangkalan Data

[1,2,3]Jl. H.R. Soebrantas no. 155 KM. 18 Simpang Baru, Pekanbaru 28293, Telp (0761) 562223

e-mail: [1]siti.ramadhani@uin-suska.ac.id, [2]diniazzahra038@gmail.com, [3]tomiz@ uin-suska.ac.id

Abstract

The thesis is one of the scientific works based on the conclusions of field research or observations compiled and developed by students as well as research carried out according to the topic containing the study program which is carried out as a final project compiled in the last stage of formal study. A large number of theses, of course, will be sought in looking for categories of thesis topics, or the titles raised have different relevance. However, the student thesis can be by topics that are almost relevant to other topics so that it can make it easier to find topics that are relevant to the group. One of the uses of techniques in machine learning is to find text processing (Text Mining). In-text mining, there is a method that can be used, namely the Clustering method. Clustering processing techniques can group objects into the number of clusters formed. In addition, there are several methods used in clustering processing. This study aims to compare 2 cluster algorithms, namely the K-Means and K-Medoids algorithms to obtain an appropriate evaluation in the case of thesis grouping so that the relevant topics in the formed groups have better accuracy. The evaluation stage used is the Davies Bouldin Index (DBI) evaluation which is one of the testing techniques on the cluster. In addition, another indicator for comparison is the computation time of the two algorithms. According to the DBI value test carried out on algorithm 2, the K-Medoids algorithm is superior to K-Means, where the average DBI value produced by K-Medoids is 1,56 while K-Means is 2,79. In addition, the computational time required in classifying documents is also a reference. In testing the computational time required to group 50 documents, K-Means is superior to K-Medoids. K-Means has an average computation time for grouping documents, which is 1 second, while K-Medoids provide a computation time of 26,7778 seconds.

Keywords: *Davies Bouldin Index (DBI), K-Means, K-Medoids, Thesis, Computed Time*

Perbandingan Algoritma K-Means Dan K-Medoids Pada Text Mining untuk Klasifikasi Tesis Mahasiswa Berdasarkan Pengujian Davies Bouldin Index

https://doi.org/10.31849/digitalzone.v13i1.9292

Abstrak

Skripsi ialah salah satu karya ilmiah yang berdasar pada kesimpulan penelitian lapangan atau observasi yang disusun serta dikembangkan oleh mahasiswa sebagai peneliti, serta penelitian dilakukan sesuai dengan topik yang mengandung program studi yang ditempuh sebagai tugas akhir yang disusun pada studi formal tahap terakhir. Banyaknya jumlah skripsi tentunya akan menyulitkan dalam mencari kategori pada topik skripsi, melihat dari topik atau judul yang diangkat memiliki tingkat kemiripan yang berbeda-beda. Namun skripsi mahasiswa dapat dikelompokkan sesuai dengan topik yang hampir relevan dengan topik skripsi lainnya sehingga dapat mempermudah dalam pencarian topik yang relevan sesuai dengan kelompoknya. Salah satu pemanfaat teknik di dalam machine learning ialah pengolahan teks (Text Mining). Didalam text mining, terdapat metode yang dapat digunakan yaitu metode Clustering. Teknik pengolahan clustering mampu mengelompokkan object kedalam jumlah cluster yang dibentuk. Selain itu, ada beberapa metode yang digunakan didalam pada pengolahan clustering. Penelitian ini bertujuan untuk melakukan melakukan perbandingan 2 algoritma cluster yaitu algoritma K-Means dan K-Medoids untuk mendapatkan evaluasi yang tepat pada kasus pengelompokkan skripsi sehingga topik yang relevan pada kelompok yang dibentuk memiliki keakuratan yang lebih baik. Tahapan evaluasi yang digunakan ialah evaluasi Davies Bouldin Index (DBI) yang merupakan salah satu Teknik pengujian pada cluster. Selain itu, indicator lain yang menjadi perbandingan ialah waktu komputasi kedua algoritma. Dilihat dari pengujian nilai DBI yang dilakukan pada 2 algoritma, algoritma K-Medoids lebih unggul terhadap K-Means, dimana hasil rata-rata nilai DBI yang dihasilkan oleh K-Medoids yaitu 1,56 sedangkan K-Means yaitu 2,79. Selain itu, waktu komputasi yang dibutuhkan dalam mengelompokkan dokumen juga menjadi acuan. Dalam pengujian waktu komputasi yang dibutuhkan dalam mengelompokkan 50 dokumen skripsi, K-Means unggul terhadap K-Medoids. Dimana K-Means memiliki rata-rata waktu komputasi untuk mengelompokkan dokumen ialah sebesar 1 detik sementara pada K-Medoids memberikan waktu komputasi rata-rata sebesar 26,7778 detik.

Kata kunci: *Davies Bouldin Index (DBI), K-Means, K-Medoids*, Skripsi, Waktu Komputasi

1. Pendahuluan

Perkembangan dunia dalam bidang pembaruan teknologi berkembang sangat pesat. Teknologi dan digital adalah satu kesatuan yang berjalan beriringan. Media digital yang mengalami perkembangan dan sedang digunakan tentunya memberikan pengaruh yang baik didalam kehidupan, tidak terkecuali pemanfaatan media digital didalam bidang Pendidikan. Teknologi yang digunakan pada dunia Pendidikan termasuk fasilitas yang tak terpisahkan[1]. Perkembangan media digital berbasis teknologi di bidang pendidikan dapat memberikan sisi baik dilihat dari kemudahan ditemukannya karya ilmiah yang dihasilkan oleh penelitian di bidang teknologi pendidikan. Saat ini, karya ilmiah dibidang Pendidikan banyak ditemukan, seperti Skripsi. Umumnya skripsi adalah salah satu karya ilmiah yang disusun oleh mahasiswa tingkat sarjana. Pengertian lain, Skripsi ditulis oleh peneliti sebagai kajian akhir dalam penelitian formal, dilakukan berdasar hasil penelitian yang ditemukan dilapangan dan dilakukan tahapan referensi dan studi kepustakaan. Hal itu tentunya sesuai bidang topik dan kajian peneliti. Menurut Munslich Mansur, skripsi menunjukkan kompetensi sebagai mahasiswa dalam bidang studi formal dalam penelitian yang berkaitan dengan bidang keahliannya. Selain itu, skripsi memiliki arti lain dimana mahasiswa mempresentasikan karya ilmiah pada topik atau bidang tertentu berdasarkan hasil tinjauan pustaka yang ditulis oleh seorang ahli, temuan penelitian atau perkembangan di bidang tersebut (melalui pengalaman peneliti).

Banyaknya jumlah karya ilmiah tentunya akan menyulitkan dalam mencari kategori untuk karya ilmiah yang akan disusun. Melihat dari topik atau judul yang diangkat memiliki tingkat kemiripan yang berbeda-beda, maka hal itu juga dapat menjadi kendala dalam mengklasifikasikan judul atau topik yang sesuai antara skripsi satu dengan skripsi lainnya[1].

Penerapan data mining dan text mining pada skripsi dapat mempermudah dalam melakukan pengelompokkan kategori topiknya. Didalam text mining, terdapat metode yang dapat digunakan yaitu metode Clustering[2]. Terdapat beberapa metode clustering (Partition-based Clustering Hierarki) yang biasa digunakan dalam penelitian, baik didalam data teks maupun diluar data teks. Adapun beberapa metode clustering yang biasanya digunakan ialah K-Medians, K-Medoids, Fuzzy K-Means, K-Means dan lainnya. Namun dalam penelitian ini, skripsi dikelompokkan menggunakan 2 Algoritma yaitu K-Means dan K-Medoids yang selanjutnya akan dibandingan keakuratan clsuternya pada 2 algoritma dalam penerapannya dalam text mining.

Algoritma K-Means clustering ialah metode pengelompokkan data yang lebih sederhana dalam mengimplementasikannya atau mudah untuk diimplementasikan. Alur proses dari algoritma ini yaitu mempartisi atau membagi-bagi sejumlah data kedalam K cluster melalui rata-rata (Mean) jarak terdekat seluruh data ke data cluster yang selanjutnya akan dihitung rata-ratanya Kembali untuk titik data pada iterasi selanjutnya. Keunggulan dari algoritma K-Means clustering yaitu memiliki waktu komputasi yang relatif cepat[3]. Sedangkan pengertian K-Medoids sama halnya dengan K-Means, namun K-Medoids bekerja sebagai algoritma clustering yang mempartisi sejumlah data kedalam K cluster melalui data medoids yang ditemukannya sebagai titik pusak pada iterasi selanjutnya [4]. Kedua algoritma ini menggunakan Eucludean Distance dalam mencari jarak rata-rata terkecil dalam pengelompokkan datanya. Algoritma K-Medoids bekerja dengan mentukan titik data sebagai data yang representatif untuk meminimalkan jauhnya range data antar data yang lainnya. Sedangkan K-Means hanya menghitung rata-rata jarak seluruh data pada setiap cluster. Hal itu mendorong K-Medoids dinilai lebih baik dibandingkan dengan Algoritma K-Means.

Berbagai penelitian terkait mengenai pengelompokkan dokumen skripsi telah banyak dilakukan sebelumnya. Penelitian yang membandingkan 2 algoritma clustering ini salah satunya dilakukan oleh Elly Muningsih yang diterapkan dalam data mining, disimpulkan bahwa algoritma K-Means dan K-Medoids dapat dilakukan perbandingan dengan kesimpulan K-Means memperoleh nilai Silhouette Coefficient yaitu sekitar 0,627 sedangkan K-Medoids sekitar 0,536. Dilihat dari nilai pengujian Silhouette Coefficient yang dilakukan pada penelitian terkait, maka dapat disimpulkan bahwa keakuratan cluster yang dihasilkan oleh K-Means lebih akurat dan berkualitas daripada K-Medoids[5]. Terdapat penelitian terkait lainnya yang dilakukan oleh Muhammad Sholeh Hudin dkk dalam penerapan metode cluster dengan teks mining, dapat disimpulkan bahwa dari hasil analisis dengan memasukkan jumlah K cluster yang berbeda telah didapatkan nilai optimal dari beberapa percobaan Jumlah K serta implementasi metode clustering dengan text mining dapat dilakukan.

Selain itu, pada penelitian lainnya yang dilakukan oleh Rima Dias dkk, dapat disimpulkan bahwa dengan menggunakan dataset berukuran berskala kecil hingga berskala medium memberikan kesimpulan bahwa algoritma K-Means memiliki performa yang relative cepat dalam pengelompokkan data, sedangkan algoritma K-Medoids memiliki waktu komputasi yang cukup lama dalam pengelompokkan data dalam K cluster[2].

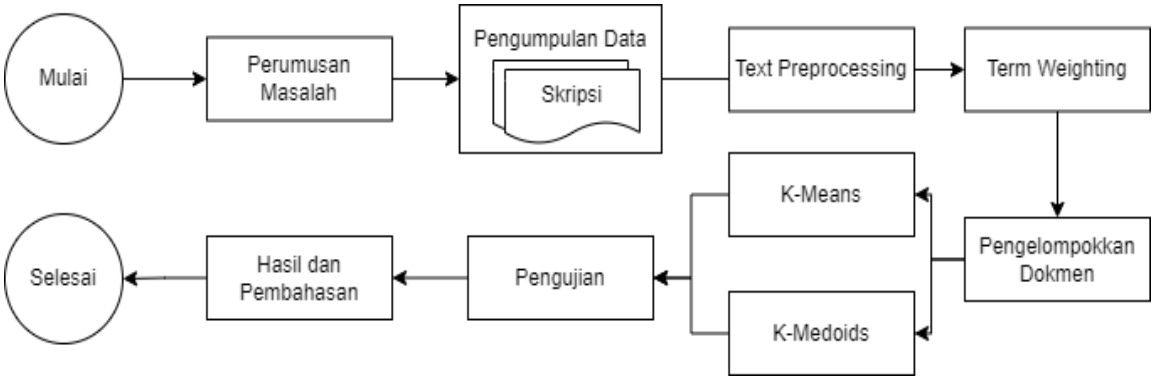
Berbeda dengan penelitian sebelumnya, maka pada penelitian ini membahas perbandingan algoritma K-Means Clustering dan algoritma K-Medoids pada text mining untuk mengklasifikasikan skripsi mahasiswa Fakultas Ekonomi Sosial dalam beberapa kelompok yang diujikan. Selanjutnya dilakukan pengujian antara 2 algoritma tersebut dengan menggunakan pengujian menghitung nilai DBI (*Davies Bouldin Index*).

2. Metode Penelitian

Perancangan alur pada tahapan ini bertujuan untuk memudahkan proses melakukan agar menghasilkan penelitian yang terarah dan terstruktur[2]. Tahapan ini merupakan pedoman utama yang disusun tahap pertahap dari awal penelitian akan dirancang sampai dengan penelitian telah selesai dilakukan[3]. Pada penelitian ini, tahapan awal penelitian dimulai dengan symbol mulai, tahap perumusan masalah, pengumpulan data, text preprocessing,

https://doi.org/10.31849/digitalzone.v13i1.9292

pengelompokkan dokumen, tahapan pengujian, tahapan hasil dan pembahasan serta diakhiri dengan symbol selesai [6]–[12]. Dibawah ini merupakan tahapan-tahapan yang menjadi pedoman utama pada peneltian ini:



Gambar 1. Metodologi Penelitian

Adapun penjabaran setiap tahap pada gambar 1 diatas yaitu:

1. Mulai
   Pada tahapan ini, dilakukan persiapan sebelum melakukan penelitian. Tahapan ini dimulai dari mengangkat topik penelitian yang akan diambil, dataset yang akan digunakan serta tahapan-tahapan yang akan dilalui selama proses penelitian.
2. Perumusan Masalah
   Pada tahapan perumusan masalah, rumusan masalah ditemukan sebagai ide atau topik penelitian yang akan diangkat. Pada rumusan masalah ini, data yang akan digunakan telah ditemukan sumber referensinya serta metode yang akan digunakan telah ditentukan.
3. Pengumpulan Data
   Pada proses pengumpulan data, data yang dikumpulkan berupa 50 dokumen skripsi yang diunduh dari website repository Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi negara. Adapun bagian teks yang diambil dan dipilih yaitu bagian latar belakang pada Bab 1 serta ekstensi file yang diunduh berupa PDF, selanjutnya file latar belakang berekstensi pdf tersebut diubah menjadi ekstensi txt untuk memudahkan proses selanjutnya.
4. Text Preprocessing
   Selanjutnya ialah tahapan Text Preprocessing. Tahapan ini merupakan salah satu komponen utama yang penting didalam text mining. Tahapan ini bertujuan untuk mengubah data sebelumnya berbentuk teks yang tidak memiliki struktur ke dalam data teks yang lebih terstruktur sehingga dapat digunakan pada tahapan selanjutnya. Terdapat beberapa tahapan dalan text preprocessing, yaitu file skripsi dilakukan proses case folding, tokenisasi, stopword removal serta stemming. Berikut ialah penjelasan mengenai setiap tahap text preprocessing pada tahapan ini, yaitu:
   a. Case Folding
      Proses case folding ialah proses yang dilakukan untuk menyeragamkan seluruh karatkter pada teks. Adapun contoh teks yang dilakukan tahapan case folding yaitu:

Teks Sebelum	Teks Sesudah
ANALISIS PENYELENGGARAAN INOVASI PELAYANAN PUBLIK PADA MAL PELAYANAN PUBLIK KOTA PEKANBARU"	analisis penyelenggaraan inovasi pelayanan publik pada mal pelayanan publik kota pekanbaru

Gambar 2. Contoh mat Tahapan Case Folding

b. Tokenisasi
Proses tokenisasi ialah proses dimana dilakukan pemisahan kata dalam mat. Adapun contoh teks yang dilakukan tahapan tokenisasi yaitu:

Teks Sebelum	Teks Sesudah
analisis penyelenggaraan inovasi pelayanan publik pada mal pelayanan publik kota pekanbaru	analisis penyelenggaraan inovasi pelayanan publik pada mal pelayanan publik kota pekanbaru

Gambar 3 Contoh mat Tahapan Tokenisasi

c. Stopword Removal
Proses stopword removal ialah proses dimana dilakukan pengahapusan kata dalam mat yang dianggap tidak mewakili isi dokumen. Adapun contoh teks yang dilakukan tahapan stopword removal yaitu:

Teks Sebelum	Teks Sesudah
analisis, penyelenggaraan, inovasi, pelayanan, publik, pada, mal, pelayanan publik, kota, pekanbaru	analisis, penyelenggaraan, inovasi, pelayanan, publik, mal, pelayanan publik, kota

Gambar 4 Contoh mat Tahapan *Stopword Removal*

d. Stemming
Proses *stemming* ialah proses dimana dilakukan perubahan kata menjadi kata dasar. Adapun contoh teks yang dilakukan tahapan stemming yaitu:

Teks Sebelum	Teks Sesudah
analisis, penyelenggaraan, inovasi, pelayanan, publik, mal, pelayanan publik, kota	analisis, selenggara, inovasi, layan, publik, mal, layan publik, kota

Gambar 5 Contoh mat Tahapan *Stemming*

5. *Term Weighting*
Selanjutnya, dilakukan pembobotan tiap kata. Pemberian bobot atau nilai pada tiap kata dilakukan melalui pembobotan kata. Metode umum pembobotan yang biasanya digunakan ialah Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini yaitu tahapan pemberian bobot atau nilai pada set term dalam teks pada dokumen. TF-IDF ini dapat memberikan pentingnya kata yang terdapat didalam tiap-tiap teks dokumen yang diukur dari nilai yang diberikan dalam bentuk angka dari 0-1. Pada proses pembobotan kata, terdapat beberapa langkah yang dilakukan yaitu mencari *term frequency, inverse document frequency* serta *term frequency – inverse document frequency*.
6. Pengelompokkan Dokumen
Pada tahapan pengelompokkan dokumen, terdapat 2 algoritma dilakukan sebagai perbandingan pada penelitian ini yaitu algoritma K-Means dan K-Medoids. Dimana

https://doi.org/10.31849/digitalzone.v13i1.9292

pengelompokkan file skripsi pada penelitian ini menggunakan Tools Rapidminer serta menggunakan code program python yang diakses di Google Colab.

7. Pengujian
   Pengujian menggunakan *Davies bouldin index* (DBI) merupakan salah satu pengujian yang dapat membantu dalam pengevaluasian keakuratan cluster pada penelitian ini. Pengujian DBI dilakukan dengan mengukur kekuatan atau keaakuratan Jumlah K (cluster) yang dibentuk pada algoritma K-Means dan K-Medoids[4]. Adapun tujuan tahapan pendekatan pengujian DBI yaitu untuk memaksimalkan jarak antara satu cluster dengan cluster lainnya serta mencari nilai untuk meminimalkan jarak antar data dokumen yang terdapat didalam cluster yang sama. Serta selain itu, dalam tahap pengujian ini juga dilakukan pengujian waktu komputasi yang dibutuhkan dalam mengelompokkan 50 file dokumen skripsi pada 2 algoritma yang akan dibandingkan.
8. Hasil dan Pembahasan
   Tahapan hasil dan pembahasan yaitu pemaparan hasil pada cluster dan pengujian yang dilakukan dengan menggunakan pengujian DBI, pengujian performance kecepatan waktu komputasi yang dibutuhkan dalam mengelompokkan dokumen skripsi serta gambaran keanggotan dokumen terhadap kedua algoritma pada cluster yang terbaik menurut pengujian nilai DBI.
9. Selesai
   Akhir tahapan penelitian ini yaitu selesai. Pada tahapan ini, akan dipaparkan penutup pada penelitian ini yang berupa point kesimpulan serta saran dan masukan yang terdapat pada penelitian ini.

4. Hasil dan Pembahasan

Hasil dari penelitian yang dilakukan ini yaitu berupa analisis perbandingan pada algoritma K-Means dan K-Medoids. Adapun analisis dilakukan melalui eveluasi keakuratan cluster melalui perhitungan nilai DBI pada kedua algoritma pada jumlah K yang dibentuk pada 50 dokumen skripsi bagian latar belakang. Jumlah K yang dibentuk ialah 10, dirujuk pada penelitian serupa[5]. Adapun nilai pengujian DBI yang didapatkan yaitu pada tabel berikut ini:

Tabel 1. Perbandingan dari Pengujian Nilai DBI pada Algoritma K-Means dan K-Medoids

Jumlah K	K-Means	K-Medoids
2	4,27	1,79
3	3,22	1,72
4	2,90	1,66
5	2,90	1,66
6	2,59	1,54
7	2,60	1,51
8	2,36	1,43
9	2,23	1,42
10	2,08	1,34

Dalam pengujian davies bouldin index (DBI), apabila nilai yang diberikan mendekati 0 maka pengujian dbi atau non-negatif >= 0, maka hasil cluster yang diberikan semakin baik. Dapat dilihat dari Tabel 1., menurut pengujian DBI yang dilakukan terhadap kedua algoritma, maka dapat disimpulkan pada penelitian ini bahwa nilai DBI yang didapatkan dari jumlah K pada algoritma K-Medoids lebih baik dibandingnkan jumlah K pada algoritma K-Means. Dibuktikan dengan nilai DBI yang dihasilkan oleh algoritma K-Medoids pada tiap tiap cluster yaitu mendekati 0. Sementara nilai DBI yang dihasilkan oleh algoritma K-Means pada tiap tiap cluster

Selanjutnya, mengukur hasil performance kecepatan pengelompokkan yang dilihat dari waktu komputasi antara algoritma K-Means dan K-Medoids dalam pengelompokkan dokumen skripsi. Dalam peneilitian ini, waktu yang diperlukan dalam pengelompokkan dokumen skripsi

dilakukan di tool RapidMiner. Berikut ialah waktu komputasi serta rata-rata waktu yang dibutuhkan pada algoritma K-Means dan K-Medoids dalam mengelompokkan 50 dokumen skripsi:

Tabel 2. Perbandingan waktu komputasi yang dibutuhkan dalam 2 kali percobaan pada Algoritma K-Means dan K-Medoids

Jumlah K	Percobaan	K-Means	K-Medoids
2	1	1 *sec*	30 *sec*
	2	1 *sec*	31 *sec*
3	1	1 *sec*	22 *sec*
	2	1 *sec*	22 *sec*
4	1	1 *sec*	28 *sec*
	2	1 *sec*	29 *sec*
5	1	1 *sec*	26 *sec*
	2	1 *sec*	25 *sec*
6	1	1 *sec*	25 *sec*
	2	1 *sec*	24 *sec*
7	1	1 *sec*	29 *sec*
	2	1 *sec*	30 *sec*
8	1	1 *sec*	24 *sec*
	2	1 *sec*	25 *sec*
9	1	1 *sec*	29 *sec*
	2	1 *sec*	29 *sec*
10	1	1 *sec*	27 *sec*
	2	1 *sec*	27 *sec*

Setelah dilakukan 2 kali pada percobaan dalam mengelompokkan 50 dokumen pada 2 algoritma, selanjutnya dihitung waktu rata-rata komputasi pada algoritma K-Means dan K-Medoid. Adapun waktu komputasi rata-rata pada kedua algoritma dapat dilihat pada tabel berikut ini:

Tabel 3. Perbandingan Rata-rata waktu komputasi yang dibutuhkan pada Algoritma K-Means dan K-Medoids

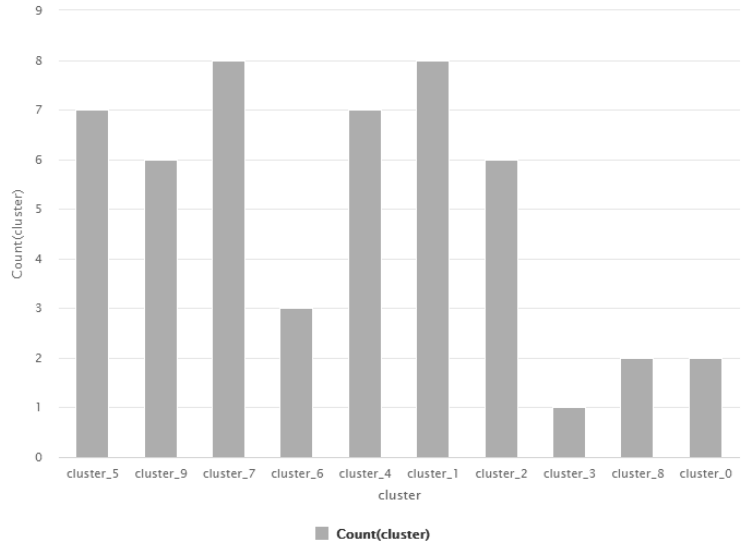
Jumlah K	K-Means	K-Medoids
2	1 *sec*	30,5 *sec*
3	1 *sec*	22 *sec*
4	1 *sec*	28,5 *sec*
5	1 *sec*	25,5 *sec*
6	1 *sec*	24,5 *sec*
7	1 *sec*	29,5 *sec*
8	1 *sec*	24,5 *sec*
9	1 *sec*	29 *sec*
10	1 *sec*	27 *sec*
Rata-rata	1 *sec*	26,7778 *sec*

Dapat dilihat dari Tabel 2., perbandingan waktu yang dibutuhkan pada algoritma K-Means dan K-Medoids untuk mengelompokkan 50 dokumen skripsi memiliki nilai rata rata yaitu 1 detik dan 26,6 detik. Nilai rata-rata waktu yang dibutuhkan diambil dari waktu yang dibutuhkan tiap tiap cluster dalam pengelompokkan dokumen. Dapat disimpulkan pada

https://doi.org/10.31849/digitalzone.v13i1.9292

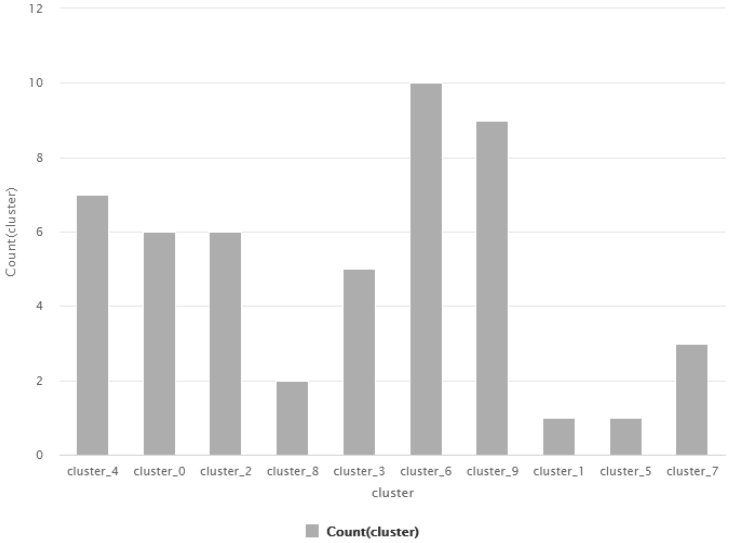
penelitian ini bahwa, waktu yang dibutuhkan oleh algoritma K-Means untuk mengelompokkan data (dokumen skripsi) lebih cepat dibandingkan dengan algoritma K-Medoids. Dibuktikan dengan waktu komputasi rata-rata pada algoritma K-Means ialah 1 detik, sementara pada K-Medoids membutuhkan waktu komputasi rata-rata ialah 26,6 detik.

Selain itu, keanggotaan cluster pada kedua algoritma berbeda. Berikut merupakan gambaran keanggotaan cluster pada kedua algoritma dengan jumlah cluster yang memiliki nilai DBI yang terbaik:



Gambar 6. Keanggotaan tiap *cluster* algoritma K-Means

Keanggotaan dokumen pada algoritma K-Means yaitu cluster 1 ialah 2 dokumen, cluster 2 ialah 8 dokumen, cluster 3 ialah 6 dokumen, cluster 4 ialah 1 dokumen, cluster 5 ialah 7 dokumen, cluster 6 ialah 7 dokumen, cluster 7 ialah 3 dokumen, cluster 8 ialah 8 dokumen, cluster 9 ialah 2 dokumen dan cluster 10 ialah 6 dokumen.



Gambar 7 Kenggotaan tiap *cluster* algoritma K-Medoids

Keanggotaan dokumen pada algoritma K-Medoids yaitu cluster 1 ialah 6 dokumen, cluster 2 ialah 1 dokumen, cluster 3 ialah 6 dokumen, cluster 4 ialah 5 dokumen, cluster 5 ialah 7 dokumen, cluster 6 ialah 1 dokumen, cluster 7 ialah 10 dokumen, cluster 8 ialah 3 dokumen, cluster 9 ialah 2 dokumen dan cluster 10 ialah 9 dokumen.

4. Kesimpulan

Kesimpulan yang didapat setelah penelitian dilakukan yaitu bahwa telah sukses dilakukan implementasi dan perbandingan terhadap metode K-Means dan K-Medoids clustering. Pada kasus pengelompokkan dokumen skripsi yang telah dilakukan, terdapat 2 poin yang dapat menjadi acuan perbandingan yaitu pengujian nilai Davies bouldin index serta pengujian yang dilihat dari waktu komputasi rata-rata yang dibutuhkan untuk mengelompokkan dokumen pada kedua algoritma. Adapun kesimpulan yang didapat yaitu dalam pengujian yang dilakukan pada perhitungan nilai Davies bouldin index (DBI), dapat disimpulkan bahwa algoritma K-Medoids memiliki nilai rata-rata DBI yang lebih mendekati 0 (dan tidak bernilai negatif) yaitu 1,56, yang diuji pada setiap cluster. Sementara rentang nilai DBI yang dimiliki oleh K-Means memiliki rentang yang cukup jauh dari angka 0 yaitu 2,79. Melihat kedua algoritma pada pengujian DBI, maka dapat dikatakan bahwa dalam pengujian pada penelitian ini, K-Medoids memiliki keunggulan dalam pengelompokkan dokumen. Serta dalam pengujian waktu komputasi yang dibutuhkan, untuk mengelompokkan 50 dokumen skripsi, algoritma K-Means memiliki waktu rata-rata sebesar 1 detik sementara pada algortima K-Medoids memiliki waktu rata-rata sebesar 26,7778 detik. Melihat kedua algoritma pada pengujian waktu komputasi yang dibutuhkan, maka dapat dikatakan bahwa dalam penelitian ini, K-Means memiliki keunggulan dalam waktu komputasi dalam mengelompokkan 50 dokumen skripsi.

Daftar Pustaka

[1] R. Siti, "Sistem Pencegahan Plagiarisme Tugas Akhir Menggunakan Algoritma Rabin-Karp (Studi Kasus: Sekolah Tinggi Teknik Payakumbuh)," *J. Teknol. Inf. Komun. Digit. Zo.*, vol. 6, no. 1, pp. 44–52, 2015.

[2] S. Andika and S. Ramadhani, "Rancang Bangun Sistem Informasi Pendayagunaan Aset Dinas Perkebunan Provinsi Riau," *J. Teknol. Dan Sist. Inf. Bisnis - JTEKSIS*, vol. 3, no. 2, pp. 387–394, Jul. 2021, doi: 10.47233/JTEKSIS.V3I2.298.

[3] S. R. Alvin Anzas Islami, "Rancang Bangun Sistem Pendataan Hardware," *J. Teknol. Dan Sist. Inf. Bisnis - JTEKSIS*, vol. 3, no. 2, pp. 412–418, 2021.

[4] Syah Maulana Ramadhan; Siti Ramadhani; Tomi Z;, "Perancangan Website Masyarakat Peduli Sampah Kelurahan Ratu Sima," *J. Has. Penelit. dan Pengkaj. Ilm. Eksakta*, vol. 01, no. 01, pp. 40–49, 2022.

[5] R. Simaremare, "Implikasi Perkembangan Dan Internet Diindonesia," *Teknologi Informasi Dan Dunia Pendidikan*, Vol. 2, 2009.

[6] D. Ayu *Et Al.*, "Pengukuran Kemiripan Dokumen Teks Bahasa Indonesia Menggunakan Metode Cosine Similarity," *E-Journal Teknik Informatika*, Vol. 9, No. 1, Pp. 1–8, 2016.

[7] F. Nur, M. Zarlis, And B. B. Nasution, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolah Menengah Kejuruan Untuk Clustering Jurusan," No. 9, Pp. 100–105, 2015.

[8] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, And D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *Jti (Jurnal Teknologi Informasi)*, Vol. 4, No. 1, Pp. 166–173, 2020.

[9] E. Muningsih, "Komparasi Metode Clustering K-Means Dan K-Medoids Dengan Model Fuzzy Rfm Untuk Pengelompokan Pelanggan," *Evolusi : Jurnal Sains Dan Manajemen*, Vol. 6, No. 2, 2018, Doi: 10.31294/Evolusi.V6i2.4600.

[10] D. Azzahra dan S. Ramadhani, "Pengembangan Aplikasi Online Public Access Catalog (Opac) Berbasis Web Pada Stai Auliaurrasyiddin Tembilanan," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, Vol. 2, No. 2, Pp. 152–160, 2020.

[11] R. A. Atmala And S. Ramadhani, "Rancang Bangun Sistem Informasi Surat Menyurat Di Kementrian Agama Kabupaten Kampar," *Jurnal Intra Tech*, Vol. 11, No. 2, Pp. 56–62, 2018.

https://doi.org/10.31849/digitalzone.v13i1.9292

[12] R. Nazwita, Siti, "Analisis Sistem Keamanan Web Server Dan Database Server Menggunakan Suricata," In *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (Sntiki) 9*, 2017, Pp. 308–317.
[13] S. Ramadhani, S. Saide, And R. E. Indrajit, "Improving Creativity Of Graphic Design For Deaf Students Using Contextual Teaching Learning Method (Ctl)," In *Acm International Conference Proceeding Series*, 2018, Pp. 136–140. Doi: 10.1145/3206098.3206128.
[14] M. R. Saputra, S. Ramadhani, And S. Baru, "Sistem Informasi Bantuan Dana Hibah Operasional Rumah Ibadah Kabupaten Bengs," *Jurnal Teknologi Dan Informasi Bisnis*, Vol. 3, No. 1, P. 148, 2021.
[15] S. Ramadhani, "A Review Comparative Mammography Image Analysis On Modified Cnn Deep Learning Method," *Indonesian Journal Of Artificial Intelligence (Ijaidm)*, Vol. 4, No. 1, Pp. 54–61, 2021.
[16] M. R. Asyari And S. Ramadhani, "Sistem Informasi Arsip Surat Menyurat," *Jurnal Teknologi Dan Informasi Bisnis*, Vol. 3, No. 1, Pp. 175–184, 2021.