

ANALISIS PERFORMA DAN KECEPATAN KOMPUTASI ALGORITMA K-MEANS DAN K-MEDOIDS PADA TEXT CLUSTERING

Karno Nur Cahyo¹, Agus Subekti², Muhammad Haris³

Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Nusa Mandiri

Jl. Raya Jatiwaringin No.2, Cipinang Melayu, Kec. Makassar, Jakarta Timur, DKI Jakarta 13620

14002459@nusamandiri.ac.id¹, agus@nusamandiri.ac.id², muhammad.uhs@nusamandiri.ac.id³

ARTICLE INFO

Article history:

Received September 2022

Received in revised form Oktober 2022

Accepted Nov 2022

Available online Desember 2022

ABSTRACT

The large number of theses will certainly make it difficult to find categories on thesis topics that have been written by students at a university. One of the uses of the Text Mining method is being able to group thesis objects into the number of clusters formed by the clustering algorithm. This study aims to compare 2 clustering algorithms, namely the K-Means and K-Medoids algorithms to obtain an accurate evaluation of the performance and computational time in the case of thesis clustering, so that relevant topics can be grouped and have better clustering accuracy. The evaluation parameter used is the Davies Bouldin Index (DBI) which is one of the testing techniques on clustering results, with the distribution of training data and testing data using cross validation using a repetition parameter of 10 folds iteration. From the results of the study with the Term Weighting condition used is Term Occurrences and using the N-Grams value is 2, it can be concluded that the K-Means algorithm has a better DBI value of -0.426. Meanwhile, the range of DBI values owned by K-Medoids with the same conditions has a DBI value of -1,631. However, from the visualization results using t-SNE with the same supporting parameters, there are options that can be used, namely the number of clusters is 6, and the DBI value is -1.110. For testing the computational time in the clustering process of 50 thesis documents, the K-Means algorithm has an average time of 2.5 seconds while the K-Medoids algorithm has an average time of 261.5 seconds. The computer specifications used are Asus ZenBook UX425EA.312 with the processor used is 11th Gen Intel® Core™ i5-1135G7 @ 2.40GHz @ 2.40GHz, the graphics card is Intel® Iris® Xe Graphics, the RAM used is 8GB, with storage of 512GB SSD.

Keywords: clustering, text mining, thesis grouping, k-means, k-medoids, Davies Bouldin Index, t-SNE visualization, cluster determination performance, cluster determination computation time

Abstrak

Banyaknya jumlah skripsi tentunya akan menyulitkan dalam mencari kategori pada topik skripsi yang sudah ditulis oleh mahasiswa di suatu universitas. Salah satu pemanfaatan metode Text Mining adalah mampu mengelompokkan obyek skripsi ke dalam jumlah kluster yang dibentuk dengan algoritma clustering. Penelitian ini bertujuan untuk melakukan perbandingan 2 algoritma clustering yaitu algoritma K-Means dan K-Medoids untuk mendapatkan evaluasi yang tepat dari sisi performa dan waktu komputasi pada kasus pengelompokan skripsi, sehingga topik yang relevan dapat dikelompokkan dan memiliki keakuratan jumlah kluster yang lebih baik. Parameter evaluasi yang digunakan adalah Davies Bouldin Index (DBI) yang merupakan salah satu teknik pengujian pada hasil clustering, dengan pembagian data training dan data testing menggunakan cross validation dengan parameter pengulangan 10 folds iteration. Dari hasil penelitian dengan kondisi Term Weighting yang digunakan adalah Term Occurrences dan menggunakan nilai N-Grams adalah 2 dapat disimpulkan bahwa algoritma K-Means memiliki nilai DBI yang lebih baik yaitu -0,426. Sementara rentang nilai DBI yang dimiliki oleh K-Medoids dengan kondisi yang sama memiliki nilai DBI yaitu -1,631. Namun dari hasil visualisasi menggunakan t-SNE dengan parameter pendukung yang sama, terdapat opsi yang bisa digunakan yaitu dengan jumlah kluster adalah 6, dengan nilai DBI adalah -1,110. Untuk pengujian waktu komputasi dalam proses clustering 50 dokumen skripsi, algoritma K-Means memiliki waktu rata-rata sebesar 2,5 detik sementara pada algoritma K-Medoids memiliki waktu rata-rata sebesar 261,5 detik. Spesifikasi komputer yang digunakan adalah Asus ZenBook UX425EA.312 dengan prosesor yang digunakan adalah 11th Gen Intel® Core™ i5-1135G7 @ 2.40GHz @ 2.40GHz, kartu grafis adalah Intel® Iris® Xe Graphics, RAM yang digunakan adalah 8GB, dengan penyimpanan sebesar 512GB SSD.

Kata kunci: clustering, text mining, pengelompokan skripsi, k-means, k-medoids, Davies Bouldin Index, visualisasi t-SNE, performa penentuan cluster, waktu komputasi penentuan cluster

1. PENDAHULUAN

Saat ini, karya ilmiah di bidang pendidikan tercipta di setiap tahun, seperti Tugas Akhir, Skripsi, Thesis, Disertasi. 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 berdasarkan hasil penelitian yang ditemukan di lapangan dan dilakukan tahapan referensi dan studi kepustakaan. Hal itu tentunya sesuai bidang topik berdasarkan program studi atau jurusan masing-masing. Skripsi menunjukkan kompetensi sebagai mahasiswa dalam bidang studi Strata 1 dalam penelitian yang berkaitan dengan bidang keahliannya/program studinya. 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 skripsi yang dibuat di setiap semesternya tentunya akan menyulitkan dalam mencari kategori skripsi/karya ilmiah yang sudah 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 text mining pada skripsi dapat mempermudah dalam melakukan pengelompokan kategori topik dari skripsi tersebut. Di dalam text mining, terdapat metode yang dapat digunakan yaitu metode clustering[2]. Terdapat beberapa metode clustering yang biasa digunakan dalam penelitian, baik di dalam data teks maupun di luar data teks. Adapun beberapa metode clustering yang biasanya digunakan adalah K-means, K-Medoids, Fuzzy K-Means, K-medians dan lainnya. Namun dalam penelitian ini, skripsi dikelompokkan/dilakukan clustering menggunakan 2 Algoritma yaitu K-Means dan K-Medoids, selanjutnya akan dibandingkan jumlah kluster terbaik pada 2 algoritma tersebut dalam penerapannya pada alur proses text mining.

Berbagai penelitian terkait mengenai pengelompokan 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 relatif cepat dalam pengelompokan data, sedangkan algoritma K-Medoids memiliki waktu komputasi yang cukup lama dalam pengelompokan data dalam parameter K cluster[2].

Berbeda dengan penelitian sebelumnya, maka pada penelitian ini membahas perbandingan algoritma K-Means dan algoritma K-Medoids pada text mining untuk mengklasifikasikan skripsi mahasiswa Fakultas Ekonomi Sosial dalam beberapa kelompok yang diujikan. Data training dan data testing diuji divalidasi menggunakan cross validation dengan parameter pengulangan 10 folds iteration. Selanjutnya dilakukan pengujian antara 2 algoritma tersebut dengan menggunakan pengujian menghitung nilai DBI (Davies Bouldin Index) dan juga dengan melihat hasil visualisasi pengelompokan kluster dengan metode t-SNE.

2. METODE PENELITIAN

2.1. Jenis Penelitian

Peneliti melakukan eksperimen terhadap data 50 skripsi yang merupakan data publik, yang diambil dari website repository Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi negara. Data tersebut digunakan untuk menganalisis kluster dengan performa dan waktu komputasi terbaik berdasarkan perbandingan 2 algoritma yaitu K-Means dan K-Medoids.

2.2. Metode Pengolahan Data

Penelitian ini menggunakan metode Knowledge Discovery in Database (KDD), tahapan yang harus dilakukan dalam metode ini, yaitu: Pemahaman tentang domain tujuan pengolahan data (Domain Understanding), proses pemilihan data (Selection & Addition Data), Pra pemrosesan data untuk membersihkan data (Data Preprocessing & Data Cleaning), mentransformasi data (Transformation), Proses Data Mining, dan proses evaluasi serta interpretasi dari hasil yang diinginkan (Evaluation & Interpretation), yang dirincikan pada Gambar 1.



Gambar 1. Metode Penelitian KDD

- A. Domain Understanding
Pada langkah ini, tujuan melakukan data mining ditentukan, tujuan penelitian ini adalah untuk menganalisis performa dan waktu komputasi terbaik dari proses *text clustering* data skripsi mahasiswa Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial Program Studi Administrasi negara Tahun 2022, dengan tujuan menentukan sebaran kluster yang paling baik untuk data 50 skripsi tersebut.
- B. Selection & Addition
Pada tahap ini, data diambil dan dipilih dari 50 skripsi terbaru dari mahasiswa Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi negara, yang diambil dari *web repository* Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi negara dengan alamat URL <https://repository.uin-suska.ac.id/view/divisions/ek=5Fan/2022.html>.
- C. Data Preprocessing
Pada tahap ini, menggunakan 6 tahapan *data preprocessing* yang terdapat di RapidMiner, antara lain dengan penentuan n-Grams, penggunaan operator *transform case*, penggunaan *stopwords removal*, penggunaan *filter tokens (by Length)*, penggunaan *Stemming*, dan juga penggunaan *Term Weighting*. Penentuan n-Grams digunakan untuk memberikan probabilitas kemungkinan pada kata berikutnya yang mungkin dapat digunakan untuk melakukan kemungkinan penggabungan pada keseluruhan kalimat. Operator *transform case* adalah proses *data preprocessing* yang mengubah isi data menjadi seragam, bisa menjadi kapital semua, atau bisa menjadi huruf kecil semua, agar proses dan isi dari data menjadi seragam. Lalu *stopwords removal* adalah metode *data preprocessing* yang sering digunakan pada mekanisme *Text Mining* untuk menghilangkan kata-kata penghubung/kata-kata yang tidak dibutuhkan seperti di, ke, dari, yang, dan sebagainya. Kemudian mekanisme *filter token (by length)*, yaitu proses *data preprocessing* untuk menghapus kata-kata yang terlalu panjang. Kemudian *Stemming* adalah suatu proses yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya atau ke kata-kata dasarnya (*root word*). Terakhir adalah penggunaan vektorisasi *Term Weighting* merupakan proses penghitungan bobot tiap *term* di dalam dokumen.
- D. Transformasi Data
Pada tahap ini, data diubah menjadi bentuk yang lain, perubahan yang dilakukan adalah perubahan bentuk data dari data tidak terstruktur yang berisi uraian latar belakang 50 skripsi menjadi data terstruktur, untuk dilakukan proses pengolahan data mining dengan algoritma *clustering*. Proses transformasi data menggunakan mekanisme tokenisasi.
- E. Data Mining
Penelitian ini menggunakan algoritma data mining untuk dilakukan proses analisis performa dan kecepatan komputasi dari proses data mining, dengan menggunakan 2 algoritma *clustering* yaitu K-Means dan K-Medoids.
- F. Interpretasi atau Evaluasi
Data yang dihasilkan dari algoritma menjadi informasi yang diharapkan dapat membantu melihat kluster pengelompokan skripsi mahasiswa Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi negara. Selain itu untuk melakukan pengukuran analisis kemampuan algoritma clustering, dilakukan pengukuran performa menggunakan *Cluster Distance Performance* yang diukur menggunakan *Davies Bouldin Index (DBI)*. Selain menggunakan DBI, parameter yang dianalisis adalah kecepatan waktu komputasi dari proses pengolahan data. Evaluasi data training dan data testing dilakukan dengan bantuan operator *cross validation* dengan parameter *10 folds iteration*. Kemudian untuk interpretasi visualisasi hasil penyebaran cluster menggunakan bantuan operator t-SNE.

2.3. Deskripsi Atribut Dataset

Dataset yang digunakan terdiri dari kumpulan dataset uraian latar belakang dari Bab 1 pada dokumen skripsi, 50 skripsi terbaru Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi Negara di tahun 2022 dengan 1 atribut yaitu Uraian Latar Belakang, kemudian dilakukan tokenisasi, memunculkan 6.302 atribut.

Tabel 2. Dataset Uraian Skripsi Mahasiswa

Atribut	Description	Type
Uraian Latar Belakang	Daftar uraian latar belakang dari 50 skripsi terbaru di tahun 2022, dari Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi Negara	text

2.4. Tahapan Preprocessing

- A. Tahap ujicoba sistem adalah tahap pengujian sistem yang sudah diintegrasikan antara rancangan perangkat lunak dan perangkat keras. Melakukan sinkronisasi keduanya, apakah sistem sudah berjalan sesuai dengan yang diinginkan. N-Grams

N-gram adalah proses yang dipakai untuk melakukan *text mining* atau pemrosesan uraian text, untuk memperoleh hingga n karakter dari sebuah string. Segmentasi kata melakukan proses ini dengan memisahkan hingga n-kata dari rangkaian kata (paragraf, kalimat, bacaan) yang dibaca secara berurutan dari awal hingga akhir dokumen. Ukuran 1 n-Gram disebut unigram, ukuran 2 n-Gram disebut bigram, dan ukuran 3 n-Gram disebut trigram [16].

B. Transform Case

Operator Transform Case adalah operator yang digunakan untuk menyeragamkan jenis dari *value/data*, penyeragaman ini bisa bersifat *Lower Case* atau *Upper Case*, *Lower Case* apabila ingin *value/data* bertuliskan seluruhnya huruf kecil, apabila *Upper Case* jika ingin *value/data* seluruhnya huruf kapital. Pada Gambar 2 dapat dilihat bahwa seluruh *value* sebelum diberikan operator transform case berawalan huruf kapital, namun setelah diberikan operator transform case dengan jenis lower case, seluruh *value* berubah menjadi huruf kecil.

Sebelum	Sesudah
Aamin	aamin
Abadi	abadi
Abah	abah
Abai	abai
Abaikan	abaikan
Covid	covid
Ekonomi	ekonomi
Masker	masker
Virus	virus

Gambar 2. Contoh Hasil Penerapan *Operator Transform Case*

C. Stopwords

Operator Stopword adalah operator yang digunakan untuk menghilangkan kata-kata penghubung/kata-kata yang tidak dibutuhkan atau kata-kata yang tidak memiliki makna begitu penting seperti di, ke, dari, yang, supaya dan sebagainya. Penggunaan *Stopword* dibantu oleh file kamus sesuai dengan sumber bahasa dari kata yang ingin kita hilangkan, termasuk file kamus stopwords Bahasa Indonesia. Pada gambar 3 adalah contoh penerapan operator *Stopword* di sebuah kumpulan data.

Sebelum	Sesudah
aamin	aamin
abadi	abadi
abah	abah
oleh	abai
olehnya	abaikan
pada	covid
padahal	ekonomi
para	masker
di	virus
abai	
abaikan	
covid	
ekonomi	
masker	
virus	
ke	
dari	
yang	

Gambar 3. Contoh Hasil Penerapan Operator Stopwords

D. Filter Tokens (By Length)

Operator Filter Tokens (By Length) adalah operator yang digunakan untuk menghilangkan kata-kata yang panjang karakternya tidak sesuai standar atau terlalu panjang. Parameter yang digunakan adalah parameter minimal jumlah karakter dan juga maksimal jumlah karakter sebagai pembatas value karakter yang digunakan. Pada Gambar 4 adalah contoh hasil penggunaan operator *Filter Tokens (By Length)* dengan penggunaan parameter minimal jumlah karakter adalah 4 dan maksimal jumlah karakter adalah 10.

Sebelum	Sesudah
aamin	aamin
abadi	abadi
abah	abah
argumentasinya	abai
mempengaruhi	abaikan
membersihkan	covid
mengundang	ekonomi
mendaurulang	masker
abai	virus
abaikan	
covid	
ekonomi	
masker	
virus	

Gambar 4. Contoh Hasil Penerapan Operator *Filter Token (By Length)*

E. Stemming

Stemming adalah suatu proses yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya atau ke kata-kata dasarnya (*root word*). Stemming menyediakan pemetaan variasi kata morfologis yang berbeda ke dalam kata dasar/umum mereka [17].

Sebelum	Sesudah
Analisis	Analisis
Penyelenggaraan	selenggara
Pelayanan	Layan
Publik	Publik
Kota	Kota
Pelayanan Publik	Layan Publik

Gambar 5. Contoh Hasil Penerapan Stemming

F. Term Weighting

Term Weighting adalah proses penghitungan bobot tiap term yang dicari pada setiap dokumen sehingga dapat diketahui ketersediaan dan kemiripan suatu term di dalam dokumen. Skema term weighting memainkan peran penting dalam klasifikasi teks. Dibutuhkan suatu metode term weighting yang dapat menghasilkan lebih banyak term yang kaya informasi dan menetapkan nilai bobot term yang sesuai untuk persyaratan klasifikasi teks [17].

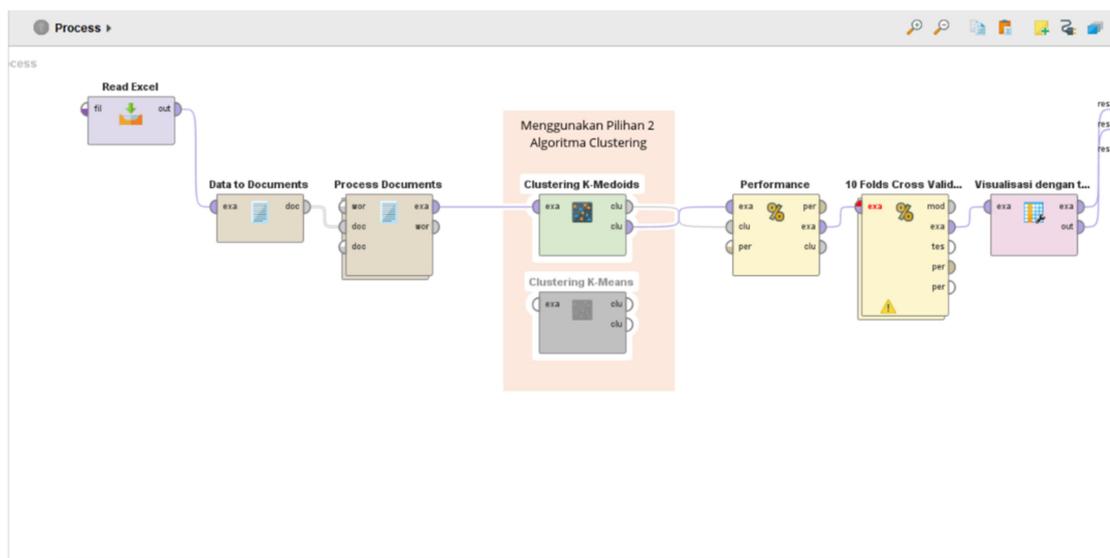
3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Tujuan dari penelitian ini adalah menganalisis performa dan waktu komputasi terbaik dari proses text clustering uraian latar belakang dari 50 skripsi mahasiswa Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi Negara tahun 2022. Selain itu untuk menentukan jumlah kluster yang paling baik dari pengelompokan skripsi tersebut. Setelah jumlah kluster yang paling baik dimunculkan, hasil penelitian selanjutnya adalah dapat memunculkan hasil visualisasi dari jumlah kluster yang sudah dibentuk.

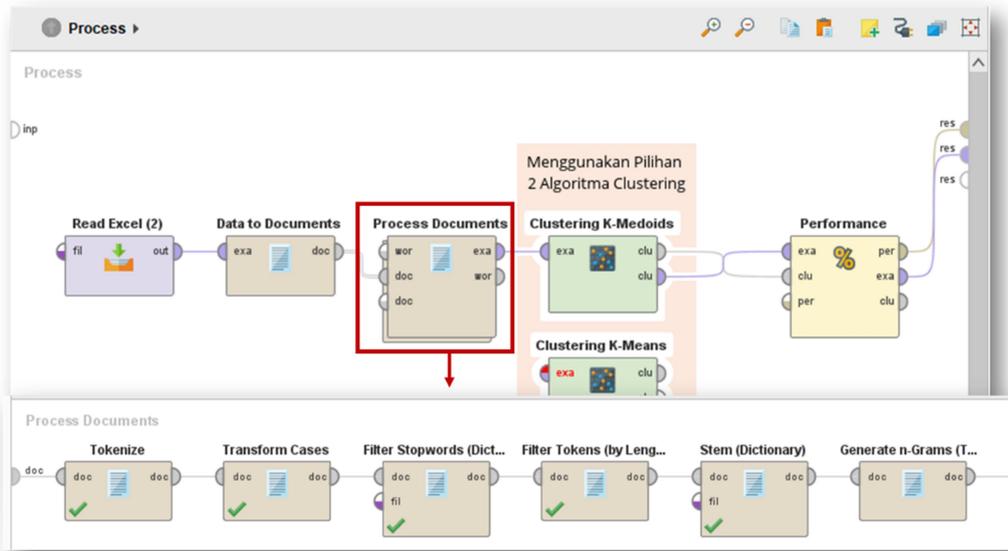
3.2. Text Processing

Dari data eksperimen akan dilakukan pembersihan data terlebih dahulu menggunakan beberapa operator di *tools RapidMiner*. Gambar penerapan model yang dibuat di *tools RapidMiner* ditampilkan pada Gambar 6.



Gambar 6. Model RapidMiner Klasterisasi

Setelah data 50 skripsi didapatkan dari web repository Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi Negara, 50 skripsi tersebut akan diambil bagian latar belakang dari Bab I untuk menjadi dataset yang berisi 50 uraian latar belakang dari skripsi yang sudah didapatkan dari web repository Universitas Islam Negeri Sultan Syarif Kasim Riau. Kemudian dataset tersebut akan diproses menggunakan metode text processing, melakukan perubahan data dari tidak terstruktur menjadi data terstruktur dengan penggunaan tokenisasi. Setiap kata pada uraian latar belakang yang terdapat di Bab I akan ditokenisasi, dan akan menjadi atribut-atribut data terstruktur. Proses *text processing* berada di dalam sub proses Process Domain di model RapidMiner klasterisasi data laporan bhabinkamtibmas. Gambar *Sub process* di operator *Process Documents* terdapat pada Gambar 7.



Gambar 7. Model Penerapan *Text Processing*

Pada Gambar di atas ditampilkan proses pertama adalah melakukan tokenisasi, kemudian melakukan pengubahan jenis huruf dari atribut, yang awalnya tidak beraturan jenis huruf kapital dan non kapitalnya, semuanya diubah menjadi huruf kecil semua, dengan menggunakan operator *Transform Case*. Kemudian dilakukan kembali pembersihan data menggunakan operator *filter stopwords removal*, fitur ini menghilangkan kata-kata hubung seperti dan, ke, dari, dan sebagainya, agar tidak masuk ke dalam proses text mining, termasuk kata-kata yang tidak dibutuhkan. Kemudian melakukan filterisasi kembali dengan menggunakan *filter tokens (by length)*, yaitu fitur untuk menghapus atribut yang terlalu panjang. Setelah menggunakan *filter tokens (by length)* selanjutnya adalah menggunakan operator Stem untuk mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya atau ke kata-kata dasarnya (*root word*). Pada Gambar 8 ditampilkan hasil dari sesudah proses *Data Preprocessing*.

Open in Turbo Prep Auto Model

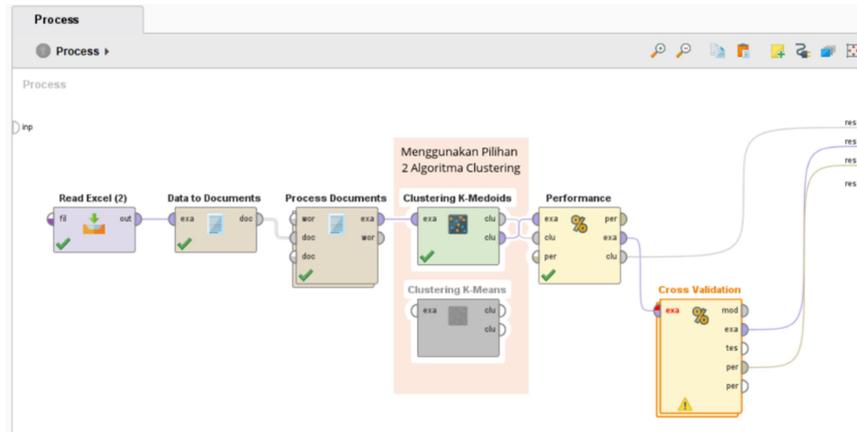
Filter (50 / 50 examples): all

Row No.	abadi	abai	abdi	abdai	abodemen	abrasi	absen	absensi	acara	access	accessories	aceh	achmad	actual	actual
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0.027	0	0	0	0	0	0
3	0.029	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0.052	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0.103	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0.027	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0.005	0.010	0	0	0	0.008	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0.157	0.172	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0.017	0	0	0	0	0	0
19	0	0	0	0	0	0.636	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0.023	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0.015	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0.011	0	0.011	0	0	0	0	0	0

ExampleSet (50 examples, 0 special attributes, 6,302 regular attributes)

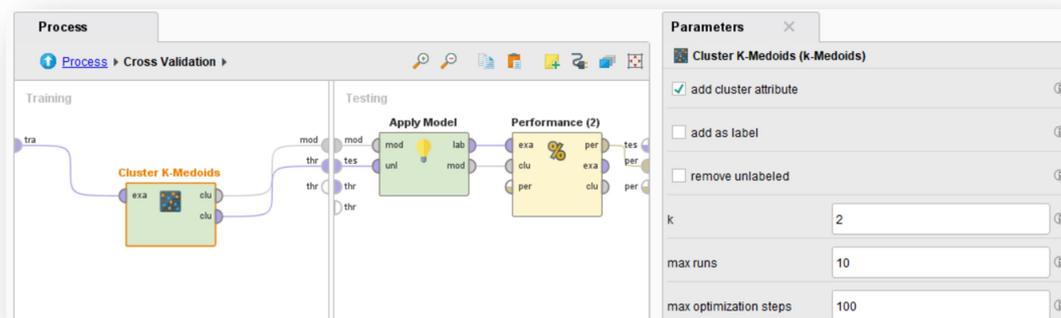
Gambar 8. Hasil Sesudah *Data Preprocessing*

Evaluasi pembagian data *testing* dan data *training* menggunakan operator 10 fold cross validation yang terdapat di tools RapidMiner, pada data *training* menggunakan algoritma cluster K-Medoids dengan jumlah folds 10. Ditambahkan operator *Cross Validation* untuk menampilkan performa dengan kondisi pembagian data *training* dan *data testing*, sekaligus dengan ExampleSet datanya yang ditampilkan pada Gambar 9.



Gambar 9. Model Baru dengan ditambahkan 10 Fold Cross Validation

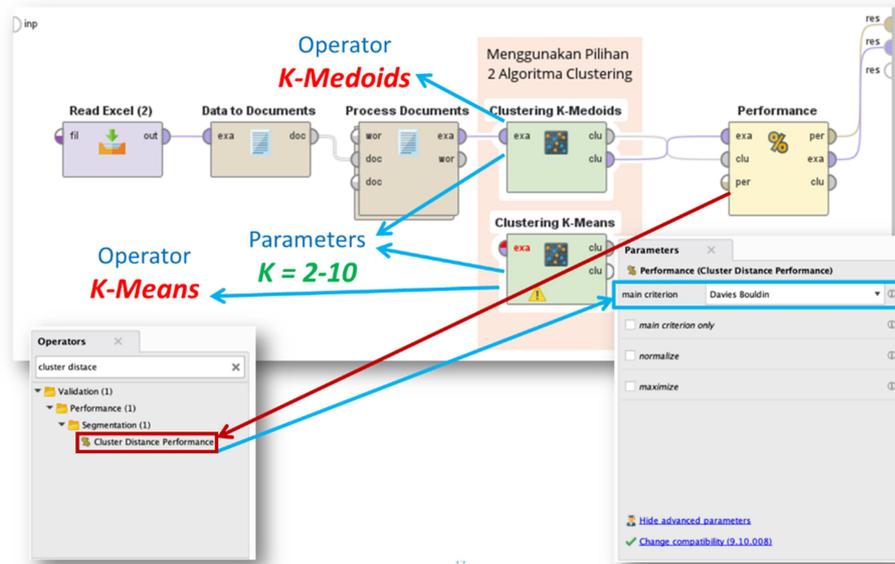
Di dalam operator Cross Validation ditambahkan evaluasi pengujian data training dan data testing, dengan penjabaran yang dapat dilihat pada Gambar 10. dengan menggunakan algoritma K-Medoids ataupun K-Means.



Gambar 10. Penerapan 10 Folds Cross Validation

3.2.1 Cluster Distance Performance

Cluster Distance Performance digunakan untuk mengukur performa dari hasil klasterisasi yang dilakukan. Mekanisme pengukurannya menggunakan kriteria utama Davies Bouldin Index (DBI). Model RapidMiner yang digunakan untuk melakukan pengukuran performa terdapat pada Gambar 11.



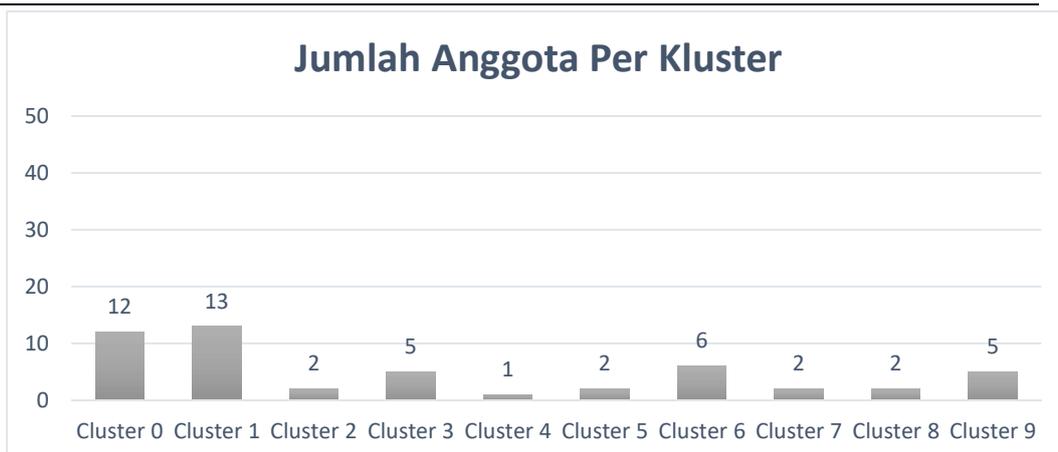
Gambar 11. Model Pengukuran Performa Clustering Menggunakan DBI

Dari model di atas, dilakukan 216 kali pengujian performa, dengan menguji parameter K berjumlah 2 sampai dengan 10, menguji 2 algoritma yaitu K-Means dan K-Medoids, dengan menguji 4 jenis term weighting yaitu TF-IDF, Term Frequency, Term Occurrences, dan Binary Term Occurrences, dengan pengujian masing-masing 1-3 n-Grams. Nilai DBI bersifat absolut, sehingga nilai negatif tidak diperhatikan, jumlah kluster yang paling baik adalah yang nilai DBInya mendekati angka 0. Hasil pengujian performa kluster untuk parameter Term-Weighting menggunakan TF-IDF dengan N-Grams adalah 1 ditampilkan pada Gambar 12.

Pengujian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-3,708	-1,905
3	-3,752	-1,748
4	-3,526	-1,705
5	-3.378	-1,616
6	-3.098	-1,646
7	-2.885	-1,587
8	-2.742	-1,561
9	-2.464	-1,484
10	-2.320	-1,443

Gambar 12. Hasil Pengujian Nilai DBI TFIDF n-Grams 1

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 10 dengan menggunakan algoritma K-Medoids dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 1.443, yang paling kecil di antara DBI kluster yang lain. Dari hasil pengujian menggunakan DBI ditemukan bahwa kluster berjumlah 10 adalah kluster yang paling baik dengan Term Weighting adalah TF-IDF, N-Grams adalah 1 untuk mengelompokkan skripsi Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi Negara. 10 kluster yang dihasilkan adalah kluster 0 sampai dengan kluster 9, Dengan kondisi jumlah pengelompokan skripsi menggunakan algoritma K-Medoids dapat dilihat pada Gambar 13.



Gambar 13. Gambar Hasil Klasterisasi Menggunakan Algoritma K-Medoids

Dari distribusi data tersebut dapat dibuatkan beberapa sample data dari hasil distribusi di atas, yang dapat dilihat pada Gambar 14.

Judul Skripsi	Kluster
1. ANALISIS PENGEMBANGAN KAPASITAS BIDANG PENGELOLAAN SAMPAH DINAS LINGKUNGAN HIDUP DAN KEBERSIHAN KOTA PEKANBARU	cluster_0
7. IMPLEMENTASI PENGARUSUTAMAAN GENDER (PUG) DALAM MEWUJUDKAN TUJUAN PEMBANGUNAN BERKELANJUTAN (SUSTAINABLE DEVELOPMENT GOALSSDGS) DI KOTA PEKANBARU	cluster_0
2. ANALISIS PELAKSANAAN FUNGSI BADAN PERMUSYAWARATAN DESA (BPD) DALAM PENYELENGGARAAN PEMERINTAH DESA DI DESA SEI KUNING KECAMATAN TANDUN KABUPATEN ROKAN HULU	cluster_1
3. ANALISIS PERANAN KEPALA DESA DALAM PEMBANGUNAN INFRASTRUKTUR "DESA PAKNING ASAL" KABUPATEN BENGKALIS	cluster_1
19. KOORDINASI BADAN PENANGGULANGAN BENCANA DAERAH (BPBD) DAN MASYARAKAT DALAM PENANGGULANGAN ABRASI DI KECAMATAN TEBING TINGGI BARAT KABUPATEN KEPULAUAN MERANTI	cluster_2
40. ANALISIS PERAN BADAN PENANGGULANGAN BENCANA DAERAH (BPBD) DALAM MENANGGULANGI KABUT ASAP DI KOTA PEKANBARU	cluster_2
4. PENERAPAN PROGRAM KAMPUNG HIJAU DI KABUPATEN SIAK	cluster_3
17. KUALITAS PELAYANAN PEMBAYARAN PAJAK KENDARAAN BERMOTOR DI KANTOR BERSAMA SAMSAT UPT KOTA PINANG KABUPATEN LABUHANBATU SELATAN	cluster_3
41. PERAN BADAN NARKOTIKA NASIONAL KOTA PEKANBARU DALAM PENCEGAHAN, PEMBERANTASAN, PENYALAHGUNAAN DAN PEREDARAN GELAP NARKOTIKA DAN PREKURSOR NARKOTIKA DI KOTA PEKANBARU	cluster_4
14. INOVASI PELAYANAN NASI UDUK (SEDERHANA SEKALI URUS ADMINISTRASI KEPENDUDUKAN) DI KABUPATEN INDRAGIRI HILIR	cluster_5
18. ANALISIS PROGRAM RIAU BERTANI OLEH DINAS PANGAN, TANAMAN PANGAN DAN HORTIKULTURA PROVINSI RIAU (STUDI KASUS DI KABUPATEN KAMPAR)	cluster_5
6. STRATEGI DINAS PANGAN TANAMAN PANGAN DAN HORTIKULTURA PROVINSI RIAU DALAM MENJAGA KETAHANAN PANGAN PADA ERA PANDEMI COVID-19	cluster_6
9. ETIKA PELAYANAN PUBLIK DI DINAS PENANAMAN MODAL DAN PELAYANAN TERPADU SATU PINTU (DPMPTSP) PROVINSI RIAU	cluster_6
25. ANALISIS PENGELOLAAN ARSIP DINAMIS BERDASARKAN PERATURAN PEMERINTAH NOMOR 28 TAHUN 2012 TENTANG KEARSIPAN DI DINAS PEMBERDAYAAN MASYARAKAT DESA KEPENDUDUKAN DAN PENCATATAN SIPIL	cluster_7
43. ANALISIS PENGELOLAAN PENGARSIPAN DI DINAS KEPENDUDUKAN DAN PENCATATAN SIPIL KOTA PEKANBARU	cluster_7
26. PENGELOLAAN ASET DAERAH TAMAN HUTAN KOTA DAN TAMAN SATWA OLEH PEMERINTAH KABUPATEN PASAMAN BARAT	cluster_8
33. PERAN DINAS PEMUDA OLAHRAHA DAN PARIWISATA DALAM PENGEMBANGAN TAMAN WISATA ALAM RIMBO PANTI KABUPATEN PASAMAN PROVINSI SUMATERA BARAT	cluster_8
13. ANALISIS KINERJA BADAN USAHA MILIK DESA (BUMDES) ARTHA MITRA KESUMA DALAM MENYEDIAKAN PELAYANAN KEPADA MASYARAKAT DESA BENCAH KESUMA KECAMATAN KABUN KABUPATEN ROKAN HULU	cluster_9
18. ANALISIS PROGRAM RIAU BERTANI OLEH DINAS PANGAN, TANAMAN PANGAN DAN HORTIKULTURA PROVINSI RIAU (STUDI KASUS DI KABUPATEN KAMPAR)	cluster_9

Gambar 14. Gambar Sample data pemetaan kluster menggunakan Algoritma K-Means

Pengujian selanjutnya adalah hasil pengujian performa kluster untuk parameter Term-Weighting menggunakan TF-IDF dengan N-Grams adalah 2 ditampilkan pada Gambar 15.

Penguian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-4,021	-1,906
3	-4,065	-1,807
4	-3,687	-1,792
5	-3,53	-1,768
6	-3,182	-1,678
7	-2,878	-1,632
8	-2,877	-1,618
9	-2,673	-1,594
10	-2,438	-1,476

Gambar 15. Hasil Penguian Nilai DBI TFIDF n-Grams 2

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 10 dengan menggunakan algoritma K-Medoids dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 1.473, yang paling kecil di antara DBI kluster yang lain. Penguian selanjutnya adalah hasil penguian performa kluster untuk parameter Term-Weighting menggunakan TF-IDF dengan N-Grams adalah 3 ditampilkan pada Gambar 16.

Penguian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-4,044	-1,905
3	-4,079	-1,816
4	-3,681	-1,801
5	-3,54	-1,776
6	-3,188	-1,729
7	-2,881	-1,659
8	-2,84	-1,631
9	-2,515	-1,561
10	-2,442	-1,526

Gambar 16. Hasil Penguian Nilai DBI TFIDF n-Grams 3

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 10 dengan menggunakan algoritma K-Medoids dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 1.526, yang paling kecil di antara DBI kluster yang lain. Penguian selanjutnya adalah hasil penguian performa kluster untuk parameter Term-Weighting menggunakan Term Frequency dengan N-Grams adalah 1 ditampilkan pada Gambar 17.

Penguian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-2,063	-1,794
3	-2,461	-1,63
4	-2,788	-1,741
5	-2,217	-1,622
6	-2,482	-1,632
7	-2,39	-1,546
8	-2,202	-1,53
9	-2,149	-1,49
10	-1,843	-1,446

Gambar 17. Hasil Penguian Nilai DBI Term Frequency n-Grams 1

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 10 dengan menggunakan algoritma K-Medoids dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 1.446, yang paling kecil di antara DBI kluster yang lain. Penguian selanjutnya adalah hasil penguian performa kluster untuk parameter Term-Weighting menggunakan Term Frequency dengan N-Grams adalah 2 ditampilkan pada Gambar 18.

Pengujian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-2,289	-1,97
3	-3,215	-1,684
4	-2,99	-1,67
5	-2,932	-1,701
6	-2,84	-1,599
7	-2,548	-1,564
8	-2,223	-1,537
9	-2,1	-1,519
10	-2,08	-1,469

Gambar 18. Hasil Pengujian Nilai DBI Term Frequency n-Grams 2

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 10 dengan menggunakan algoritma K-Medoids dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 1.469, yang paling kecil di antara DBI kluster yang lain. Pengujian selanjutnya adalah hasil pengujian performa kluster untuk parameter Term-Weighting menggunakan Term Frequency dengan N-Grams adalah 3 ditampilkan pada Gambar 19.

Pengujian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-2.447	-1.957
3	-3.069	-1.680
4	-3.048	-1.674
5	-2.831	-1.707
6	-2.904	-1.612
7	-2.623	-1.555
8	-2.278	-1.567
9	-2.164	-1.540
10	-2.163	-1.499

Gambar 19. Hasil Pengujian Nilai DBI Term Frequency n-Grams 3

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 10 dengan menggunakan algoritma K-Medoids dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 1.499, yang paling kecil di antara DBI kluster yang lain. Pengujian selanjutnya adalah hasil pengujian performa kluster untuk parameter Term-Weighting menggunakan Term Occurrences dengan N-Grams adalah 1 ditampilkan pada Gambar 20.

Pengujian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-1.593	-1.547
3	-1.164	-1.166
4	-1.134	-1.337
5	-1.077	-1.139
6	-1.007	-1.338
7	-1.339	-1.051
8	-1.324	-1.373
9	-1.395	-1.281
10	-1.435	-1,263

Gambar 20. Hasil Pengujian Nilai DBI Term Occurrences n-Grams 1

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 6 dengan menggunakan algoritma K-Means dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 1.007, yang paling kecil di antara DBI kluster yang lain. Pengujian selanjutnya adalah hasil pengujian performa kluster untuk parameter Term-Weighting menggunakan Term Occurrences dengan N-Grams adalah 2 ditampilkan pada Gambar 21.

Pengujian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-0,426	-1.631
3	-1,267	-1.236
4	-1,247	-1.449
5	-1,254	-1.150
6	-1,18	-1.110
7	-1,685	-1.122
8	-1,362	-1.200
9	-1,525	-1.334
10	-1,372	-1.284

Gambar 21. Hasil Pengujian Nilai DBI Term Occurrences n-Grams 2

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 2 dengan menggunakan algoritma K-Means dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 0,426, yang paling kecil di antara DBI kluster yang lain. Pengujian selanjutnya adalah hasil pengujian performa kluster untuk parameter Term-Weighting menggunakan Term Occurrences dengan N-Grams adalah 3 ditampilkan pada Gambar 22.

Pengujian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-0,437	-1.653
3	-1,358	-1.257
4	-1,468	-1.267
5	-1,332	-1.177
6	-1,123	-1.133
7	-1,57	-1.310
8	-1,48	-1.101
9	-1,362	-1.264
10	-1,17	-1.220

Gambar 22. Hasil Pengujian Nilai DBI Term Occurrences n-Grams 3

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 2 dengan menggunakan algoritma K-Means dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 0,437, yang paling kecil di antara DBI kluster yang lain. Pengujian selanjutnya adalah hasil pengujian performa kluster untuk parameter Term-Weighting menggunakan Binary Term Occurrences dengan N-Grams adalah 1 ditampilkan pada Gambar 23.

Pengujian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-3,832	-1,73
3	-3,319	-1,247
4	-3,504	-0,914
5	-1,321	-1,094
6	-1,977	-0,896
7	-2,839	-1,2
8	-2,706	-1,143
9	-2,295	-1,115
10	-2,259	-1,143

Gambar 23. Hasil Pengujian Nilai DBI Binary Term Occurrences n-Grams 1

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 6 dengan menggunakan algoritma K-Medoids dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 0,896, yang paling kecil di antara DBI kluster yang lain. Pengujian selanjutnya adalah hasil pengujian performa kluster untuk parameter Term-Weighting menggunakan Binary Term Occurrences dengan N-Grams adalah 2 ditampilkan pada Gambar 24.

Penguian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-5,551	-0,827
3	-1,596	-0,938
4	-3,573	-0,924
5	-3,524	-0,908
6	-3,103	-0,909
7	-2,211	-0,898
8	-1,515	-0,93
9	-1,674	-0,887
10	-2,151	-0,889

Gambar 24. Hasil Penguian Nilai DBI Binary Term Occurrences n-Grams 2

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 2 dengan menggunakan algoritma K-Medoids dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 0,827, yang paling kecil di antara DBI kluster yang lain. Penguian selanjutnya adalah hasil penguian performa kluster untuk parameter Term-Weighting menggunakan Binary Term Occurrences dengan N-Grams adalah 3 ditampilkan pada Gambar 25.

Penguian Nilai DBI		
Jumlah K	K-Means	K-Medoids
2	-0,717	-0,818
3	-1,656	-0,938
4	-2,359	-0,915
5	-0,802	-0,917
6	-2,659	-0,909
7	-0,865	-0,904
8	-0,912	-0,896
9	-2,149	-0,91
10	-1,614	-0,91

Gambar 25. Hasil Penguian Nilai DBI Binary Term Occurrences n-Grams 3

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 5 dengan menggunakan algoritma K-Medoids dengan nilai DBI yang paling mendekati angka 0, yaitu: absolut 0,717, yang paling kecil di antara DBI kluster yang lain. Setelah seluruh percobaan sudah dilakukan, dilakukan perekapan seluruh percobaan, yang ditampilkan pada Gambar 26.

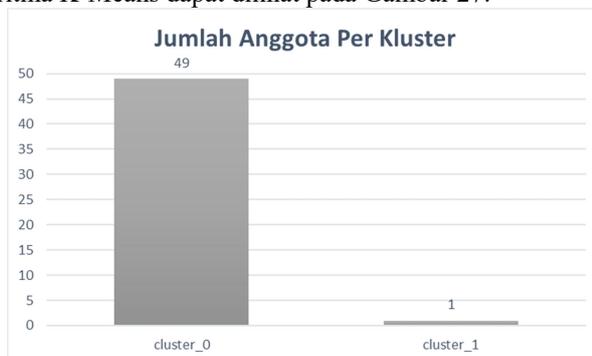
Rekapitulasi Penguian DBI				
Term Weighting	N-Grams	Jumlah K	Algoritma	Nilai DBI
TF-IDF	1	10	K-Medoids	-1,443
TF-IDF	2	10	K-Medoids	-1,476
TF-IDF	3	10	K-Medoids	-1,526
Term Frequency	1	10	K-Medoids	-1,446
Term Frequency	2	10	K-Medoids	-1,469
Term Frequency	3	10	K-Medoids	-1,499
Term Occurrences	1	6	K-Means	-1,338
Term Occurrences	2	2	K-Means	-0,426
Term Occurrences	3	2	K-Means	-0,437
Binary Term Occurrences	1	6	K-Medoids	-0,896
Binary Term Occurrences	2	2	K-Medoids	-0,827
Binary Term Occurrences	3	2	K-Means	-0,802

Gambar 26. Hasil Rekapitulasi Penguian DBI

Dari gambar di atas ditemukan bahwa kluster yang paling bagus adalah kluster berjumlah 2 dengan menggunakan algoritma K-Means dengan parameter Term Weighting menggunakan Term Occurrences serta penggunaan N-Grams 2 dengan nilai DBI yang paling baik yaitu -0,426, yang paling baik di antara DBI kluster yang lain. Dengan waktu komputasi menggunakan algoritma K-Means pada percobaan pertama adalah 3 detik, waktu komputasi pada percobaan kedua adalah 2 detik, dengan total rata-rata adalah 2,5 detik waktu komputasi. Dengan kondisi yang sama menggunakan Term Weighting yaitu Term Occurrences serta penggunaan N-Grams yang berjumlah 2 namun menggunakan algoritma K-Medoids nilai DBI yang muncul adalah -1,631 dengan waktu komputasi pada percobaan pertama adalah 274 detik, dan waktu komputasi pada percobaan kedua adalah 249 detik, dengan total rata-rata adalah 261,5 detik.

3.2.2 Informasi Kluster

Dari hasil pengujian menggunakan DBI ditemukan bahwa kluster berjumlah 2 adalah kluster yang paling baik dengan Term Weighting adalah Term Occurrences, N-Grams adalah 2 untuk mengelompokkan skripsi Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial dengan Program Studi Administrasi Negara. 2 kluster yang dihasilkan adalah kluster 0 dan kluster 1, Dengan kondisi pengelompokan jumlah pengelompokan skripsi menggunakan algoritma K-Means dapat dilihat pada Gambar 27.



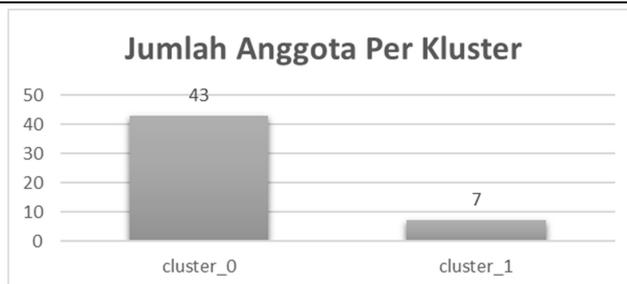
Gambar 27. Gambar Hasil Klasterisasi Menggunakan Algoritma K-Means

Pada gambar di atas dengan menggunakan algoritma K-Means jumlah kluster 0 adalah 49 items skripsi, sedangkan kluster 1 adalah 1 items skripsi. Dari distribusi data tersebut dapat dibuatkan beberapa sample data dari hasil distribusi di atas, yang dapat dilihat pada Gambar 4.28

Judul Skripsi	Kluster
8. ANALISIS PELAKSANAAN PENGGANTIAN PEJABAT STRUKTURAL DI LINGKUNGAN PEMERINTAH DAERAH PROVINSI RIAU JELANG PILKADA TAHUN 2018	cluster_0
11. PERANAN CAMAT DALAM MENCEGAH PERILAKU MALADMINISTRASI BIROKRASI DI KANTOR CAMAT MINAS KABUPATEN SIAK	cluster_0
12. ANALISIS PELAKSANAAN PEMANFAATAN DANA DESA UNTUK PEMBANGUNAN DESA TANJUNG BERULAK KECAMATAN KAMPAR KABUPATEN KAMPAR	cluster_0
20. ANALISIS PENERIMAAN RETRIBUSI PARKIR DI KECAMATAN BANGKINANG KOTA KABUPATEN KAMPAR	cluster_0
38. PERAN KEPALA DESA DALAM PEMBANGUNAN DI DESA PULO LIMAN KECAMATAN DOLOK SIGOMPULON KABUPATEN PADANG LAWAS UTARA	cluster_0
30. ANALISIS POTENSI PAJAK SARANG BURUNG WALET DIKECAMATAN TEMBILAHAN KABUPATEN INDRAGIRI HILIR	cluster_1

Gambar 28. Gambar Sample data pemetaan kluster menggunakan Algoritma K-Means

Pada Gambar di atas dapat dilihat bahwa sample data dari cluster_0 adalah judul skripsi nomor 8, 11, 12, dan 20 seperti Judul Skripsi ANALISIS PENERIMAAN RETRIBUSI PARKIR DI KECAMATAN BANGKINANG KOTA KABUPATEN KAMPAR atau PERAN KEPALA DESA DALAM PEMBANGUNAN DI DESA PULO LIMAN KECAMATAN DOLOK SIGOMPULON KABUPATEN PADANG LAWAS UTARA. Untuk sample data dari cluster_1 adalah judul skripsi nomor 30 yaitu ANALISIS POTENSI PAJAK SARANG BURUNG WALET DIKECAMATAN TEMBILAHAN KABUPATEN INDRAGIRI HILIR. Selain itu dilakukan ujicoba dengan menggunakan Term Weighting adalah Term Occurrences, N-Grams adalah 2 namun menggunakan algoritma K-Medoids, dengan kondisi jumlah pengelompokan skripsi yang dapat dilihat pada Gambar 29.



Gambar 29. Gambar Hasil Klasterisasi Menggunakan Algoritma K-Means

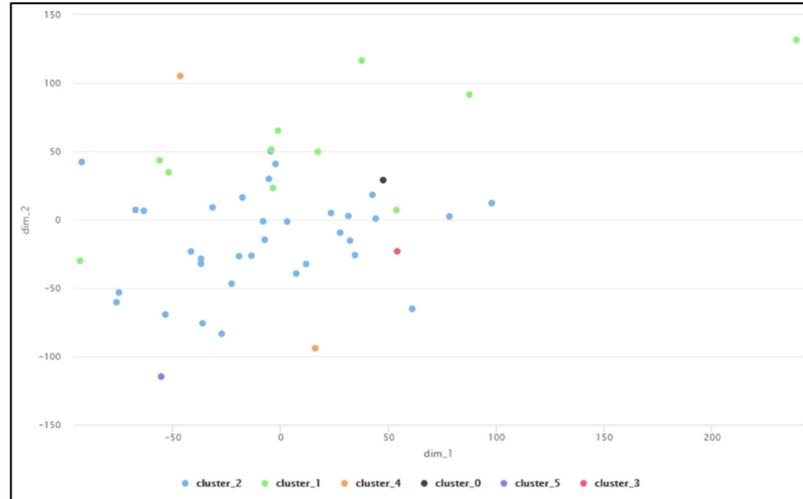
Pada gambar di atas dengan menggunakan algoritma K-Medoids jumlah kluster 0 adalah 43 items skripsi, sedangkan kluster 1 adalah 7 items skripsi. Dari distribusi data tersebut dapat dibuatkan beberapa sample data dari hasil distribusi di atas, yang dapat dilihat pada Gambar 30.

Judul Skripsi	Kluster
1. ANALISIS PENGEMBANGAN KAPASITAS BIDANG PENGELOLAAN SAMPAH DINAS LINGKUNGAN HIDUP DAN KEBERSIHAN KOTA PEKANBARU	cluster_0
3. ANALISIS PERANAN KEPALA DESA DALAM PEMBANGUNAN INFRASTRUKTUR “DESA PAKNING ASAL” KABUPATEN BENGKALIS	cluster_0
8. ANALISIS PELAKSANAAN PENGGANTIAN PEJABAT STRUKTURAL DI LINGKUNGAN PEMERINTAH DAERAH PROVINSI RIAU JELANG PILKADA TAHUN 2018	cluster_0
11. PERANAN CAMAT DALAM MENCEGAH PERILAKU MALADMINISTRASI BIROKRASI DI KANTOR CAMAT MINAS KABUPATEN SIAK	cluster_0
20. ANALISIS PENERIMAAN RETRIBUSI PARKIR DI KECAMATAN BANGKINANG KOTA KABUPATEN KAMPAR	cluster_0
22. Implementasi Retribusi Pelayanan Persampahan dan Kebersihan di Kabupaten kampar	cluster_0
30. ANALISIS POTENSI PAJAK SARANG BURUNG WALET DIKECAMATAN TEMBLAHAN KABUPATEN INDRAGIRI HILIR	cluster_0
2. ANALISIS PELAKSANAAN FUNGSI BADAN PERMUSYAWARATAN DESA (BPD) DALAM PENYELENGGARAAN PEMERINTAH DESA DI DESA SEI KUNING KECAMATAN TANDUN KABUPATEN ROKAN HULU	cluster_1
5. ANALISIS PELAKSANAAN PROGRAM PEMBANGUNAN DESA DI DESA KELESA KECAMATAN SEBERIDA KABUPATEN INDRAGIRI HULU	cluster_1
12. ANALISIS PELAKSANAAN PEMANFAATAN DANA DESA UNTUK PEMBANGUNAN DESA TANJUNG BERULAK KECAMATAN KAMPAR KABUPATEN KAMPAR	cluster_1
13. ANALISIS KINERJA BADAN USAHA MILIK DESA (BUMDES) ARTHA MITRA KESUMA DALAM MENYEDIAKAN PELAYANAN KEPADA MASYARAKAT DESA BENCAH KESUMA KECAMATAN KABUN KABUPATEN ROKAN HULU	cluster_1
38. PERAN KEPALA DESA DALAM PEMBANGUNAN DI DESA PULO LIMAN KECAMATAN DOLOK SIGOMPULON KABUPATEN PADANG LAWAS UTARA	cluster_1
46. ANALISIS PENGELOLAAN DESA BERDASARKAN UNDANG–UNDANG NOMOR 6 TAHUN 2014 (STUDI KASUS DESA LIMAU MANIS KECAMATAN KAMPAR KABUPATEN KAMPAR)	cluster_1
47. ANALISIS PERAN BADAN USAHA MILIK DESA (BUMDes) MAJU BERSAMA DALAM PENGEMBANGAN EKONOMI MASYARAKAT DI DESA PEBAUN HILIR KECAMATAN KUANTAN MUDIK KABUPATEN KUANTAN SINGINGI	cluster_1

Gambar 30. Gambar Sample data pemetaan kluster menggunakan Algoritma K-Medoids

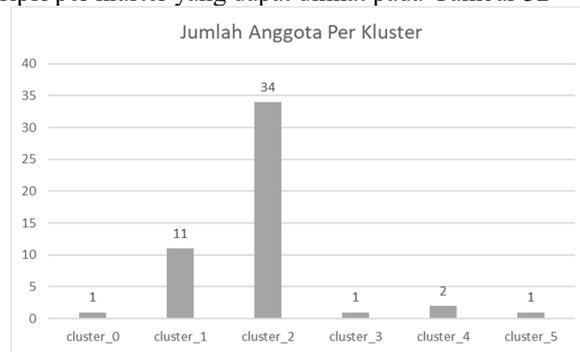
Pada Gambar di atas dapat dilihat bahwa sample data dari cluster_0 adalah judul skripsi nomor 1, 3, 8, 11, 20, 22, dan 30 seperti Judul Skripsi ANALISIS PENGEMBANGAN KAPASITAS BIDANG PENGELOLAAN SAMPAH DINAS LINGKUNGAN HIDUP DAN KEBERSIHAN KOTA PEKANBARU atau ANALISIS PERANAN KEPALA DESA DALAM PEMBANGUNAN INFRASTRUKTUR “DESA PAKNING ASAL” KABUPATEN BENGKALIS. Untuk sample data dari cluster_1 adalah judul skripsi dengan nomor 2, 5, 12, 13, 38, 46, dan 47 yaitu dengan contoh seperti judul skripsi ANALISIS PELAKSANAAN FUNGSI BADAN PERMUSYAWARATAN DESA (BPD) DALAM PENYELENGGARAAN PEMERINTAH DESA DI DESA SEI KUNING KECAMATAN TANDUN KABUPATEN ROKAN HULU atau ANALISIS PELAKSANAAN PROGRAM PEMBANGUNAN DESA DI DESA KELESA KECAMATAN SEBERIDA KABUPATEN INDRAGIRI HULU, atau ANALISIS PELAKSANAAN PEMANFAATAN DANA DESA UNTUK PEMBANGUNAN DESA TANJUNG BERULAK KECAMATAN KAMPAR KABUPATEN KAMPAR.

Kondisi parameter yang digunakan adalah dengan parameter Term Weighting: Term Occurrences, kemudian dengan N-Grams: 2. Dengan parameter tersebut terdapat opsi DBI dengan jumlah klaster yang secara visualisasi lebih baik, yaitu dengan jumlah klaster 6 dengan nilai DBI -1,110, menggunakan algoritma K-Medoids, hasil visualisasi dibuat dengan menggunakan t-SNE, dan dapat dilihat pada Gambar 31



Gambar 31. Visualisasi Penyebaran Cluster dengan t-SNE

Pada tampilan visualisasi penyebaran cluster yang terdapat pada Gambar 31, dapat ditampilkan distribusi jumlah skripsi per kluster yang dapat dilihat pada Gambar 32



Gambar 32. Distribusi Jumlah Anggota Per Kluster

Pada tampilan visualisasi penyebaran cluster yang terdapat pada Gambar 31, dapat ditampilkan contoh data skripsi yang merupakan hasil dari distribusi per kluster, yang dapat dilihat pada Gambar 33

Judul Skripsi	Kluster	Penjelasan Kluster
19. KOORDINASI BADAN PENANGGULANGAN BENCANA DAERAH (BPBD) DAN MASYARAKAT DALAM PENANGGULANGAN ABRASI DI KECAMATAN TEBING TINGGI BARAT KABUPATEN KEPULAUAN MERANTI	cluster_0	Skripsi dengan topik administrasi negara terkait penanggulangan bencana
2. ANALISIS PELAKSANAAN FUNGSI BADAN PERMUSYAWARATAN DESA (BPD) DALAM PENYELENGGARAAN PEMERINTAH DESA DI DESA SEI KUNING KECAMATAN TANDUN KABUPATEN ROKAN HULU	cluster_1	Skripsi dengan topik administrasi negara terkait pembangunan desa
3. ANALISIS PERANAN KEPALA DESA DALAM PEMBANGUNAN INFRASTRUKTUR "DESA PAKNING ASAL" KABUPATEN BENGKALIS	cluster_1	Skripsi dengan topik administrasi negara terkait pembangunan desa
5. ANALISIS PELAKSANAAN PROGRAM PEMBANGUNAN DESA DI DESA KELESA KECAMATAN SEBERIDA KABUPATEN INDRAGIRI HULU	cluster_1	Skripsi dengan topik administrasi negara terkait pembangunan desa
1. ANALISIS PENGEMBANGAN KAPASITAS BIDANG PENGELOLAAN SAMPAH DINAS LINGKUNGAN HIDUP DAN KEBERSIHAN KOTA PEKANBARU	cluster_2	Skripsi dengan topik administrasi negara terkait penanganan sampah dan program hijau
4. PENERAPAN PROGRAM KAMPUNG HIJAU DI KABUPATEN SIAK	cluster_2	Skripsi dengan topik administrasi negara terkait penanganan sampah dan program hijau
6. STRATEGI DINAS PANGAN TANAMAN PANGAN DAN HORTIKULTURA PROVINSI RIAU DALAM MENJAGA KETAHANAN PANGAN PADA ERA PANDEMI COVID-19	cluster_2	Skripsi dengan topik administrasi negara terkait penanganan sampah dan program hijau
49. PENGARUH ANALISIS JABATAN TERHADAP KINERJA PEGAWAI PADA DINAS KEPEMUDAAN DAN OLAHRAGA KOTA PEKANBARU	cluster_3	Skripsi dengan topik administrasi negara terkait kinerja pegawai
13. ANALISIS KINERJA BADAN USAHA MILIK DESA (BUMDES) ARTHA MITRA KESUMA DALAM MENYEDIAKAN PELAYANAN KEPADA MASYARAKAT DESA BENCAH KESUMA KECAMATAN KABUN KABUPATEN ROKAN HULU	cluster_4	Skripsi dengan topik administrasi negara terkait Bumdes
47. ANALISIS PERAN BADAN USAHA MILIK DESA (BUMDes) MAJU BERSAMA DALAM PENGEMBANGAN EKONOMI MASYARAKAT DI DESA PEBAUN HILIR KECAMATAN KUANTAN MUDIK KABUPATEN KUANTAN SINGINGI	cluster_4	Skripsi dengan topik administrasi negara terkait Bumdes
30. ANALISIS POTENSI PAJAK SARANG BURUNG WALET DIKECAMATAN TEMBILAHAN KABUPATEN INDRAGIRI HILIR	cluster_5	Skripsi dengan topik administrasi negara terkait isu khusus seperti pajak sarang burung walet

Gambar 33. Data Hasil Distribusi

Pada hasil di Gambar 4.27 dapat diberikan nama per masing-masing kluster sesuai dengan topik penelitian yang digunakan dengan penamaan sebagai berikut:

- Cluster_0 = Topik administrasi negara terkait penanggulangan bencana
- Cluster_1 = Topik administrasi negara terkait pembangunan desa
- Cluster_2 = Topik administrasi negara terkait dengan penanganan sampah dan program hijau
- Cluster_3 = Topik administrasi negara terkait perbaikan kinerja pegawai
- Cluster_4 = Topik administrasi negara terkait Badan Usaha Milik Desa (Bumdes)
- Cluster_5 = Topik administrasi negara terkait dengan isu khusus seperti pajak sarang burung walet

3.2. Implikasi Penelitian

Dari hasil penelitian, implikasi kebijakan yang akan digunakan oleh Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Ekonomi dan Sosial adalah dapat mengelompokkan 50 skripsi terbaru di tahun 2022 untuk Program Studi Administrasi Negara menjadi 2 kluster dengan jumlah kluster pertama adalah 43 skripsi, dan jumlah kluster kedua adalah 7 skripsi atau dapat opsi yang lain dengan kluster pertama berjumlah 49 skripsi, dan jumlah kluster kedua adalah 1 skripsi.

Selain itu secara visualisasi bisa menerapkan 6 kluster dengan parameter yang sama, bisa dibagi menjadi topik terkait penanggulangan bencana, topik terkait pembangunan desa, topik terkait dengan penanganan sampah dan program hijau, topik terkait perbaikan kinerja pegawai, topik terkait Badan Usaha Milik Desa (Bumdes), dan topik khusus semisal terkait dengan pajak sarang burung walet.

Untuk implikasi teoritis yang dirujuk kepada pengembangan teoritis dalam menghasilkan pola prediksi yang lebih baik, dapat menggunakan pengecekan algoritma yang lain yang sering digunakan di proses clustering seperti X-Means dan DBSCAN, selain itu parameter pengukuran bisa menggunakan parameter lain yaitu silhouette coefficient sebagai opsi lain selain Davies Bouldin Index (DBI).

4. KESIMPULAN DAN SARAN

4.1. Kesimpulan

Kesimpulan yang didapat setelah penelitian pada kasus pengelompokan dokumen skripsi selesai dilakukan, terdapat 2 poin yang dapat menjadi acuan perbandingan yaitu pengujian nilai Davies Bouldin Index (DBI) 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-Means memiliki nilai DBI yang lebih mendekati 0 yaitu -0,426, yang diuji dengan kondisi Term Weighting menggunakan Term Occurrences dan juga menggunakan N-Grams adalah 2. Sementara rentang nilai DBI yang dimiliki oleh K-Medoids dengan kondisi yang sama memiliki rentang yang cukup jauh dari angka 0 yaitu -1,631.

Melihat kedua algoritma pada pengujian DBI, maka dapat disimpulkan bahwa dalam pengujian pada penelitian ini, K-Means memiliki keunggulan dalam pengelompokan dokumen. Serta dalam pengujian waktu komputasi yang dibutuhkan, untuk mengelompokkan 50 dokumen skripsi, algoritma K-Means memiliki waktu rata-rata sebesar 2,5 detik sementara pada algoritma K-Medoids memiliki waktu rata-rata sebesar 261,5 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. Dari hasil visualisasi pengelompokan kluster, terdapat opsi untuk menggunakan jumlah kluster 6 dengan menggunakan parameter yang sama, dengan hasil DBI adalah -1,110 namun hasil visualisasi penyebaran klusternya lebih baik, dengan pembagian 6 kluster yaitu topik terkait penanggulangan bencana, pembangunan desa, dengan penanganan sampah dan program hijau, perbaikan kinerja pegawai, Badan Usaha Milik Desa (Bumdes), dan terkait topik khusus semisal terkait dengan pajak sarang burung walet.

4.2. Saran

Proses penelitian ini melakukan analisis 2 algoritma yang sering digunakan di proses clustering yaitu K-Means dan K-Medoids, analisis pengukuran yang dilakukan adalah menggunakan pengamatan performa dengan parameter Davies Bouldin Index (DBI) dan juga pengamatan waktu

komputasi prosesnya, ada beberapa hal yang dapat dianalisis lebih lanjut untuk meningkatkan kegunaan dari insight/pengetahuan klaster yang dihasilkan, yaitu:

1. Untuk penelitian selanjutnya dapat dikembangkan dengan menambahkan algoritma klusterisasi ke proses analisis yaitu algoritma X-Means dan juga DBSCAN untuk dapat menentukan performa dan juga waktu komputasi yang lebih baik lagi.
2. Pada penelitian selanjutnya dapat dikembangkan pengujian model penelitian dengan menggunakan parameter performa yang lain selain Davies Bouldin Index (DBI) yaitu silhouette coefficient.

DAFTAR PUSTAKA

- [1] S. Lialiyah and R. Andrea, "Penerapan Algoritma K-Medoids Clustering Dalam Pembentukan Zona Cluster Vaksin Booster," vol. 4, no. 1, pp. 124–129, 2022, doi: 10.47065/bits.v4i1.1617.
- [2] S. Ramadhani, D. Azzahra, U. I. Negeri, and S. S. Kasim, "Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student 's Thesis," vol. x, no. x, pp. 24–33, 2022.
- [3] F. Nur, M. Zarlis, and B. B. Nasution, "Penerapan Algoritma K-Means Pada Siswa Baru Sekolahmenengah Kejuruan Untuk Clustering Jurusan," *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 2, pp. 100–105, 2017, doi: 10.30743/infotekjar.v1i2.70.
- [4] M. Arifandi, A. Hermawan, and D. Avianto, "Implementasi Algoritma K-Medoids Untuk Clustering Wilayah Terinfeksi Kasus Covid19 Di Dki Jakarta," *JIT (Jurnal Teknol. Ter.)*, vol. 7, no. September, pp. 120–128, 2021.
- [5] B. Wira, A. E. Budianto, and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang," *RAINSTEK J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 53–68, 2019, doi: 10.21067/jtst.v1i3.3046.
- [6] Y. Elda, S. Defit, Y. Yunus, and R. Syaljumairi, "Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means," *J. Inf. dan Teknol.*, vol. 3, pp. 103–108, 2021, doi: 10.37034/jidt.v3i3.130.
- [7] C. Purnama, W. Witanti, and P. Nurul Sabrina, "Klasterisasi Penjualan Pakaian untuk Meningkatkan Strategi Penjualan Barang Menggunakan K-Means," *J. Inf. Technol.*, vol. 4, no. 1, pp. 35–38, 2022, doi: 10.47292/joint.v4i1.79.
- [8] N. Nurahman, A. Purwanto, and S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, 2022, doi: 10.30812/matrik.v21i2.1411.
- [9] S. U. Tarigan, M. Yetri, and Saniman, "Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat," *Jurnam Sist. Inf. TGD*, vol. 1, p. 14, 2022.
- [10] Noviyanto and P. Ekasari, "Algoritma K-Means Untuk Klasterisasi Jabatan Fungsional Dosen Pada Perguruan Tinggi Swasta Di Lingkungan LLDikti Wilayah III," *Paradigma*, vol. 24, no. 1, pp. 103–107, 2022.
- [11] A. Wahyu and Rushendra, "Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 1, pp. 175–179, 2022.
- [12] H. N. Putra, A. Wisandra, and Fransiska, "Penerapan Algoritma K-Means untuk Klasterisasi Data Obat Pasien Rawat Jalan Berdasarkan 3 Penyakit Terbanyak Di Rumah Sakit M. Natsir Solok," *Ensikolemediaku J.*, vol. 4, no. 3, pp. 304–312, 2022.
- [13] M. R. Nugroho, I. E. Hendrawan, and Purwantoro, "Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI," *J. Nuansa Inform.*, vol. 16, pp. 125–133, 2022.
- [14] S. A. Rahmah and J. Antares, "Klasterisasi Seleksi Mahasiswa Calon Penerima Beasiswa Yayasan Menggunakan K-Means Clustering," *I N F O R M a T I K a*, vol. 13, no. 2, p. 25, 2022, doi: 10.36723/juri.v13i2.282.
- [15] M. A. Hairudin, Y. Wabula, and Hazriani, "Rekomendasi Strategi Sosialisasi Program Studi Melalui Jalur Undangan Menggunakan Algoritma ID3 dan K-Means," *J. Inf. Technol. Comput. Eng.*, vol. 01, pp. 14–18, 2022.
- [16] R. Chairunnisa and P. P. Adikara, "Analisis Sentimen terhadap Karyawan Dirumahkan pada Media Sosial Twitter menggunakan Fitur N-Gram dan Pembobotan Augmented TF – IDF Probability dengan

-
- K-Nearest Neighbour,” vol. 6, no. 4, pp. 1960–1965, 2022.
- [17] A. T. Ni'mah and A. Z. Arifin, “Perbandingan Metode Term Weighting terhadap Hasil Klasifikasi Teks pada Dataset Terjemahan Kitab Hadis,” *Rekayasa*, vol. 13, no. 2, pp. 172–180, 2020, doi: 10.21107/rekayasa.v13i2.6412.