

KOMPARASI ALGORITMA CLUSTERING DATA MEDIA *ONLINE* PADA PROSES BISNIS

Iqbal Dzulfiqar Iskandar¹, Melisa Winda Pertiwi², Mira Kusmira³, Imam Amirulloh⁴

¹AMIK BSI Tasikmalaya, ^{2,3,4}STMik Nusa Mandiri Jakarta

Jl. Tanuwijaya No.4, Empangsari, Tawang, Tasikmalaya

E-mail : iqbal.iql@bsi.ac.id¹, melisa.mwp@bsi.ac.id², mira.mik@bsi.ac.id³, imam.iau@bsi.ac.id⁴

ABSTRAK

Media *online* saat ini telah dipergunakan untuk berbagai macam bidang karena sangat mudah dalam menjangkau informasi. Penelitian berfokus pada komparasi performa terhadap 5 algoritma clustering algoritma K-Means, K-Medoids, Self Organizing Maps, Fuzzy C-Means, dan Density-Based Spatial Clustering of Applications with Noise dengan metode komparasi Davies–Bouldin index Clustering pada dataset penggunaan media *online* pada proses bisnis, di ambil dari 94 perusahaan bisnis, baik itu bidang pendidikan, dinas pemerintahan, dan perusahaan swasta yang diambil di wilayah Tasikmalaya-Indonesia. Dengan menghasilkan data hasil pengolahan data Komparasi Davis-Bouldin Indeks SOM memiliki nilai rata-rata terkecil dari algoritma yang lainnya yaitu sebesar 1.073 yang artinya algoritma tersebut memiliki proses clustering dengan data kuantitatif yang lebih baik. Sedangkan DBSCAN dan K-Means memiliki nilai yang paling besar diantara keempat lagoritma cluster yang berarti memiliki proses clustering yang kurang baik. Sedangkan F-CM dan K-Medoids tidak memiliki nilai perbedaan yang signifikan bahkan hampir sama.

Kata kunci: *Clustering, Davies–Bouldin index, Media Online*

ABSTRACT

The online media is currently used for various areas because very easy to reach the information. The research focuses on the comparison of performance against 5 the algorithm clustering algorithm K-Means, , K-Medoids, Self Organizing Maps, Fuzzy C-Means, and Density-Based Spatial Clustering of Applications with Noise With the method comparison Davies-Bouldin index Clustering on the dataset using the media online on business process , In take from 94 companies business, good education in the district of the government and private companies that were taken in the region Tasikmalaya-Indonesia. To produce data processing results comparable data Davis-Bouldin SOM Index has the smallest average value from the other algorithm amounting 1.073 which means that the algorithm has the process of clustering with a better quantitative data. While DBSCAN and K-Means has the greatest value among the four lagoritma cluster which means have the process of clustering that less good. While the F-CM and K-Medoids does not have the value of significant differences even almost the same

Keywords: *Clustering, Davies–Bouldin index, Online Media*

1. PENDAHULUAN

Media *Online* merupakan layanan yang telah banyak dipergunakan pada berbagai macam bidang karena kemudahannya dalam menjangkau informasi. Contoh penggunaan media *online* di bidang pendidikan adanya program e-learning yang membantu peserta didik mendapatkan materi secara mudah kapanpun dan dimanapun, selain pendidikan media *online* dapat dipergunakan pada sarana hiburan, dan termasuk dalam proses bisnis. Saat ini banyak sekali perusahaan yang memanfaatkan media *online* dalam proses bisnisnya, tetapi apakah

perusahaan-perusahaan tersebut benar-benar mengandalkan media *online* sebagai perangkat utama dalam menjalankan proses bisnis, atau hanya digunakan sebagai media sekunder saja.

Berkaitan dengan media *online* telah banyak dilakukan penelitian dengan menggunakan pendekatan data mining. (Dimitrov, 2016) telah melakukan penelitian terhadap media *online* berupa objek website berjudul “*Visual Positions Of Links and Clicks on Wikipedia dan What Makes a Link Successful on Wikipedia*”. Penelitian tersebut melakukan identifikasi posisi link yang akan berpotensi lebih banyak di klik oleh pengunjung

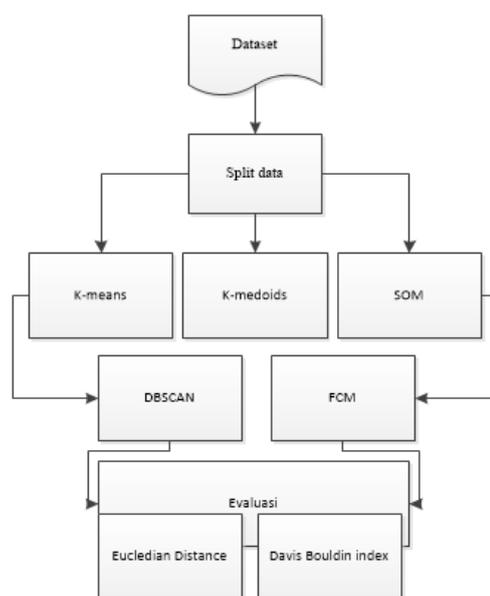
web, (Ren, 2016) dengan judul “*What Makes a Music Track Popular in Online Social Networks*”, melakukan mining terhadap data website yang berisikan isi konten untuk mengetahui apa yang menjadikan musik track seseorang menjadi populer. (Dholakia, F.2016). “*What makes commercial Web pages popular? An empirical investigation of Web page*” melakukan riset dengan pendekatan datamining untuk mengidentifikasi penyebab website komersial menjadi begitu populer. Penelitian yang telah dijadikan acuan masih terdapat kekurangan yaitu hanya berfokus pada media *online* konten website, serta tidak ada evaluasi terhadap algoritma datamining cluster. akan .Berkaitan dengan hal tersebut, akan dilakukan penelitian menggunakan metode algoritma clustering pada data Penggunaan media *online* pada proses bisnis, baik itu bidang pendidikan, dinas pemerintahan, dan perusahaan swasta yang diambil di wilayah Tasikmalaya-Indonesia, penelitian ini bertujuan untuk mengetahui perilaku serta mengelompokkan perusahaan yang memiliki karakteristik sama dalam penggunaan media *online* untuk proses bisnis berdasarkan data penggunaan Media *Online* yang diambil 94 perusahaan bisnis dan lembaga.

Selain melakukan clustering, terhadap data. Penelitian ini juga membahas komparasi performa terhadap 5 algoritma clustering dengan metode komparasi Davies–Bouldin index tentunya dengan sumber dataset yang sama. Saat ini telah banyak dilakukan penelitian yang berkaitan dengan komparasi algoritma clustering, tetapi hanya mencakup algoritma clustering yang sudah populer saja contohnya K-Means dan K-medoids, atau sekedar membahas tentang clustering tanpa komparasi algoritma terbaik. Lima Algoritma yang di komparasi pada penelitian ini adalah algoritma K-Means, K-Medoids, Self Organizing Maps, Fuzzy C-Means, dan Density-Based Spatial Clustering of Applications with Noise. Sehingga diharapkan memberikan informasi algoritma mana yang memiliki performa yang lebih baik diantara algoritma lainnya. Metode Clustering adalah proses untuk menemukan pengetahuan group pada data. Objek pada clustering tidak digunakan untuk prediksi sebuah target pada variable kelas. Tetapi hanya melakukan pengelompokan data secara alami (Vijay Kotu, Bala Deshpande, PhD.,2015). Struktur paper ini dimulai dari pengenalan ruang lingkup masalah setelah itu related work untuk menjabarkan keterkaitan penelitian yang di bahas dengan penelitian lain yang berhubungan, Theoretical Foundation menjabarkan landasan teori yang di gunakan, Proposed Method: Metode yang diusulkan, Experimental Results: Hasil eksperimen, Kesimpulan (Conclusion) Kesimpulan dan penelitian yang harus dilakukan setelahnya. Paper ini fokus membahas bagaimana mengkomparasikan

algoritma clustering pada data set penggunaan media *online* pada proses bisnis.

2. METODOLOGI

Metode yang di usulkan pada penelitian ini bertujuan untuk melakukan prosedur penelitian komparasi algoritma *clustering* terhadap data media *online* pada proses bisnis , sehingga dapat dilakukan komparasi lima algoritma *clustering* dengan metode evaluasi algoritma *Euclidian Distance*, sedangkan metode komparasi menggunakan algoritma *Davies-Bouldin index* (DBI). kerangka pemikiran yang ditujukan pada penelitian ini dapat dilihat pada gambar 4.



Gambar 1. Metodologi empiris yang diusulkan

Tahapan pertama untuk melakukan komparasi algoritma adalah menentukan objek data yang akan di olah, tahap kedua dilakukan pemisahan data otomatis training dan testing melalui operator split data. Tahap ketiga dilakukan proses ekstraksi data mining terhadap data set yang telah di siapkan sebelumnya dengan lima algoritma *clustering* *K-Means*, *K-Medoids*, *Self Organizing Maps*, *Fuzzy C-Means*, dan *Density-Based Spatial Clustering of Applications with Noise*. Tahap ke empat melakukan komparasi hasil clutering menggunakan algoritma Davies Bouldin index. Nilai terkecil dari hasil perhitungan Davies-Bouldin index mengartikan nilai untuk algoritma clustering yang lebih baik. (Vijay Kotu, Bala Deshpande, PhD.,2015).

3. LANDASAN TEORI

Penelitian tentang komparasi algoritma telah banyak dilakukan dan telah dipublikasikan. Sehingga dilakukan pengkajian terhadap penelitian

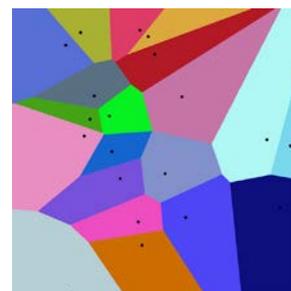
yang terkait sebelumnya agar dapat mengetahui metode yang digunakan, data seperti apa yang diproses, serta model yang dihasilkan. Penelitian yang dilakukan oleh (Mamta Gupta dan Anand Rajavat, 2014) dengan judul “*Comparision Of Algorithms For Document Clustering*”. Menghasilkan kesimpulan bahwa algoritma BIRCH memproses data clustering dokumen dengan efisiensi yang lebih bagus dibandingkan dengan K-means. Penelitian berikutnya oleh (Szil’agyi L. Enesi D, K. Levente dan M. Szil’agyi,2014) dengan judul “*Comparison of Various Improved-Partition Fuzzy c-Means Clustering, Algorithms in Fast Color Reduction*”. Mengkomparasikan algoritma clustering Fuzzy C-mans dengan algoritma Varoius improved-partition, dengan objek penelitian pengolahan pada warna citra, menghasilkan kesimpulan pengolahan data pada masalah color reduction lebih efisiensi menggunakan algoritma FCM. Penelitian yang dilakukan (R.Delshi Howsalya Devi, P.Deepika, 2015) “*Performance Comparison of Various Clustering Techniques for Diagnosis of Breast Cancer*”. Mengkomparasikan lima algoritma clustering terhadap objek penelitian data *Breast Cancer*. Menyimpulkan algoritma Farthest First memiliki akurasi yang tinggi dan tingkat efisiensi yang rendah terhadap data breast cancer dibandingkan ke empat algoritma yang lainnya yaitu DBSCAN, Canopy, LVQ, *Hierarchy*. Penelitian yang dilakukan oleh (H.Haripriya, R DeviSree, Dinesh.Pooja, Prema Nedungadi, 2015) dengan judul “*Comparative Performance Analysis of Self Organizing Maps on Weight Initializations using different Strategies*”. Melakukan komparasi algoritma SOM PCI, KPCI, RI, FCMI terhadap data *asli UCI BreastCancer, Blood Transfusion, Wine, Iris, Computer*, hasil penelitian menyimpulkan Strategi *initialization* berlaku bagi datasets setelah dimensionality reduksi dan sebelum dimensionality di lakukan pengurangan. . Maka dapat terlihat apakah PCA atau KPCA *Outperformed* untuk satu daset tertentu, sehingga dapat diidentifikasi apakah dataset bersifat linier atau bukan linear. Penelitian yang dilakukan oleh (Dyah Herawatie, Eto W., Purbandini, 2014) yang berjudul “Perbandingan Algoritma Pengelompokan *Non-Hierarki* untuk Dataset Dokumen”. Mengkomparasikan algoritma *clustering* K-Mean, Bi-secting,, K-Median, K-Medoid dengan objek pengolahan data set dokumen, penelitian tersebut menyimpulkan Algoritma pengelompokan yang memberikan hasil pengelompokan yang terbaik adalah K-Mean. Disamping itu Bi-secting K-Mean juga menghasilkan pengelompokan yang memuaskan. Sedangkan hasil pengelompokan Kmedoid memberikan hasil yang tidak memuaskan. Berdasarkan penelitian yang telah di kaji dari beberapa penelitian yang terkait. Maka dilakukan penelitian komparasi *performance* lima algoritma *clustering* dengan objek data penggunaan media

online pada proses bisnis. Algoritma yang di komparasikan K-Means, K-Medoids, Self Organizing Maps, Fuzzy C-Means, dan *Density-Based Spatial Clustering of Applications with Noise(DBSCAN)* Penelitian ini bertujuan mengelompokan data penggunaan Media Online pada proses bisnis serta mengidentifikasi algoritma yang memiliki performa terbaik terhadap pengolahan dataset dengan metode *Davies–Bouldin index* tabel diutamakan garis horizontal saja sedangkan garis vertikal dihilangkan.

Metode *Clustering* adalah proses untuk menemukan pengetahuan group pada data. Objek pada *clustering* tidak digunakan untuk prediksi sebuah target pada *variable* kelas. Tetapi hanya melakukan pengelompokan data secara alami (Vijay Kotu, Bala Deshpande,2015). Lima Algoritma *Clustering* yang di gunakan pada penelitian ini adalah algoritma *K-Means, K-Medoids, Self Organizing Maps, Fuzzy C-Means, dan Density-Based Spatial Clustering of Applications with Noise .*

a. K-Means

Clustering K-Means merupakan *prototype-based metode clustering* dimana kumpulan data set dibagi kedalam k *Cluster*. k-means adalah salah satu algoritma yang paling sederhana dan algoritma clustering yang paling umum digunakan. Pada teknik ini, pengguna menetapkan angka cluster (k) kebutuhan tersebut di kelompokan pada kumpulan data. Objektivitas clustering k-means adalah untuk menemukan satu *contoh asli poin* data pada masing-masing semua data poin cluster kemudian akan dimasukan kedalam *prototype* yang terdekat, kemudian terbentuk sebuah *cluster. prototype* di sebut sebagai *centroid*, merupakan pusat pada *cluster*.



Gambar 1. *Voronoi partition. under Creative Tahapan Algoritma K-means:*

a. *Initiate Centroids*

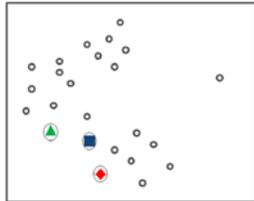
Tahapan pertama inialisasi nilai k secara random untuk centroid, nilai cluster di tetapkan oleh *user*.

b. *Assign Data Points*

Centroids telah memiliki inialisasi, tahap dua ini semua data point di tugaskan untuk nearest centroid untuk membentuk cluster. didalam konteks "nearest" adalah

kalkulasi dari pengukuran proximity. Pengukuran Euclidean distance secara umum mengukur proximity, kedua pengukuran tersebut seperti Manhattan measure dan Jaccard Coefficient yang dapat digunakan. Euclidean distance yang mana dua data points X (X_1, X_2, \dots, X_n) dan C (C_1, C_2, \dots, C_n) dengan atribut n di dapat dari rumus distance berikut :

$$d = \sqrt{(X_1 - C_1)^2 + (X_2 - C_2)^2 + \dots + (X_n - C_n)^2} \quad (1)$$



Gambar2. Initial Random Centroid

- c. *Step 3: Calculate New Centroids*
Mengkalkulasikan centroid baru, serta prototype pada cluster.
Dengan formula :

$$SSE = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (2)$$

Dimana C_i adalah i th cluster, j merupakan data points yang di beri dari cluster, μ_i adalah centroid untuk i th cluster, dan x_j adalah sebuah data objek yang spesifik. Centroid dengan minimal SSE untuk memberi cluster i adalah mean baru pada cluster. mean pada cluster dapat dikalkulasikan dengan formula

$$\mu_i = \frac{1}{j} \sum_{x \in c_i} X \quad (3)$$

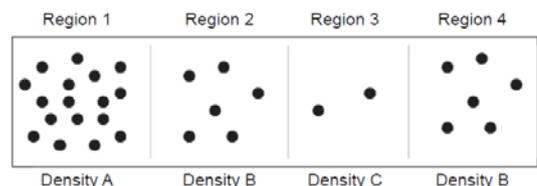
dimana X adalah data objek vektor (x_1, x_2, \dots, x_n).

- d. *Repeat Assignment and Calculate New Centroids*
Setelah centroid baru telah diidentifikasi, masukan data point ke nearest centroid ulangi hingga semua titik data yang dimasukan kembali ke pusat centroid baru. catat perubahan pada penugasan dari dua point data yang pantas masuk ke cluster yang berbeda pada langkah sebelumnya.
- e. *Termination*
Langkah 3, kalkulasikan centroid baru dan langkah 4 penugasan data point untuk centroid baru, hingga tidak ada perubahan selanjutnya pada data point yang terjadi, atau dengan kata lain tidak ada perubahan berpengaruh nyata pada centroid. centroid akhir menerangkan prototype data objek atau vektor dan mereka dipergunakan untuk menerangkan keseluruhan model cluster. Masing-masing data pada dataset

sekarang menghubungkan sebuah clustering ID baru, dimana atribut yang mengidentifikasi cluster

- b. *DBSCAN*
Cluster dapat juga didefinisikan sebagai satu area konsentrasi tinggi (atau density) dari objek data yang terkumpul oleh area dari konsentrasi rendah (atau density) dari objek data. Satu algoritma cluster density mengidentifikasi cluster pada data berlandaskan pengukuran dari distribusi kepadatan pada ruang n-dimensional. tidak sama dengan metode centroid, menetapkan angka dari parameter cluster (k), hal itu tidak cocok untuk algoritma density-based. Algoritma DBSCAN membuat cluster dari identifikasi high-density dan low-density space pada data set. seperti algoritma clustering pada K-means, menerima atribut numerik karena perhitungan distance masih tetap digunakan. (Vijay Kotu, Bala Deshpande, PhD.,2015).. Tahapan:

- Step 1: Defining Epsilon and MinPoints*
Step 2: Classification of Data Points
Step 3: Clustering
k-distribution chart for Iris data set with k = 4.



Gambar 3. Density dengan 4 wilayah

Sumber: Kotu predictive Analytic using Rapid Miner

- c. *Self Organizing Map*
Self Organizing map adalah algoritma Jaringan Syaraf Tiruan yang menggunakan topologi jaringan kohonen dengan cara memetakan korteks untuk mengelompokkan record kedalam suatu himpunan dengan vector visual dua dimensi, berdasarkan karakteristik atau fitur data yang di inputkan. Algoritma SOM mirip seperti pada centroid-based clustering tetapi dengan fondasi neural network. sejak SOM algoritma neural network model ini hanya menerima atribut numeric. SOM tidak memiliki target variable karena termasuk kedalam unsupervised learning Tahapan Algoritma SOM
- Inisialisasi bobot matriks w_{ij} . Set topologi tetangga. Set parameter learning rate.
 - Lakukan langkah 2 s/d 8 untuk

- kondisi batas berhenti belum tercapai
- Untuk setiap vektor input x lakukan langkah 3 sampai 5.
 - Untuk setiap j , hitung :

$$D(j) = \sum_i w_{ij} - Xi)^2 \quad (4)$$
 - Temukan index J yang membuat $D(J)$ minimum
 - Untuk setiap unit j pada tetangga J dan semua i hitung. Rumus yang digunakan:

$$W_{ij} = W_{ij}(\text{old}) + \alpha [X_i - W_{ij}(\text{old})] \quad (5)$$
 - Update learning rate nya
 - Kurangi jarak/radius dari topologi tetangga
 - test kondisi stop/Selesai.

Keterangan:

d_i = Jarak Euclidean

I_j = Input node

X : Vektor inputan pembelajaran (X_1, X_2, \dots, X_n)

α : learning rate

R : radius neighborhood

X_i : neuron / node input

Y_j : neuron / node output

W_{ij} : bias pada neuron output ke- j

M : Cluster

d. *Fuzzy C-Mean*

Fuzzy C-Means adalah teknik klusterisasi yang di tentukan oleh derajat dalam mengelompokan keanggotaannya.pertama kali menentukan pusat kluster, yang akan menandai lokasi rata-rata untuk tiap-tiap kluster. Pada kondisi awal, pusat kluster ini masih belum akurat. Sehingga dilakukan perbaikan pusat kluster dan derajat keanggotaan tiap-tiap titik data secara berulang. Maka terlihat pusat kluster akan bergerak menuju lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan kepusat cluster yang terbobot oleh derajat keanggotaan titik data. Output dari Fuzzy C-Means adalah deretan pusat kluster dan beberapa derajat kelompok untuk tiap-tiap titik data.

e. *K-Meoids*

K-medoid merupakan salah satu teknik penggerombolan yang mirip dengan *k-means*. Algoritma ini memiliki kemiripan dengan Algoritma *K-Means Clustering*, tetapi terdapat beberapa perbedaan utama, dimana apabila pada Algoritma *K-Means Clustering*, nilai tengah dihitung dengan rata-rata (*mean*) dan perhitungan jarak dihitung dari data pada masing-masing *mean*, sedangkan pada algoritma ini, data akan digunakan sebagai nilai tengah / disebut dengan medoid, dan

perhitungan jarak dihitung dari jarak antar masing-masing data. Tahapan algoritma *K-Medoid* sebagai berikut:

- Inisialisasi: memilih objek k secara acak yang akan berfungsi sebagai medoids.
- Asosiasikan setiap titik data dengan medoid yang paling serupa dengan menggunakan ukuran jarak dan menghitung biaya.
- Secara acak memilih objek k baru yang akan berfungsi sebagai medoid dan menyimpan salinan dari set asli.
- Gunakan set medoids baru untuk menghitung ulang biaya.
- Jika biaya yang baru lebih besar dari pada biaya lama kemudian hentikan algoritma tersebut.
- Ulangi langkah kedua hingga kelima sampai tidak ada perubahan dalam medoid.

f. *Euclidian Distance*

Euclidian Distance merupakan distance matrix dengan perhitungan menggunakan aturan pangkat dan akar kuadrat. Euclidian akan memberikan hasil jarak yang relatif kecil. Jarak antara Nilai Random/ Bobot dan data dihitung dengan menggunakan rumus *Euclidean Distance*.

$$d_{euclidian} = \sqrt{\sum_{i=1}^N (P_i - q_i)^2} \quad (6)$$

Keterangan :

$d_{euclidian}$: Jarak Euclidian Distance

p_i : Titik Awal

q_i : Titik Awal

N : Jumlah Data

g. *Davies-Bouldin index*.

Davies-Bouldin index (DBI) merupakan salah satu metode validasi cluster untuk evaluasi kuantitatif dari hasil clustering. (Davies & Bouldin, 1979). Evaluasi clustering berbeda dengan regresi dan algoritma klasifikasi karena algoritma cluster tidak mengetahui label untuk komparasi. "Nilai terkecil dari hasil perhitungan Davies-Bouldin index mengartikan nilai untuk algoritma clustering yang lebih baik. (Vijay Kotu, Bala Deshpande, PhD.,2015)". Kalkulasi algoritma *Davies Bouldin index* dapat dilakukan dengan rumus berikut:

$$var(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (7)$$

$$R_i = j = 1, \dots, k, i \neq j R^{ij}$$

$$R_{ij} = \frac{var(C_i) + var(C_j)}{\|C_i - C_j\|}$$

$$DB = \frac{1}{k} \sum_{i=1}^k R_i \quad (8)$$

Dimana

DB : validasi davies bouldin

Var : variance dari data

N : Banyaknya data

X : data ke-i

X : rata-rata dari tiap Cluster

R : jarak antar Cluster

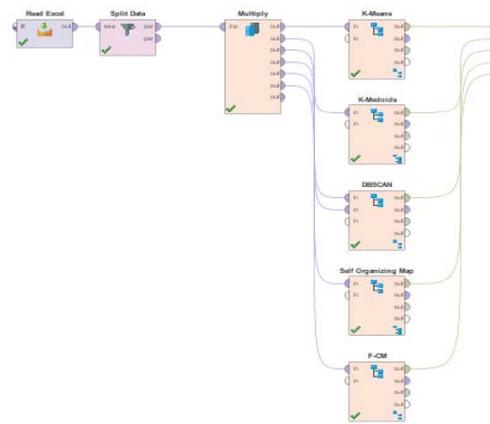
4. HASIL DAN PEMBAHASAN

Data set yang digunakan pada penelitian ini adalah data set penggunaan media online pada proses bisnis yang diambil dari 100 perusahaan di daerah Tasikmalaya-Indonesia. Eksperimen yang dilakukan untuk pemrosesan data mining komputer dengan platform Intel Core i5 1.80 Ghz dengan Ram 4GB, Sistem operasi menggunakan windows 10 64 bit, software rapid Miner 7.4. Komparasi yang

Nomor	Penggunaan Media Online	Jumlah data pengguna Media online dari 100 perusahaan
1	Transaksi Jual Beli	98
2	Pemasaran online	6
3	Kirim E-mail	98
4	Payroll	71
5	Monitoring User	63
6	Pendaftaran Anggota online	87
7	Prediksi Saham	9
8	Jasa Ekspedisi	15
9	Absensi	90
10	Penggunaan Server	88
11	Sosialisasi	14
12	Pengembangan Website	95
13	Portal berita	98
14	Download Data	85
15	Upload Data	99
16	Jasa Keuangan	98
17	Pendidikan	18
18	Pembayaran Jasa	57
19	Creditur	41
20	Jejaring Sosial	72
21	Penilaian Kinerja	94

dilakukan terhadap algoritma klastering adalah dengan menggunakan metode Davis-Bouldin Index, karena sangat berbeda dengan algoritma data mining Klasifikasi, Algoritma Cluster tidak dapat mengetahui label untuk di komparasi

Tabell. Data Penggunaan media online pada proses bisnis



Gambar 4. komparasi algoritma clustering

Tabel2. Hasil Proses Clustering

Perusahaan pengguna media online	Cluster
Netral net	Cluster_0
Honda	Cluster_0
Dhemi	Cluster_0
Sky komputer	Cluster_0
Dev design	Cluster_0
3g komputer	Cluster_0
Three six komputer	Cluster_0
Notaris dan ppat ida aryani sh	Cluster_0
Notaris dan ppat sh	Cluster_0
Notaris dan ppat	Cluster_0
Kecamatan cibereum	Cluster_0
Notaris dan ppat	Cluster_0
Dishub kominfo	Cluster_0
Kantor pelayanan perizinan terpadu	Cluster_0
Bpn ri kota tasikmalaya	Cluster_0
Kpkn tasikmalaya	Cluster_0
Kpkn tasikmalaya	Cluster_0
Cafa comunika	Cluster_0
Kppn tasikmalaya	Cluster_0
Cv sinar mas tsm (toyota)	Cluster_0
Kbpn tasikmalaya	Cluster_0
Kantor pelayanan perizinan terpadu	Cluster_0
Queen computer	Cluster_0
Dishub kominfo	Cluster_0
Kementrian agama	Cluster_0
Arba nova	Cluster_0
Stimik dci	Cluster_0
Ide komunika	Cluster_0
Notaris dan ppat ida aryani sh	Cluster_0
Badan pertahanan nasional	Cluster_0
Bpjs kesehatan	Cluster_0
Kandaga net	Cluster_0
Setda kota tasikmalaya	Cluster_0
Kantor pelayanan perizinan terpadu	Cluster_0
Culiner	Cluster_0
Lion net	Cluster_0
King net	Cluster_0
Fawaz net	Cluster_0
Smk angkasa tasikmalaya	Cluster_0
SMA Tasikmalaya	Cluster_0
Rs. Tasik medika citratama	Cluster_0
Pt. Pln tasikmalaya	Cluster_0
Kedai x treme	Cluster_0
Rc technology	Cluster_0
Net 73	Cluster_0
Rusdi net	Cluster_0

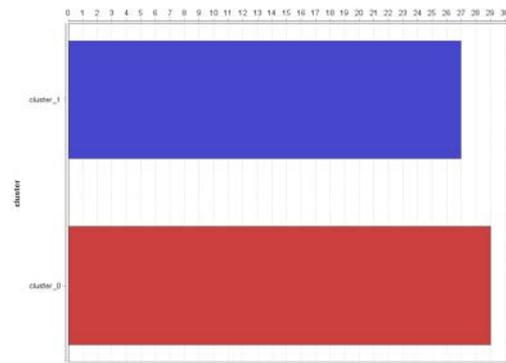
Kantor pelayanan pajak	Cluster_0
Notaris dan ppat	Cluster_0

Tabel 2 menunjukkan Perusahaan yang memiliki karakteristik sama berjumlah 50 perusahaan, sehingga di kelompokkan kedalam klastering 0.

Tabel 3. Clustering 1

Dirgantara pilot school	Cluster_1
Bill komputer	Cluster_1
Kpp pratama tasikmalaya	Cluster_1
Yayasan sumber barokah	Cluster_1
Mometo	Cluster_1
Disduk capil	Cluster_1
Kantor kelurahan tasikmalaya	Cluster_1
Pos indonesia	Cluster_1
Kpkn tasikmalaya	Cluster_1
Kppn tasikmalaya	Cluster_1
Kantor pajak tasikmalaya	Cluster_1
Toko sa	Cluster_1
Kantor kelurahan cikalang	Cluster_1
Bpjs kesehatan	Cluster_1
Kantor kecamatan cibereum	Cluster_1
All komputer	Cluster_1
Dinas kependudukan dan pencatatan sipil	Cluster_1
Bpjs kesehatan	Cluster_1
Setda kota tasikmalaya	Cluster_1
Kementrian agama	Cluster_1
Kementrian agama	Cluster_1
Bpn ri kota tasikmalaya	Cluster_1
Kpp pratama tasikmalaya	Cluster_1
Jhon cell	Cluster_1
Technozone computer	Cluster_1
Cv kawilang	Cluster_1
Pt kai	Cluster_1
Sukmara	Cluster_1
Jne	Cluster_1
Bpjs kesehatan	Cluster_1
Tiki cibereum	Cluster_1
Jne siliwangi	Cluster_1
Lp3i tasikmalaya	Cluster_1
Mukena desy collection	Cluster_1
Tdc tasikmalaya	Cluster_1
Theater cinema 21	Cluster_1
Indosat	Cluster_1
Telkomsel	Cluster_1
Rp media	Cluster_1
Kantor pajak pratama i	Cluster_1
Kantor pajak pratama ii	Cluster_1
Kantor pembendaharaan negara	Cluster_1

Tabel 3 menunjukkan Jumlah perusahaan yang masuk kedalam Cluster 1 sebanyak 44 perusahaan yang memiliki karakteristik sama.



Gambar 5. Diagram Bar

Diagram Bar pada gambar 6 mengartikan jumlah cluster 0 dan 1, Cluster 0 memiliki persentase lebih tinggi dari cluster 1.

Hasil komparasi lima algoritma data mining *K-Means*, *K-Medoids*, *Self Organizing Maps*, *Fuzzy C-Means*, dan *Density-Based Spatial Clustering of Applications with Noise* akan di

jabarkan pada tabel 4 dan 5.

Tabel 4 dan 5 adalah nilai rata-rata hasil evaluasi setiap cluster menggunakan metode *Centroid Distance* dari 5 algoritma clustering. *Algoritma clustering* DBSCAN memiliki nilai rata-rata 1.610 pada cluster0 dan *K-Means* 1.697 pada cluster1, sehingga dapat dikatakan cluster yang terdapat pada algoritma *K-means* dan *SOM* memiliki performa yang baik.

Tabel4. Hasil Komparasi cluster dengan *centroid distance*

<i>K-means</i>		<i>K-Medoids</i>	
C_0	C_1	C_0	C_1
2.8	1.6	3.5	1.8
12	97	48	96

Tabel5. Hasil komparasi

<i>SOM</i>		<i>DBSCAN</i>		<i>F-CM</i>	
C_0	C_1	C_0	C_1	C_0	C_1
90.	82.	1.	2.	2.	2.
770	513	610	111	028	222

Tabel4. Dan Tabel 5 adalah nilai rata-rata hasil evaluasi setiap cluster menggunakan metode *Centroid Distance* dari 5 algoritma clustering. *Algoritma clustering* DBSCAN memiliki nilai rata-rata 1.610 pada cluster0 dan *K-Means* 1.697 pada cluster1, sehingga dapat dikatakan cluster yang terdapat pada algoritma *K-means* dan *SOM* memiliki performa yang baik.

Tabel6. Nilai rata-rata hasil komparasi *Centroid Distance*.

K-means	K-Medoids	SOM	DBSCAN	F-CM
-1.969	-2.667	-86.917	-2.034	-1.911

Tabel6 adalah nilai rata-rata hasil evaluasi dengan algoritma centroid distance dengan menghitung nilai *cluster* keseluruhan pada masing-masing algoritma. FCM memiliki performa yang baik, karena memiliki nilai terendah diantara algoritma lainnya.

Tabel7. Komparasi Davis-Bouldin

K-means	K-Medoids	SOM	DBSCAN	F-CM
-2.382	-1.483	-1.073	-2.202	-1.128

Hasil pengolahan data Komparasi *Davis-Bouldin Indeks* SOM memiliki nilai rata-rata terkecil dari algoritma yang lainnya yaitu sebesar -1.073 yang artinya algoritma tersebut memiliki performa yang lebih baik. Sedangkan DBSCAN dan K-Means memiliki nilai yang paling besar diantara ke empat algoritma cluster yang berarti DBSCAN dan K-means memiliki proses *clustering* yang kurang baik. Sedangkan K-Medoids dan F-CM tidak memiliki nilai perbedaan yang signifikan.

5. KESIMPULAN

Komparasi lima Algoritma *Clustering* menggunakan metode *Davis-Bouldin Index* menggunakan dataset Pengguna data *media online* pada proses bisnis yang diambil dari 94 perusahaan yang ada di wilayah tasikmalaya dapat dilakukan. Dari 94 Perusahaan telah di ketahui 50 perusahaan yang memiliki karakteristik sama dan masuk kedalam *Cluster 0*, dan 44 perusahaan masuk kedalam *cluster 1*. Hasil evaluasi yang di dapatkan dari pengolahan data Komparasi *Davis-Bouldin Indeks* SOM memiliki nilai rata-rata terkecil dari algoritma yang lainnya yaitu sebesar 1.073 yang artinya algoritma tersebut memiliki proses *clustering* dengan data kuantitatif yang lebih baik. Sedangkan DBSCAN dan K-Means memiliki nilai yang paling besar diantara keempat algoritma cluster yang berarti DBSCAN dan K-means memiliki proses *clustering* yang kurang baik. Sedangkan K-Medoids dan F-CM tidak memiliki nilai perbedaan yang signifikan bahkan hampir sama. Ini menyimpulkan bahwa dataset tidak mempengaruhi kualitas pada algoritma *clustering*. Saran untuk penelitian berikutnya diharapkan dapat mengkomparasi algoritma *clustering* menggunakan metode komparasi dengan pola untuk mengetahui akurasi terhadap proses pengolahan data *clustering*. Serta melakukan validasi dengan metode *External Evaluation*.

Saran untuk penelitian dimasa mendatang, dapat dilakukan komparasi terhadap algoritma *clustering*

dengan dataset penggunaan media *online* yang lebih luas.

DAFTAR PUSTAKA

Prosiding

- Dholakia, F. (2016). *What makes commercial Web pages popular? An empirical investigation of Web page*.
- Dimitrov, D., Lemmerich, F., & Strohmaier, M. (2016). *Visual Positions of Links and Clicks on Wikipedia*, 3–4. <https://doi.org/10.6084/m9.figshare.1305770>.
- Delshi, R., D., P.Deepika.,(2015). *Performance Comparison of Various Clustering Techniques for Diagnosis of Breast Cancer. International Conference on Computational Intelligence and Computing Research*. 978-1-4799-7849-6.
- Haripriya H., DeviSree R., Pooja D., Nedungadi.(2015). *Fifth International Conference on Advances in Computing and Communication*.
- Herawatie D, Wuryanto E..(2015) Perbandingan Algoritma Pengelompokan Non-Hierarki untuk Dataset Dokumen. Seminar Nasional Aplikasi Teknologi Informasi (SNATI).
- Kotu V.,Deshpande B.(2015). *Predictive analytic Using Rapid Miner*. Elsevier Inc. ISBN: 978-0-12-801460-8
- L'aszl, SA., Gell'ert D., Levente K., Andor S., M. Agyi.(2014). *Comparison of Various Improved-Partition Fuzzy c-Means Clustering Algorithms in Fast Color Reduction. International Symposium on Intelligent Systems and Informatics.Serbia*.
- Mamta, G., Anand R.(2014). Comparison Of Algorithms For Document Clustering. *Sixth International Conference on Computational Intelligence and Communication Networks*.
- Rahmawati L., Cahyani A.D., Putro S.2014. Pemanfaatan metode cluster som – idb sebagai analisa Pengelompokan penerima beasiswa.
- Ren, J. (2016). *What Makes a Music Track Popular in Online Social Networks ?*, 95–96