JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 2, Juni 2024, Pp. 878-888

ANALISIS PERILAKU PENGGUNA INTERNET DENGAN METODE K-MEANS CLUSTERING DAN PENDEKATAN DAVIES BOULDIN INDEX MENGGUNAKAN DATA LOG UNIVERSITAS XYZ

Angga Zakharia Wijaya*[1)], Irwan Sembiring[2)]
1. Universitas Kristen Satya Wacana, Indonesia
2. Universitas Kristen Satya Wacana, Indonesia

Article Info
Kata Kunci: Algoritma *K Means*; Log Akses; Pengelompokan; *Davies Bouldin Index;*

Keywords: *Algoritma K-Means; Access Log; Clustering; Davies Bouldin Index;*

Article history:
Received 12 March 2024
Revised 26 March 2024
Accepted 9 April 2024
Available online 1 June 2024

DOI :
https://doi.org/10.29100/jipi.v9i2.4750

* Corresponding author.
Angga Zakharia Wijaya
E-mail address:
672019244@student.uksw.edu

ABSTRAK
Aktivitas penggunaan jaringan Internet sangat berdampak pada penggunanya, perubahan perilaku menjadi penyebab dari penggunaan jaringan internet, Informasi yang dicari terkadang tidak sesuai dengan kebutuhan terhadap penggunaan jaringan Internet. Sehingga, situs website yang tidak memberikan manfaat, perlu diidentifikasi, dan aksesnya diblokir. Hal ini dilakukan bertujuan untuk meminimalisir dari penggunaan jaringan Internet yang menyimpang dari penggunaan nya. Sehingga dapat menunjang kinerja, baik dari bagian administrasi, maupun dalam proses pembelajaran. Penelitian ini memiliki tujuan untuk melakukan *clustering* data *Access Log* jaringan Internet di universitas xyz dengan menggunakan *Algoritma K-Means*, dan melakukan uji validasi hasil *clustering* berdasarkan *Davies Bouldin Index*. Hasil dari penelitian ini menunjukan bahwa perilaku penggunaan jaringan internet di universitas xyz masih menyimpang dari kebutuhan Informasi yang dicari. Dengan penggunaan *Algoritma K-Means Clustering* menghasilkan tingkat kualitas *cluster* yang baik, berdasarkan uji validasi data *Davies Bouldin Index*, yang mendapatkan nilai DBI 0,110369132, sehingga hasil dari *clustering* yang dilakukan sudah cukup baik. Dengan dilakukannya penelitian ini, diharapkan dapat memberikan gambaran terhadap pengelola jaringan Internet, berdasarkan metode *Algoritma K-Means Clustering.*

ABSTRACT
The activity of using the Internet network has a huge impact on its users, changes in behavior are the cause of using the internet network, the information sought is sometimes not in accordance with the needs of using the Internet network. Thus, websites that do not provide benefits need to be identified and access blocked. This is done to minimize the use of the Internet network that deviates from its use. So that it can support performance, both from the administration, and in the learning process. This research has a goal to cluster the data Access log Internet network at xyz university using the K-Means Algorithm, and test the validation of clustering results based on Davies Bouldin Index. The results of this study indicate that the behavior of Internet network usage at xyz university still deviates from the needs of the information sought. The use of the K-Means Clustering Algorithm produces a good level of cluster quality, based on the Davies Bouldin Index data validation test, which gets a DBI value of 0,110369132, so the results of the clustering done are good enough. By doing this research, it is expected to provide an overview of the Internet network manager, based on the K-Means Clustering Algorithm method.

I. PENDAHULUAN

KEMAJUAN dalam dunia teknologi informasi diharapkan dapat meningkatkan produktivitas dan menjadi media yang paling efektif untuk mencari dan menyebarkan informasi dengan cepat, dan untuk dapat meningkatkan produktivitas juga diperlukannya manajemen jaringan internet yang baik. Dalam hal ini, sebagai salah satu contoh penggunaan jaringan internet di institusi pendidikan, yang memanfaatkan teknologi internet sebagai salah satu sarana untuk mengirim dan menerima informasi. kendala yang sering terjadi adalah pemanfaatan dalam penggunaan jaringan internet yang tersedia, sering disalahgunakan oleh pengguna jaringan

internet untuk mengunjungi situs website yang tidak berguna bagi institusi tersebut. Situs website yang mengalihkan perhatian pengguna jaringan internet, dapat mengurangi produktivitas pembelajaran, dan dapat mengganggu kinerja institusi.

Berdasarkan permasalah tersebut, diperlukan untuk menganalisis pola penggunaan jaringan internet dalam menilai seberapa besar penggunaan data yang digunakan untuk mengakses situs-situs tersebut [1]. Sehingga dalam hal ini, peran *Access Log* sangatlah dibutuhkan, dimana *Access Log* dapat menunjukkan aktivitas yang diterima oleh suatu situs, serta menunjukkan halaman mana saja dalam situs yang diakses, maupun waktu dalam mengakses internet [2]. Dalam penelitian ini penulis ingin menggunakan metode *Data Mining,* yang mana *Data Mining* merupakan suatu proses yang dilakukan dengan cara mengidentifikasi data yang sangat besar yang disimpan dalam penyimpanan untuk menemukan suatu pola data [3]. *Algoritma K-Means* merupakan salah satu algoritma *Data Mining*, yang nantinya akan digunakan untuk mempermudah proses *Clustering* data yang didapat pada *Access Log* Universitas xyz. *Clustering* merupakan suatu pengelompokan data yang didasarkan hanya pada informasi yang ditemukan dalam data, yang menggambarkan objek tersebut dan hubungan diantaranya [4]. Salah satu pendekatan untuk melakukan *Clustering* yaitu dengan menggunakan *Algoritma K-Means*, dengan menggunakan *Algoritma K-Means* data dapat dikelompokan dengan mudah, *Algoritma K-Means* sendiri merupakan salah satu *Algoritma Clustering* yang populer, dan juga merupakan *Algoritma* tanpa pengawasan yang digunakan dalam pengelompokan [5]. Algoritma ini merupakan salah satu metode yang dapat digunakan dalam membantu mengelompokkan data, berdasarkan website yang sering dikunjungi oleh pengguna jaringan internet dan website yang memiliki kunjungan terbanyak.

Pada penelitian terdahulu yang berjudul "Pengelompokan Pengguna Internet Menggunakan Metode K-Means Pada Data Log Akses Server", menjelaskan bahwa pengelompokan pengguna internet dengan menggunakan *Algoritma K-Means* untuk proses *profiling* menunjukan hasil yang sesuai, karena hasil dari pengelompokan ini memiliki tingkat akurasi yang baik. Namun dalam pembahasan penelitian tersebut, hanya befokus pada hasil yang diperoleh dengan menggunakan aplikasi *Rapidminer*, tanpa menjelaskan proses perhitungan dengan metode *K-Means* [6].

Pada penelitian terdahulu yang berjudul "Perancangan Deteksi Anomali Traffic Untuk Investigasi Log Menggunakan Metode K-Means Clusters", menjelaskan mengenai penggunaan *Algoritma K-Means Clustering*, yang mana penggunaan metode Algoritma ini merupakan metode yang efektif dalam mendeteksi *Anomaly Traffic*, yang dapat menentukan pola data statistik tertinggi dan rendahnya Traffic suatu serangan [7].

Pada Penelitian terdahulu yang berjudul "Deteksi Intrusi Jaringan Dengan K-Means Clustering Pada Akses Log Dengan Teknik Pengolahan Big Data", penggunaan *Algoritma K-Means* dalam *Spark* memberikan hasil yang lebih baik dalam hal akurasi dan deteksi aktivitas normal dan anomali [8].

Pada penelitian selanjutnya yang berjudul "Analisis Perilaku Pengguna pada PT Antar Surya Jaya : Harian Surya Surabaya Menggunakan Metode K-Means Clustering", penerapan dari *Algoritma K-Means Clustering* terhadap website Harian Surya, membuktikan bahwa dalam klasterisasi tersebut dapat melakukan pengelompokan berdasarkan data log terkait pengunjung yang sedang mengakses website Harian Surya. Penelitian ini menghasilkan tiga pengelompokan yang terbagi menjadi high, moderate dan low berdasarkan tingkat persebarannya [9].

Berdasarkan hasil kesimpulan penelitian terdahulu, penelitian ini memiliki perbedaan pada konsep *Clustering Access Log*, yang mana dalam melakukan *Clustering Access Log*, diperlukannya uji validasi dalam sebuah hasil *cluster*, sehingga hasil dari *clustering* yang dilakukan bisa dapat diketahui, data yang telah di *clustering* apakah valid atau tidak. Dalam melakukan pengujian terhadap hasil *clustering*, penulis menggunakan pendekatan dengan metode *Davies Bouldin Index* yang digunakan untuk mengukur kinerja algoritma clustering setelah proses pengelompokan data selesai. Penelitian ini juga akan menjelaskan proses perhitungan dengan metode *K-Means Clustering.* Sehingga penelitian ini diharapkan dapat untuk mengetahui perilaku pengguna jaringan internet di institusi pendidikan, berdasarkan kategori website apa saja, dan banyak atau sedikitnya website yang dikunjungi, penelitian ini bisa menjadikan wawasan atau evaluasi terhadap pengelola jaringan internet terutama didalam bidang institusi pendidikan.

II. METODE PENELITIAN

Penelitian yang dilakukan, diselesaikan melalui tahapan penelitian yang difokuskan pada analisis pengelompokan data *Access Log* dengan menggunakan metode *Algoritma K-Means*, yang dapat diuraikan dengan tahapan sistematis seperti pada contoh Gambar 1.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 2, Juni 2024, Pp. 878-888

Gambar. 1. Diagram Alir Penelitian

A. *Studi Pustaka*

Tahap berikutnya adalah melakukan studi pustaka dengan merujuk pada berbagai macam sumber jurnal maupun karya ilmiah penelitian, yang telah ditemukan dalam proses penelitian literatur pada tahap identifikasi masalah sebelumnya. Penulis akan terlebih dahulu melakukan pengumpulan pengetahuan dan mempelajari jurnal ataupun karya ilmiah penelitian yang ditemukan, kemudian akan dilakukan proses tinjauan. Tinjauan ini terdiri dari perbandingan, perbedaan, kritik, sintesis, maupun ringkasan. Tujuan akhir dari tahap ini adalah untuk mencapai kesimpulan terbaik yang muncul dari tinjauan yang dilakukan terhadap jurnal maupun karya ilmiah penelitian.

B. *Pengumpulan Data*

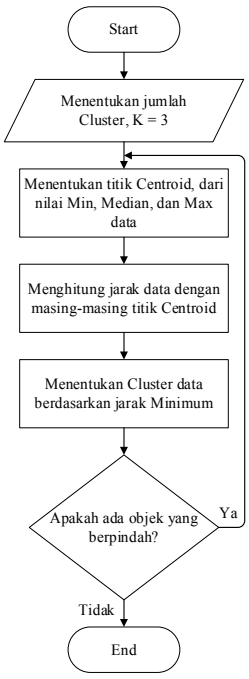
Metode pengumpulan data ini meliputi data yang dapat menunjang penelitian, yang berkaitan dengan informasi data *Access Log,* Data *Access Log* ini tersimpan dalam komputer yang berperan sebagai server web, dan digunakan untuk melacak serta menganalisis interaksi perilaku pengguna jaringan internet yang mengunjungi situs tersebut. *Access Log* berisi tentang informasi pengguna jaringan internet, seperti alamat *IP Address,* URL halaman yang diakses, Jumlah penggunaan *bandwidth*, *Time Stamp*, kode kesalahan, dan lain sebagainya [10].

C. *Clustering*

Kemudian pada tahapan Clustering peneliti akan melakukan pengelompokan data menggunakan metode *Algoritma K-Means Clustering*.

1) *Algoritma K-Means*

Algoritma K-Means merupakan salah satu teknik pengelompokan data non hierarki yang melakukan pengelompokan data ke dalam dua kelompok atau lebih. Metode ini mengorganisir data ke dalam beberapa kelompok sehingga data dengan karakteristik yang sama akan ditempatkan dalam satu kelompok, dan data yang memiliki karakteristik berbeda akan ditempatkan kedalam kelompok lain. Tujuan dari pengelompokan data ini adalah untuk meminimalkan fungsi objektif dari tujuan yang ditetapkan dalam proses pengelompokan [11].



Gambar. 2. Diagram Alir Algoritma K-Means

Langkah pertama dalam menggunakan *Algoritma K-Means* adalah memilih nilai acak untuk K, dalam artian menentukan banyaknya jumlah *Cluster* yang ingin dibentuk. Kemudian, nilai-nilai K tersebut diinisialisasikan berdasarkan tingkatan kunjungan tinggi, sedang, dan rendah, untuk menjadi pusat-pusat *Cluster* atau *Centroid*. Nilai K tersebut bisa dilakukan melalui metode rata-rata atau *Means.* Kemudian, jarak antara setiap data dengan masing-masing *Centroid* dihitung dengan menggunakan metode *Euclidean*, hingga ditemukan jarak yang paling dekat dari setiap data dengan *Centroid*. Kemudian data diklasifikasikan berdasarkan kedekatannya dengan *Centroid*. Selanjutnya, lakukan langkah-langkah tersebut hingga nilai *Centroid* tidak mengalami perubahan [12]. *Algoritma K-Means* memiliki rumus bentuk umum seperti pada rumus (1).

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^{p} |x_{2j} - x_{1j}|^2} \quad (1)$$

Berdasarkan pada rumus (1), diketahui bahwa *D* merupakan *Euclidean Distance,* kemudian *j* merupakan banyaknya objek yang akan dihitung, X_2 dan X_1 merupakan koordinat *Centroid* nya.

2) Clustering

Clustering merupakan proses pembagian data dalam satu kumpulan menjadi beberapa kelompok, dengan artian bahwa data dalam setiap kelompok memiliki kesamaan yang lebih signifikan, daripada data dalam satu kumpulan. Kelebihan dari teknik pengelompokan ini adalah untuk mengungkapkan struktur bawaan dalam data yang dapat dimanfaatkan dalam beragam aplikasi, seperti klasifikasi, manipulasi gambar, dan identifikasi pola [13]. Tujuan dilakukannya pengelompokan ini adalah untuk membagi keseluruhan data menjadi tiga tingkat *Cluster*, sebagai berikut.

Cluster 1, tingkat intensitas kunjungan tinggi, yaitu dengan mengambil nilai tertinggi atau nilai maksimum.

Cluster 2, tingkat intensitas kunjungan sedang, yaitu dengan mencari nilai rata rata atau nilai mean. Rumus (2) merupakan contoh rumus untuk mencari nilai rata-rata atau *mean*.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad (2)$$

Cluster 3, tingkat intensitas kunjungan rendah, yaitu dengan mengambil titik nilai terendah atau nilai minimum.

3) Davies Bouldin Index

Davies Bouldin Index (DBI) adalah sebuah metrik yang digunakan untuk mengukur kinerja *algoritma clustering* dalam menentukan jumlah *cluster* yang optimal setelah proses pengelompokan data selesai. Prinsip dasar DBI adalah untuk memaksimalkan jarak antara *cluster* satu dengan yang lain, dan juga meminimalkan jarak antar objek dalam satu klaster. Semakin rendah nilai DBI yang didapat (dengan nilai non-negatif ≥ 0), maka semakin baik nilai *clustering* yang dihasilkan dari pengelompokan metode *Algoritma K-Means*. [14]. Adapun beberapa rumus yang akan digunakan untuk melakukan pencarian nilai dari *Davies Bouldin Index,* sebagai berikut.

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (3)$$

Berdasarkan rumus (3), diketahui bahwa m_i merupakan jumlah data dalam *cluster* ke – i, kemudian c_i merupakan *centroid cluster* ke – i, dan $d(x_j,c_i)$ merupakan jarak dari data ke – i ke titik *cluster* i, rumus ini digunakan untuk mengetahui matrik kohesi dalam sebuah *Cluster*, yang mana sekecil mungkin nilai kohesi yang diperoleh maka nilai *cluster* nya sudah sangat optimal.

$$SSB_{i,j} = d(c_i, c_j) \quad (4)$$

Berdasarkan rumus (4), digunakan untuk mengetahui nilai separasi antar *Cluster* yang dihitung menggunakan persamaan tersebut, dan sebesar mungkin nilai separasi yang diperoleh, maka nilai *cluster* nya sudah sangat optimal.

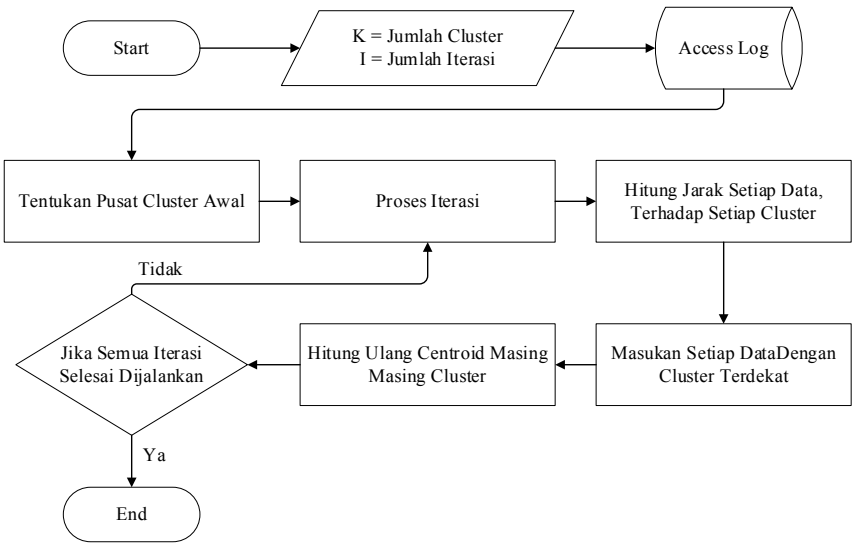
$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (5)$$

Setelah dilakukannya perhitungan nilai kohesi dan separasi, kemudian melakukan pengukuran ratio (Rij) untuk mengidentifikasi nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j, dengan menggunakan rumus (5). Kemudian hasil yang diperoleh dari perhitungan nilai ratio tersebut akan digunakan untuk mencari nilai *Davies Bouldin Index.*

$$DBI = \frac{1}{K} \sum_{i=1}^{K} \max_{i \neq j} (R_{i,j}) \quad (6)$$

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 2, Juni 2024, Pp. 878-888

Berdasarkan pada rumus (6), diketahui bahwa untuk mencari nilai *Davies Bouldin Index* adalah dengan menggunakan persamaan tersebut, dimana semakin rendah nilai DBI yang didapat (dengan nilai non-negatif ≥ 0), maka semakin baik nilai *clustering* yang dihasilkan dari pengelompokan metode *Algoritma K-Means.*



Gambar. 3. Diagram Alir Proses Clustering

Alur proses *Clustering* dalam penelitian ini dapat diperhatikan pada Gambar 3. Proses *Clustering* membutuhkan dua parameter masukan penting, yaitu jumlah *Cluster* yang sudah ditentukan dan jumlah Iterasi yang akan dilakukan. Jumlah Iterasi ini berperan dalam menentukan kapan proses *Clustering* akan berhenti. Selain itu, dalam proses *Clustering*, titik pusat awal atau *Centroid* untuk setiap *Cluster* juga harus ditentukan [15].

D. Hasil Analisis Data

Pada tahapan pengujian ini, data *Access Log* dianalisis dengan menggunakan metode *K-Means Clustering* untuk mengidentifikasi halaman yang sering diakses, dan karakteristik pengguna seperti apa saja yang ada di Universitas xyz. Dengan adanya tahapan ini, penulis dapat mengevaluasi keberhasilan dari hasil proses *Clustering* yang telah dilakukan.

E. Kesimpulan

Langkah berikutnya merupakan tahapan terakhir, dimana dapat ditarik kesimpulan yang didapat dari penelitian, dan hasil dari proses Clustering data Access Log dengan Algoritma K-Means, yang sudah didapatkan kesimpulan hasil dari intensitas kunjungan terhadap perilaku pengguna jaringan internet, yang terbagi dalam Clustering intensitas kunjungan tinggi, sedang, dan rendah.

III. HASIL DAN PEMBAHASAN

Data *Access Log* yang digunakan masih dalam bentuk file extensi *.txt kemudian dilakukan *Cleansing* dan *Preprocessing Data*, dengan menghilangkan tanda baca dan kata-kata yang tidak diperlukan, dan hanya mengambil atribut data yang akan diteliti, seperti *IP Address,* URL halaman yang diakses, dan *Timestamp.* Berikut merupakan contoh data mentah *Access Log,* dapat diperhatikan pada Tabel 1.

TABEL I
CONTOH DATA MENTAH ACCESS LOG

61.245.171.187 ix.cs.uoregon.edu - [01/May/2020:00:06:42 -0700] "GET /~nisansa/pdf/ToastMasters/ACB7%20-%20TecPre%20-%20P1%20-%20A%20WordNet%20for%20Sinhala.pdf HTTP/1.1" 200 1612724 "https://www.google.com/" "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.113 Safari/537.36" -pct
45.152.32.204 ix.cs.uoregon.edu - [01/May/2020:00:14:54 -0700] "GET /~michal/autograde12S/index.cgi?n=PmWiki HTTP/1.0" 302 194 "https://ix.cs.uoregon.edu/" "Mozilla/5.0 (Windows NT 6.2; Win64; x64; rv:53.0) Gecko/20100101 Firefox/53.0" -pct
216.244.66.234 ix.cs.uoregon.edu - [08/May/2020:23:39:27 -0700] "GET /%7Ehornof/downloads/ToCHI_2001.pdf HTTP/1.1" 200 1019753 "-" "Mozilla/5.0 (compatible; DotBot/1.1; http://www.opensiteexplorer.org/dotbot, help@moz.com)" -pct

Pada pengambilan data *Access Log* didapatkan 10000 data mentah, dan setelah dilakukan tahap *Cleansing* dan *Preprocessing Data*, dihasilkan data yang siap untuk dilakukan proses analisis sebanyak 898 data bersih dari simbol

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 2, Juni 2024, Pp. 878-888

dan karakter data mentah, dan setelah melakukan pemilihan atribut yang akan digunakan, hanya dipilih atribut *IP Address,* URL halaman yang diakses, dan *Timestamp* untuk dilakukannya analisis *K-Means Clustering*, contoh hasil dari data yang sudah siap untuk dilakukan analisis *K-Means Clustering* bisa diperhatikan pada Tabel 2.

TABEL II
CONTOH HASIL PREPROCESSING DAN CLEANSING DATA

IP Address	Date	URL
61.245.171.187	01/May/2020:00:06:42	https://www.google.com/
45.152.32.204	01/May/2020:00:14:54	https://ix.cs.uoregon.edu/
103.121.208.245	01/May/2020:00:17:36	https://www.google.com/

Kemudian dari hasil *Cleansing* dan *Preprocessing* data tersebut, maka dapat dilakukan ke tahapan selanjutnya yaitu melakukan *Clustering* dengan menerapkan metode *Algoritma K-Means* untuk menentukan pengelompokan data. Dari keseluruhan data *Access Log* ini akan dikelompokkan menjadi tiga *Cluster*, yaitu *Cluster* 1 tingkat intensitas kunjungan tinggi, *Cluster* 2 tingkat intensitas kunjungan sedang, dan *Cluster* 3 tingkat intensitas kunjungan rendah.

Dalam implementasi *Algoritma K-Means Clustering*, ditentukan nilai titik tengah atau *Centroid* dari data yang didapat, dengan mengelompokkan data ke dalam tiga tingkat *Cluster*, Penentuan *Cluster* dibagi menjadi tiga kelompok yang berbeda, yaitu kelompok dengan *Cluster* tingkat intensitas kunjungan tinggi (Cluster 1), *Cluster* tingkat intensitas kunjungan sedang (Cluster 2), dan *Cluster* tingkat intensitas kunjungan rendah (Cluster 3). Oleh karena itu, terdapat tiga titik pusat yang berkaitan dengan *Cluster* ini. Penetapan nilai-nilai pusat *Cluster* ini dilakukan dengan mengambil nilai tertinggi (maksimum), untuk kelompok dengan tingkat intensitas kunjungan tinggi (Cluster 1), nilai rata-rata (mean), untuk kelompok dengan tingkat intensitas kunjungan sedang (Cluster 2), dan nilai terendah (minimum), untuk kelompok dengan tingkat intensitas kunjungan rendah (Cluster 3). Nilai *Centroid* tersebut dapat diperhatikan pada Tabel 3.

TABEL III
CENTROID DATA ITERASI I

K	1 May 2020	2 May 2020	3 May 2020	4 May 2020	5 May 2020	6 May 2020	7 May 2020	8 May 2020	9 May 2020
C 1	442	345	183	282	756	937	641	649	31
C 2	88	67	43	58	130	150	116	142	7
C 3	1	1	2	1	3	1	3	1	1

Diketahui C1, memiliki nilai 442, 345, 183, 282, 756, 937, 641, 649, 31, dimana nilai ini didapatkan pada nilai tertinggi data di Tabel 3. Kemudian C2, memiliki nilai 88, 67, 43, 58, 130, 150, 116, 142, 7. Dan C3, memiliki nilai 1, 1, 2, 1, 3, 1, 3, 1, 1, dimana nilai ini didapatkan pada data di Tabel 3. Perhitungan (7), merupakan salah satu contoh dari pencarian nilai mean untuk C2, dengan menggunakan rumus (2).

$$\bar{X} = \frac{442 + 77 + 1 + 4 + 87 + 1 + 4}{7} \quad (7)$$

$$\bar{X} = \frac{616}{7}$$

$$\bar{X} = 88$$

TABEL IV
PERHITUNGAN JARAK PUSAT CLUSTER ITERASI I

Web Category	1 May 2020	2 May 2020	3 May 2020	4 May 2020	5 May 2020	6 May 2020	7 May 2020	8 May 2020	9 May 2020
Education	442	345	183	282	756	937	641	649	31
Portal Site	77	73	71	67	72	71	70	165	15
E Commerce	1	2	0	2	5	3	3	1	0
Medical Care	4	4	6	2	3	0	0	0	0
Search Engine	87	47	38	54	71	41	72	177	2
Blog	1	1	2	2	5	1	21	5	3
Lain - Lain	4	1	2	1	0	1	5	1	1

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 2, Juni 2024, Pp. 878-888

Berikut merupakan langkah pencarian nilai *Centroid*, dan pengelompokan data pada Iterasi 1, dengan menggunakan rumus (1), dapat diperhatikan pada salah satu contoh perhitungan (8), dan pada tabel 4.

$$D(Portal\ Site, C1) = \sqrt{\begin{matrix}(77-442)^2+(73-345)^2+(71-183)^2+(67-282)^2+\\(72-756)^2+(71-937)^2+(70-641)^2+(165-649)^2+\\(15-31)^2\end{matrix}} \quad (8)$$

$$= 1429{,}805$$

$$D(Portal\ Site, C2) = \sqrt{\begin{matrix}(77-88)^2+(73-67)^2+(71-43)^2+(67-58)^2+\\(72-130)^2+(71-150)^2+(70-116)^2+(165-142)^2+\\(15-7)^2\end{matrix}}$$

$$= 155{,}4816$$

$$D(Portal\ Site, C3) = \sqrt{\begin{matrix}(77-1)^2+(73-1)^2+(71-2)^2+(67-1)^2+\\(72-3)^2+(71-1)^2+(70-3)^2+(165-1)^2+\\(15-1)^2\end{matrix}}$$

$$= 247{,}6267$$

Proses *Clustering* dilakukan dengan menentukan jarak terdekat dari setiap data yang sedang diproses dan perhitungan *K-Means Clustering* akan terus berlanjut dengan Iterasi berulang, hingga hasil pengelompokan data Iterasi ditemukan berjumlah sama dengan pengelompokan data dari Iterasi sebelumnya. Pada langkah Iterasi 1, terbentuk *Cluster* data, tingkat kunjungan pengguna internet, dengan intensitas kunjungan tertinggi (Cluster 1), dengan *Web Category Education,* kemudian intensitas kunjungan sedang (Cluster 2), dengan *Web Category Search Engine* dan *Portal Site,* kemudian intensitas kunjungan terendah (Cluster 3), dengan *Web Category E Commerce, Medical Care, Blog, Lain – Lain,* dapat dilihat pada Tabel 5.

TABEL V
HASIL CLUSTERING ITERASI I

Web Category	Cluster 1	Cluster 2	Cluster 3	Centroid
Education	0	1347,9136	1641,795	0
Portal Site	1429,805	155,4816	247,6267	155,4816
E Commerce	1639,599	294,46901	3,872983	3,872983
Medical Care	1641,912	296,35789	6,855655	6,855655
Search Engine	1451,642	137,80421	235,752	137,8042
Blog	1632,175	287,01568	18,68154	18,68154
Lain - Lain	1641,595	296,10978	4,690416	4,690416

Namun proses Iterasi ini belum valid, sehingga dilanjutkan proses perhitungan Iterasi berikut nya. Proses perhitungan Iterasi ini akan berhenti pada iterasi ke 2, dimana pada proses Iterasi ini dilakukan perhitungan nilai pusat atau *Centroid*.

TABEL VI
CENTROID DATA ITERASI II

Cluster	1 May 2020	2 May 2020	3 May 2020	4 May 2020	5 May 2020	6 May 2020	7 May 2020	8 May 2020	9 May 2020
C 1	442	345	183	282	756	937	641	649	31
C 2	82	60	54,5	60,5	71,5	56	71	171	8,5
C 3	2,5	2	2,5	1,75	3,25	1,25	7,25	1,75	1

Berikut merupakan langkah pencarian *Centroid* data Iterasi 2, dimana salah satu pencarian *Centroid,* dilakukan dengan menghitung data yang terdapat dibagian *Cluster*. Seluruh data *Cluster* dijumlah dan dibagi dengan jumlah banyaknya *Web Category* yang terdapat dibagian *Cluster* tersebut. Perhitungan (9), merupakan salah satu contoh dari pencarian nilai mean untuk C3, dengan menggunakan rumus (2), dan hasil keseluruhan dapat dilihat pada Tabel 5.

$$\bar{X} = \frac{1+4+1+4}{4} \quad (9)$$

$$\bar{X} = 2{,}5$$

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 2, Juni 2024, Pp. 878-888

TABEL VII
PERHITUNGAN JARAK PUSAT CLUSTER ITERASI II

Web Category	1 May 2020	2 May 2020	3 May 2020	4 May 2020	5 May 2020	6 May 2020	7 May 2020	8 May 2020	9 May 2020
Education	442	345	183	282	912	937	641	649	31
Portal Site	77	73	71	67	72	71	70	165	15
E Commerce	1	2	0	2	5	3	3	1	0
Medical Care	4	4	6	2	3	0	0	0	0
Search Engine	87	47	38	54	71	41	72	177	2
Blog	1	1	2	2	5	1	21	5	3
Lain - Lain	4	1	2	1	0	1	5	1	1

Setelah menemukan nilai pusat atau *Centroid*, proses selanjutnya akan dilakukan dengan mencari nilai *Centroid* nya. Berikut merupakan langkah pencarian nilai *Centroid*, dan pengelompokan data pada Iterasi ke 2. Perhitungan (10), merupakan salah satu contoh dari pencarian nilai *Centroid* nya, dengan menggunakan rumus (1).

$$D(Blog, C1) = \sqrt{\begin{matrix}(1-442)^2 + (1-345)^2 + (2-183)^2 + (2-282)^2 + \\ (5-756)^2 + (1-937)^2 + (21-641)^2 + (5-649)^2 + \\ (3-31)^2\end{matrix}} \quad (10)$$

$$= 1632{,}175$$

$$D(Blog, C2) = \sqrt{\begin{matrix}(1-82)^2 + (1-60)^2 + (2-54{,}5)^2 + (2-60{,}5)^2 + \\ (5-71{,}5)^2 + (1-56)^2 + (21-71)^2 + (5-171)^2 + \\ (3-8{,}5)^2\end{matrix}}$$

$$= 231{,}8491$$

$$D(Blog, C3) = \sqrt{\begin{matrix}(1-2{,}5)^2 + (1-2)^2 + (2-2{,}5)^2 + (2-1{,}75)^2 + \\ (5-3{,}25)^2 + (1-1{,}25)^2 + (21-7{,}25)^2 + (5-1{,}75)^2 + \\ (3-1)^2\end{matrix}}$$

$$= 14{,}50216$$

Dari hasil perhitungan pada Iterasi ke 2 ditemukan kesamaan pengelompokan dengan data dari Iterasi ke 1. dimana pada langkah Iterasi ke 2, terbentuk *Cluster* data, tingkat kunjungan pengguna internet, dengan intensitas kunjungan tertinggi (Cluster 1), dengan *Web Category Education,* kemudian intensitas kunjungan sedang (Cluster 2), dengan *Web Category Search Engine* dan *Portal Site,* kemudian intensitas kunjungan terendah (Cluster 3), dengan *Web Category E Commerce, Medical Care, Blog, Lain – Lain,* sehingga proses perhitungan *K-Means Clustering* telah selesai dan berhenti di Iterasi ke 2, data dapat dilihat pada Tabel 7.

TABEL VIII
HASIL CLUSTERING ITERASI II

Web Category	Cluster 1	Cluster 2	Cluster 3	Centroid
Education	0	1440,483	1638,799	0
Portal Site	1429,805	28,51315	244,7791	28,51315
E Commerce	1639,599	239,0502	5,857687	5,857687
Medical Care	1641,912	239,1171	8,764274	8,764274
Search Engine	1451,642	28,51315	232,8772	28,51315
Blog	1632,175	231,8491	14,50216	14,50216
Lain - Lain	1641,595	239,3972	4,506939	4,506939

Setelah dilakukan nya pengelompokan *Algoritma K-Means,* data *cluster* akan diuji berdasarkan *Davies Bouldin Index* untuk menunjukan validasi dari data yang sudah dilakukan *Clustering.* Untuk menghitung *Davies Bouldin Index,* hal pertama yang harus dilakukan adalah mencari nilai ratio untuk mengetahui nilai ratio dari beberapa

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 9, No. 2, Juni 2024, Pp. 878-888

Cluster. Perhitungan (11), merupakan salah satu contoh langkah pencarian nilai kohesi, dengan menggunakan rumus (3) dan hasil keseluruhan dapat diperhatikan pada tabel 8.

$$SSW_3 = \frac{1}{4}(5{,}857687 + 8{,}764274 + 14{,}50216 + 4{,}506939) \quad (11)$$
$$SSW_3 = 8{,}407765$$

TABEL VIII
HASIL PERHITUNGAN KOHESI

Web Category	Cluster 1	Cluster 2	Cluster 3	Centroid	SSW
Education	0	1440,483	1638,799	0	0
Portal Site	1429,805	28,51315	244,7791	28,51315	28,51315
E Commerce	1639,599	239,0502	5,857687	5,857687	8,407765
Medical Care	1641,912	239,1171	8,764274	8,764274	
Search Engine	1451,642	28,51315	232,8772	28,51315	
Blog	1632,175	231,8491	14,50216	14,50216	
Lain - Lain	1641,595	239,3972	4,506939	4,506939	

Berdasarkan Tabel 8, diketahui nilai kohesi dari *Cluster* 1 memiliki nilai 0, *Cluster* 2 memiliki nilai 28,51315, dan *Cluster* 3 memiliki nilai 8,407765. Dimana nilai ini didapatkan pada perhitungan average di setiap *cluster*. Setelah mendapatkan nilai kohesi, proses selanjutnya akan dilakukan dengan mencari nilai separasinya. Perhitungan (12), merupakan salah satu contoh langkah pencarian nilai separasi, dengan menggunakan rumus (4), dan hasil keseluruhan dapat diperhatikan pada tabel 9.

$$SSB_{1,3} = \sqrt{\begin{matrix}(442-2{,}5)^2 + (345-2)^2 + (183-2{,}5)^2 + (282-1{,}75)^2 + \\ (756-3{,}25)^2 + (937-1{,}25)^2 + (641-7{,}25)^2 + (649-1{,}75)^2 + \\ (31-1)^2\end{matrix}} \quad (12)$$
$$SSB_{1,3} = 1638{,}798893$$

TABEL IX
HASIL PERHITUNGAN SEPARASI

SSB	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0	1440,482905	1638,798893
Cluster 2	1440,482905	0	237,1946722
Cluster 3	1638,798893	237,1946722	0

Berdasarkan Tabel 9, diketahui nilai separasi yang diperoleh pada masing masing *cluster*. Dimana salah satu nilai yang didapatkan ini berdasarkan perhitungan jarak *centroid* 1 dengan *centroid* 3, seperti pada contoh perhitungan diatas, dengan menggunakan data yang ada pada Tabel 5 centroid data iterasi ke 2. Setelah mendapatkan nilai separasinya, proses selanjutnya akan dilakukan dengan mencari nilai rationya. Perhitungan (13), merupakan salah satu contoh langkah pencarian nilai ratio, dengan menggunakan rumus (5), dan hasil keseluruhan dapat diperhatikan pada tabel 10.

$$R_{2,3} = \frac{28{,}51315 + 8{,}407765}{237{,}1946722} \quad (13)$$
$$R_{2,3} = 0{,}155656615$$

TABEL X
HASIL PERHITUNGAN RATIO

R	Cluster 1	Cluster 2	Cluster 3	Ratio Max
Cluster 1	0	0,019794164	0,005130443	0,019794164
Cluster 2	0,019794164	0	0,155656615	0,155656615
Cluster 3	0,005130443	0,155656615	0	0,155656615

Berdasarkan Tabel 10, dapat diketahui nilai ratio yang diperoleh pada masing masing *cluster.* Seperti pada contoh perhitungan diatas, nilai kohesi *cluster* ke 2 ditambah dengan nilai kohesi *cluster* ke 3 yang ada di Tabel 8, kemudian dibagi dengan nilai separasi *cluster* 2 ke 3 yang ada di Tabel 9, sehingga mendapatkan nilai ratio *cluster* 2 ke 3 adalah 0,155656615. Kemudian dari masing masing *cluster* diperoleh nilai ratio maksimalnya, sehingga proses selanjutnya akan dilakukan perhitungan dengan menentukan nilai dari *Davies Bouldin Index*, dengan menggunakan rumus (6), dan dapat diperhatikan pada perhitungan (14).

$$DBI = \frac{1}{3}(0{,}019794164 + 0{,}155656615 + 0{,}155656615) \quad (14)$$
$$DBI = 0{,}110369132$$

Hasil perhitungan dari nilai *Davies Bouldin Index* mendapatkan nilai 0,110369132 dalam artian jika nilai *DBI* mendekati angka nol, maka hasil perhitungan *Cluster* sudah cukup optimal. Dapat diketahui bahwa pada proses Iterasi ke 2, hasil yang sama ditemukan saat menerapkan teknik pengelompokan data *Access Log* ke dalam 3 *cluster* dengan melalui proses 2 Iterasi. Dari 7 *Web Category* dengan atribut data tanggal 1 May 2020 hingga tanggal 9 May 2020, terdapat 3 tingkatan *Cluster* yaitu sebagai berikut: *Cluster* 1, tingkat intensitas pengunjung tertinggi, dengan *Category Education*. *Cluster* 2, tingkat intensitas pengunjung sedang, dengan *Category Portal Site,* dan *Search Engine. Cluster* 3, tingkat intensitas pengunjung terendah, dengan *Category E Commerce, Medical Care, Blog,* dan *Lain – Lain.* Kemudian dari perhitungan setiap *cluster* dilakukan uji validasi data, berdasarkan *Davies Bouldin Index* untuk menunjukan akurasi dari data yang sudah dilakukan *Clustering,* dan hasil dari perhitungan *cluster* yang dilakukan sudah cukup baik.

Penelitian ini memiliki perbedaan pada konsep *Clustering Access Log*, yang mana dalam melakukan *Clustering Access Log*, diperlukannya uji validasi dalam sebuah hasil *cluster*, sehingga hasil dari *clustering* yang dilakukan bisa dapat diketahui, data yang telah di *clustering* apakah valid atau tidak.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilakukan dalam penelitian ini, dapat disimpulkan bahwa implementasi *Algoritma K-Means Clustering* pada data *Access Log* telah berhasil di *Clustering* dan menunjukan hasil yang sesuai dengan harapan penulis, dengan mendapatkan hasil *Cluster* 1, intensitas kunjungan tertinggi, dengan *Web Category Education,* kemudian *Cluster* 2, intensitas kunjungan sedang, dengan *Web Category Portal Site,* dan *Search Engine,* dan *Cluster* 3, intensitas kunjungan terendah, dengan *Web Category E Commerce, Medical Care, Blog,* dan *Lain - Lain.* Berdasarkan hasil *Clustering* ini, dapat diketahui bahwa perilaku pengguna jaringan internet, sangat berpengaruh oleh faktor lingkungan dan aktivitas sehari-hari. Dalam penelitian ini dapat diketahui bahwa, *Access log* jaringan Internet memiliki jenis data yang sangat besar, sehingga penggunaan *Algoritma Data Mining* sangatlah diperlukan. Dengan penggunaan *Algoritma K-Means, Access log* dapat dengan mudah di *Clustering,* hal ini dapat mempermudah analisis dalam penggunaan jaringan Internet. Adapun kelemahan yang diperoleh penulis dari penggunaan Metode *Algoritma K-Means* ini, dimana *Algoritma K-Means* bisa berlaku jika nilai *mean* bisa dihitung, yang berarti atributnya bernilai integer numerik, jika nilainya bersifat kategorikal, mean tidak dapat dihitung, dan *Algoritma K-Means* memerlukan nilai K yang harus ditentukan.

DAFTAR PUSTAKA

[1] Prabawa, Kushendra Satria. Penerapan K-Means Untuk Pengelompokan Pengguna Internet Berdasarkan Elapsed dan Byte Transferred. Diss. Program Studi Teknik Informatika FTI-UKSW, 2016.
[2] Ferdinan, Yongki. "Analisis Traffic Pengguna Pada Jaringan Internet Nirkabel Di Bawah Pengelolaan BSI Universitas Islam Indonesia." (2012).
[3] Maulida, Linda. "Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means." JISKA (Jurnal Informatika Sunan Kalijaga) 2.3 (2018): 167-174.
[4] Prasetyo, Eko. "Data Mining Processing Data Into Information using MATLAB." Yogyakarta: ANDI Publishers (2014).
[5] Gunawan, Gun Gun Indra. ANALISIS PERILAKU PENGGUNA INTERNET DENGAN DATA MINING METODE K-MEANS. Diss. Universitas Siliwangi, 2021.
[6] Kuncoro, A. P., Hutomo, D. P., & Zulfadhilah, M. (2017). Pengelompokan Pengguna Internet Menggunakan Metode K-Means pada Data Log Akses Server. *Citisee 2017*, 149–153.
[7] Aini, Fadhilah Dhinur, Imam Riadi, and Rusydi Umar. "Perancangan Deteksi Anomali Traffic Untuk Investigasi Log Menggunakan Metode K-Means Clusters." Prosiding Seminar Sains Nasional dan Teknologi. Vol. 1. No. 1. 2018.
[8] Ridho, Farid, and Arya Aji Kusuma. "Deteksi Intrusi Jaringan dengan K-Means Clustering pada Akses Log dengan Teknik Pengolahan Big Data." Jurnal Aplikasi Statistika & Komputasi Statistik 10.1 (2018): 53-66.
[9] Fari'Utomo, Kevin Gusti Farras, Fitra Abdurrachman Bachtiar, and Nanang Yudi Setiawan. "Analisis Perilaku Pengguna pada PT Antar Surya Jaya: Harian Surya Surabaya menggunakan Metode K-Means Clustering." Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer 5.11 (2021): 4828-4835.
[10] Goel, Neha, and C. K. Jha. "Analyzing users behavior from web access logs using automated log analyzer tool." International Journal of Computer Applications 62.2 (2013).
[11] Sari, Riyani Wulan, Anjar Wanto, and Agus Perdana Windarto. "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)." KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer) 2.1 (2018).

[12] Heni, Sulastri, and A. Irham Gufroni. "Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia." J. Nas. Teknol. dan Sist. Inf 3.02 (2017): 299-305.
[13] Sadewo, Mhd Gading, et al. "Algoritma K-Means Dalam Mengelompokkan Desa/Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi." Seminar Nasional Teknologi Komputer & Sains (SAINTEKS). Vol. 1. No. 1. 2019.
[14] Ghufron, Ghufron, Bayu Surarso, and Rahmat Gernowo. "The implementations of K-medoids clustering for higher education accreditation by evaluation of Davies Bouldin index clustering." *Jurnal Ilmiah KURSOR* 10.3 (2020).
[15] Prabawa, Kushendra Satria. Penerapan K-Means Untuk Pengelompokan Pengguna Internet Berdasarkan Elapsed dan Byte Transferred. Diss. Program Studi Teknik Informatika FTI-UKSW, 2016.