

Web Usage Mining Menggunakan Algoritma Clustering K-Mean

Mira Orisa
Teknik Informatika S1
(Fakultas Teknologi Industri)
ITN Malang
(ITN Malang)
Malang, Indonesia
mir4orisa@gmail.com

Michael Ardita
Teknik Informatika S1
(fakultas teknologi Industri)
ITN Malang
(ITN Malang)
Malang, Indonesia
michael.ardita@lecturer.itn.ac.id

Abstrak— *Big data* merupakan sumber informasi yang besar. Sebelum menjadi informasi tentu *big data* tersebut harus di ekstrasi terlebih dahulu untuk mendapatkan pengetahuan. Disiplin ilmu yang mempelajari pengekstraksian *big data* menjadi pengetahuan disebut *data mining*. Banyak teknik pengolahan data dalam *data mining*, salah satunya adalah teknik *clustering*. Teknik *clustering* dapat digunakan untuk mengekstraksi *big data* menjadi pengetahuan. Pengetahuan yang dihasilkan pada teknik *clustering* berupa *cluster-cluster*. Teknik *clustering* dapat mengclusterkan tingkah laku dari pengguna suatu *website*. Pada suatu *log server web* mengandung banyak sekali data yang dapat diambil pengetahuannya. Ada beberapa tahap untuk mendapatkan pengetahuan. Tahap pertama aka dilakukan *preprocessing* data yaitu untuk membersihkan data dari atribut-atribut data yang tidak diperlukan. Pada tahap *pattern discovery* dilakukan pengekstraksian *dataset* dengan algoritma *K-means clustering*. Pada tahap *pattern analysis* dilakukan analisis *cluster-cluster* yang dihasilkan oleh algoritma *K-means clustering*. Hasil yan diperoleh bergantung pada nilai *centroid* awal. Pada skenario 1 menghasilkan *web* 7,14,18,19 termasuk *cluster* 1. *web* 1,3,5,10 termasuk *cluster* 2. Dan *web* 2,4,6,8,9,11,12,13,15,16,17,dan 20 termasuk *cluster* 3. Hasil pada skenario 2 yaitu *web* 1,3,5,7,10,14,18 dan 19 termasuk *cluster* 1. Hanya *web* 2 yang termasuk *cluster* 2. sedangkan *web* 4,6,8,9,11,12,13,15,16,17,dan 20 termasuk *cluster* 3.

Keywords—komponen; *big data*, *clustering*, algoritma *k-means*,

PENDAHULUAN

Pengolahan data dalam jumlah besar sering dikaji dalam disiplin ilmu data scientist. Pengolahan data yang tak terstruktur dalam jumlah yang besar tanpa bantuan algoritma tertentu sungguh merupakan pekerjaan yang sangat sulit dan memakan waktu yang lama. Dalam bidang ilmu data scientist, data-data tak terstruktur tersebut dapat diolah dengan bantuan *machine learning*. *machine learning* akan membuat data-data dapat diolah oleh komputer sehingga manusia mendapatkan informasi tertentu. Informasi ini seringkali digunakan untuk menentukan arah kebijakan suatu perusahaan.

Usaha untuk mendapatkan informasi dengan pola tertentu dari sebuah data yang berukuran besar disebut juga *data mining*. *Data mining* atau pengalian data merupakan bagian proses dari ilmu data scientist. Ada banyak penggunaan *machine learning* untuk mendapatkan informasi dari data yang diolah tersebut. Seperti penggunaan *machine learning naïve bayes* untuk mendapatkan informasi mengenai kalsifikasi kelayakan dari data-data program raskin[1]. Penggunaan *machine learning* algoritma genetika untuk mendapatkan informasi mengenai keragaman populasi[2]. Penggunaan *Machine learning K-means* untuk pengolahan

data mahasiswa untuk pengambilan keputusan tertentu seperti keputusan penentuan kebijakan untuk uang kuliah[3]. Penggunaan *machine learning K-means* juga diterapkan untuk pengelompokan data-data obat-obatan sebagai bahan pengambil keputusan dalam analisis dan pengendalian obat-obatan[4]. Kemudian untuk analisis data kerusakan tanaman cabe. Jadi dengan menggunakan *machine learning K-means* dapat menggali informasi untuk menentukan pengelompokan penyebab kerusakan pada tanaman tersebut[5]. Penggunaan *machine learning K-means* juga diterapkan untuk menentukan strategi penjualan variasi mobil[6]. Penggunaan *machine learning K-means* untuk menggali informasi mengenai pengelompokan penumpang bandara[7]. Penggunaan *machine learning fuzzy C-means Clustering* untuk pengelompokan kejadian gempa bumi[8]. Penggunaan *K-means Clustering* untuk menganalisis bisnis di perusahaan asuransi[9].

Berdasarkan penelitian pada 5 tahun terakhir bahwa *machine learning K-means* sudah terbukti mampu memberikan pengetahuan tertentu berdasarkan kasus yang diteliti oleh masing-masing peneliti. Oleh sebab itu maka pada penelitian ini akan dikembangkan *machine learning K-means* pada pengolahan data *log server web* untuk menggali pengetahuan tentang *cluster-cluster* pengguna *web*.

DATA MINING

Data mining merupakan disiplin ilmu yang mempelajari cara menganalisis data dalam ukuran yang besar yang tersimpan di sebuah database (knowledge discovery in database). Atau disebut penambangan pengetahuan dari data, ekstraksi pengetahuan, analisis data / pola, data arkeologi, dan pengerukan data. Ada beberapa langkah dalam proses penemuan pengetahuan yang ditunjukkan seperti Gambar 1. Dalam proses penemuan pengetahuan langkah demi langkah yang harus dijalankan yaitu[10]:

1. Pembersihan data (*data cleaning*)
2. Intergrasi data
3. Menyeleksi data
4. Transformasi data
5. *Data mining* atau proses mengekstraksi *big data* dengan algoritma tertentu
6. Evaluasi pola
7. Presentasi pengetahuan

Sumber data pada *data mining* dapat berupa database, data gudang, *Web*, repositori informasi lainnya[10].

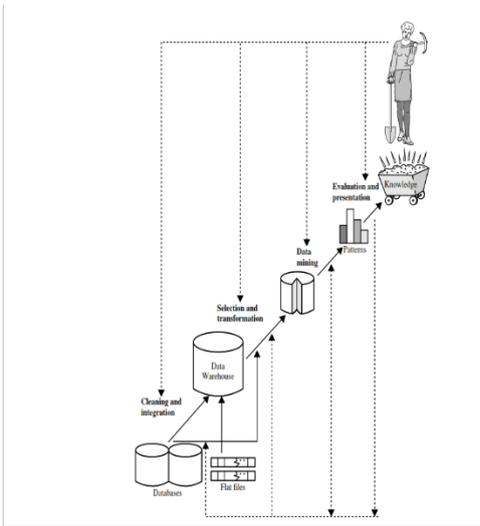


Fig. 1. Tahapan proses Penambangan pengetahuan dalam data mining(morgan ...page 7)

WEB USAGE MINING

Web usage mining merupakan teknik untuk menemukan pola aktifitas dari log pengguna web. Adapun tujuan untuk menangkap dan memodelkan pola perilaku dan profil pengguna yang berinteraksi dengan situs web. Dalam proses web usage mining terdapat beberapa tahap seperti:[11]

1. Pengumpulan data dan pra-pemrosesan data
2. Penemuan pola (Pattern discovery)
3. Analisis pola (Pattern analysis)

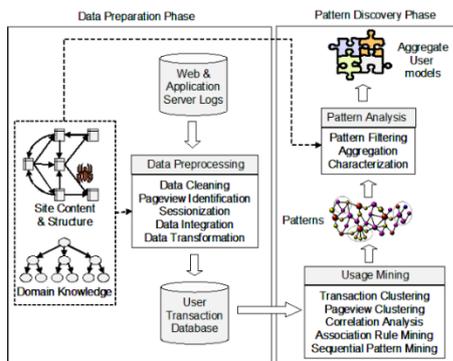


Fig. 2. Tahapan proses pada web usage mining (bing liu.2011.web data mining.second edition.springer.page 528)[11]

ALGORITMA CLUSTERING K-MEANS

Clustering adalah teknik mengelompokkan atau membagi data dalam cluster-cluster dimana objek dalam suatu cluster memiliki kemiripan yang tinggi. Tetapi memiliki perbedaan yang tinggi pula dengan objek pada cluster lainnya. Data yang dikelompokkan adalah data-data yang memiliki atribut numerik. Ada data yang memiliki label dan ada juga data yang tidak memiliki label. Pada clustering akan dicari terlebih dahulu adalah centroid awal atau titik pusat dari kelompok data. Ada beberapa kelompok metode clustering menurut para ahli yaitu[10]:

1. Metode berbasis partisi (partitioning methods)
2. Metode berbasis hirarki (hierarchical methods)
3. Metode berbasis kepadatan (density-based methods)

4. Metode berbasis kisi (grid-based methods)

Algoritma K-means adalah algoritma yang termasuk kedalam metode partisi (partitioning methods). Metode ini mempartisi data ke dalam kelompok-kelompok atau kluster. Metode ini bekerja dengan menentukan nilai centroid tiap kelompok/cluster. Centroid adalah titik pusat kluster. Algoritma K-means meminimalkan sum of squared error (SSE) antar objek-objek data dengan sejumlah k centroid. Ada beberapa langkah kerja dari algoritma K-means yaitu diperlihatkan oleh Fig 1 berikut[10].

Fig. 3. Gambar ilustrasi pseudo code algoritma K-means clustering(morgan)

Algorithm: k-means. The k-means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster.

Input:

- k: the number of clusters,
- D: a data set containing n objects.

Output: A set of k clusters.

Method:

- (1) arbitrarily choose k objects from D as the initial cluster centers;
- (2) repeat
- (3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
- (4) update the cluster means, that is, calculate the mean value of the objects for each cluster;
- (5) until no changes;

Berdasarkan fig 3 diatas bahwa algoritma K-means akan memilih secara acak untuk centroid awal(k). kemudian algoritma ini akan melihat objek yang bukan termasuk centroid awal akan dimasukkan kedalam kluster terdekat. Penentuan kluster terdekatnya berdasarkan ukuran jarak tertentu. Setelah itu centroid akan diperbaharui berdasarkan nilai rata-rata objek-objek yang ada di masing-masing kluster. Dua langkah tersebut yaitu langkah mengelompokkan objek kedalam kluster berdasarkan kedekatan ukuran jaraknya dan memperbaharui nilai centroid akan selalu diulang hingga semua centroid yang dihasilkan tidak berubah lagi atau bernilai tetap atau bernilai sama dengan centroid sebelumnya. Algoritma K-means akan menghasilkan sebuah output berdasarkan nilai centroid awalnya,dimana kita ketahui bahwa nilai centroid awal dipilih secara acak. Oleh sebab itu perlu melakukan beberapa percobaan pengambilan nilai centroid awalnya untuk mendapatkan hasil terbaik.[10]

EUCLIDEAN DISTANCE

Euclidean distance merupakan salah satu formula yang biasa digunakan untuk menghitung jarak antara dua objek yang memiliki atribut numerik selain formula manhattan distance, minkowski distance, dan supremum distance. Formula Euclidean distance [9]:

$$d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (1)$$

Dimana i dan j merupakan objek data memiliki p atribut bernilai numerik

$$i = (x_{i1}, x_{i2}, \dots, x_{ip})^2$$

$$j = (x_{j1}, x_{j2}, \dots, x_{jp})^2$$

METODE PENELITIAN

A. Pra-pemrosesan

Tahap pra-pemrosesan adalah tahap melakukan pembersihan data,intergrasi data, menyeleksi data dan melkukan transformasi data sehingga hanya data yang diperlukan dan relevan saja yang diolah. Data dari Table I

akan di ekstraksi sehingga didapatkanlah beberapa atribut seperti berikut.

TABLE I. DATASET

akses gam	akses text	akses audio/video
156	150	44
46	104	100
179	169	102
75	50	50
190	100	75
125	80	50
301	105	50
99	70	30
107	71	56
175	100	49
90	80	19
75	75	51
100	65	37
189	56	22
80	50	37
87	60	40
125	48	36
203	75	31
275	75	50
103	80	24

(peneliti.2021)

B. Pattern Discovery.

Untuk mengelompokkan data-data berdasarkan waktu akses dan kategori akses maka digunakan algoritma K-mean.

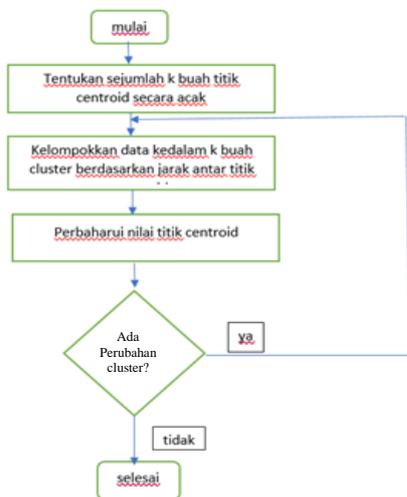


Fig. 4. Flowchar kerja algoritma K-means (peneliti,2021)

Alur kerja dari algoritma K-means adalah menentukan terlebih dahulu jumlah k buah titik centroid. Penentuan ini dilakukan secara acak. Algoritma K-means mengelompokkan data-data ke dalam sejumlah k buah yang sudah ditentukan sebelumnya berdasarkan nilai jarak terdekat dari dua buah titik centroid. Algoritma K-means akan mengulang langkah tersebut hingga titik centroid tidak lagi mengalami perubahan.

C. Pattern Analysis

Setelah pola ditemukan maka akan dilakukan analisis terhadap pola tersebut untuk menentukan pola-pola yang tidak relevan.

HASIL DAN PEMBAHASAN

Data yang diolah menggunakan Algoritma K-means menghasilkan pengetahuan berupa cluster-cluster yaitu cluster cluster pengakses gambar/image, cluster pengakses teks, dan cluster pengakses audio/video. Hasil dari algoritma K-means bergantung pada nilai centroid awal.

A. Hasil pengujian skenario 1

Nilai centroid awal pada skenario 1 untuk ke tiga cluster ditunjukkan oleh Tabel 2. Nilai centroid pada cluster 1 akses gambar sebesar 275, sedangkan nilai centroid pada cluster 2 sebesar 75, dan nilai centroid pada cluster 3 sebesar 50. Nilai centroid pada cluster 1 akses teks sebesar 179, sedangkan nilai centroid pada cluster 2 sebesar 169, dan nilai centroid pada cluster 3 sebesar 102. Nilai centroid pada cluster 1 akses video sebesar 46, sedangkan nilai centroid pada cluster 2 sebesar 104, dan nilai centroid pada cluster 3 sebesar 100.

TABLE II. TABEL CENTROID AWAL SKENARIO 1

atribut	c1	c2	c3
akses gambar	275	75	50
akses teks	179	169	102
akses video	46	104	100

(peneliti.2021)

TABLE III. TABEL CENTROID PADA ITERASI 1 SKENARIO 1

atribut	c1	c2	c3
akses gambar	242	77,75	38,25
akses teks	175	129,75	67,5
akses video	92,66666667	69,41666667	44,16666667

(peneliti.2021)

Hanya diperlukan satu kali iterasi karena pembagian cluster awal dengan pembagian cluster pada iterasi 1 sudah sama tidak ada perubahan. Dan Nilai centroid pada iterasi 1 ditunjukkan oleh Tabel 3. Nilai centroid pada cluster 1 akses gambar sebesar 242, sedangkan nilai centroid pada cluster 2 sebesar 77,75, dan nilai centroid pada cluster 3 sebesar 38,25. Nilai centroid pada cluster 1 akses teks sebesar 175, sedangkan nilai centroid pada cluster 2 sebesar 129,75, dan nilai centroid pada cluster 3 sebesar 67,5. Nilai centroid pada cluster 1 akses video sebesar 92,666, sedangkan nilai centroid pada cluster 2 sebesar 69,416, dan nilai centroid pada cluster 3 sebesar 44,166. Hasil implementasi algoritma K-means clustering ditunjukkan pada Fig 5. Pengetahuan yang dihasilkan oleh aplikasi berbasis android tersebut yaitu web 7,14,18,19 termasuk ke dalam cluster Gambar. Sedangkan web 1,3,5,10 termasuk ke dalam cluster Teks. Dan selebihnya termasuk kedalam cluster video yaitu 2,4,6,8,9,11,12,13,15,16,17,dan 20.

web5	TEXT
web4	VIDEO
web6	VIDEO
web7	GAMBAR
web8	VIDEO
web9	VIDEO
web10	TEXT
web11	VIDEO
web12	VIDEO
web13	VIDEO
web14	GAMBAR
web15	VIDEO
web16	VIDEO
web17	VIDEO
web18	GAMBAR
web19	GAMBAR
web20	VIDEO

Fig. 5. Hasil implementasi algoritma K-means pada OS android

TABLE IV. TABEL PENGETAHUAN

nama website	Gambar	Teks	Video	D1	D2	D3	MIN	CLUSTER
website 1	156	150	44	112,4683289	36,37736247	102,4929876	36,37736247	C2
website 2	46	104	100	207,1669013	135,5002306	80,56764756	80,56764756	C3
website 3	179	169	102	127,9047497	52,41004198	143,9269925	52,41004198	C2
website 4	75	50	50	169,6971567	129,0980732	26,89137098	26,89137098	C3
website 5	190	100	75	67,45098007	34,15131769	106,5824212	34,15131769	C2
website 6	125	80	50	117,6100548	72,67263928	34,51780941	34,51780941	C3
website 7	301	105	50	66,04259989	129,5948012	211,4307905	66,04259989	C1
website 8	99	70	30	143,44729	103,693358	15,52887096	15,52887096	C3
website 9	107	71	56	136,3291055	90,59697843	18,65419971	18,65419971	C3
website 10	175	100	49	71,41165871	35,03302014	87,96294201	35,03302014	C2
website 11	90	80	19	153,2306268	109,7830246	27,43135372	27,43135372	C3
website 12	75	75	51	167,5085819	115,1946722	19,74789018	19,74789018	C3
website 13	100	65	37	142,5767337	103,6716572	11,16448984	11,16448984	C3
website 14	189	56	22	59,54934928	87,77990943	99,75710083	59,54934928	C1
website 15	80	50	37	164,3643057	127,7314076	24,26545899	24,26545899	C3
website 16	87	60	40	156,0228349	115,6084448	11,75354556	11,75354556	C3
website 17	125	48	36	120,7440475	100,872754	39,63347701	39,63347701	C3
website 18	203	75	31	39,76336253	71,51092574	111,2563669	39,76336253	C1
website 19	275	75	50	35,13723097	115,342154	182,512043	35,13723097	C1
website 20	103	80	24	139,7466458	97,73081653	25,0095815	25,0095815	C3

Hasil perhitungan ditunjukkan pada Tabel 4. Pada tabel tersebut ditunjukkan nilai D1 merupakan jarak website dengan nilai centroid 1. D2 merupakan jarak website dengan nilai centroid 2. D3 adalah jarak website dengan nilai centroid 3. Algoritma K-means mengelompokkan web berdasarkan jarak terdekat dengan titik centroid. Website 1 termasuk ke dalam cluster 2 yaitu cluster teks berdasarkan nilai minimumnya adalah di D2 sebesar 36,3777 seperti terlihat pada Tabel 4. Begitu juga website 2 termasuk ke dalam cluster 3 berdasarkan nilai minimumnya berada di D3 sebesar 80,567. Website 3 termasuk ke dalam cluster 2 juga sama seperti website 1 karena website 3 memiliki nilai minimum yang berada di D2 sebesar 52,410. Website 4 termasuk ke dalam cluster 3 yaitu cluster video karena memiliki nilai minimum di D3 sebesar 26,891. Website 5 termasuk ke dalam cluster 2 yaitu cluster teks karena memiliki nilai minimum di D2 sebesar 34,151. Website 6 termasuk ke dalam cluster 3 yaitu cluster video karena memiliki nilai minimum di D3 sebesar 34,517. Website 7 termasuk ke dalam cluster 1 yaitu cluster gambar karena

memiliki nilai minimum di D1 sebesar 66,042. Website 8 termasuk ke dalam cluster 3 yaitu cluster video karena memiliki nilai minimum di D3 sebesar 15,528. Website 9 termasuk ke dalam cluster 3 juga yaitu cluster video karena memiliki nilai minimum di D3 sebesar 18,654. Website 10 termasuk ke dalam cluster 2 yaitu cluster teks karena memiliki nilai minimum di D2 sebesar 35,033. Website 11 termasuk ke dalam cluster 3 yaitu cluster video karena memiliki nilai minimum di D3 sebesar 27,431. Website 12 termasuk ke dalam cluster 3 juga yaitu cluster video karena memiliki nilai minimum di D3 sebesar 19,747. Website 13 termasuk ke dalam cluster 3 yaitu cluster video karena memiliki nilai minimum di D3 sebesar 11,164. Website 14 termasuk ke dalam cluster 1 yaitu cluster gambar karena memiliki nilai minimum di D1 sebesar 54,549. Website 15 termasuk ke dalam cluster 3 yaitu cluster video karena memiliki nilai minimum di D3 sebesar 24,265. Website 16 termasuk ke dalam cluster 3 juga yaitu cluster video karena memiliki nilai minimum di D3 sebesar 11,753. Website 17 termasuk ke dalam cluster 3 yaitu cluster video karena memiliki nilai minimum di D3 sebesar 39,633. Website 18 termasuk ke dalam cluster 1 yaitu cluster gambar karena memiliki nilai minimum di D1 sebesar 39,763. Website 19 termasuk ke dalam cluster 1 juga yaitu cluster gambar karena memiliki nilai minimum di D1 sebesar 35,137. Website 20 termasuk ke dalam cluster 3 yaitu cluster video karena memiliki nilai minimum di D3 sebesar 25,009.

B. Hasil pengujian skenario 1

Pada skenario 2 ini mengambil nilai centroid yang berbeda dengan skenario 1 seperti yang ditunjukkan pada Tabel 5.

TABLE V. TABEL CENTROID AWAL SKENARIO 2

atribut	c1	c2	c3
akses gambar	175	100	49
akses teks	156	150	44
akses video	179	169	102

Nilai centroid pada cluster 1 akses gambar sebesar 175, sedangkan nilai centroid pada cluster 2 sebesar 100, dan nilai centroid pada cluster 3 sebesar 49. Nilai centroid pada cluster 1 akses teks sebesar 156, sedangkan nilai centroid pada cluster 2 sebesar 150, dan nilai centroid pada cluster 3 sebesar 44. Nilai centroid pada cluster 1 akses video sebesar 179, sedangkan nilai centroid pada cluster 2 sebesar 169, dan nilai centroid pada cluster 3 sebesar 102. Pada skenario 2 membutuhkan 11 kali iterasi sehingga menghasilkan cluster-cluster seperti yang ditunjukkan pada Tabel 6.

TABLE VI. TABEL PENGETAHUAN

nama website	akses gambar	akses text	akses audio/vid	D1	D2	D3	MIN	CLUSTER
website 1	156	150	44	70,52714	131,727	102,5967	70,52714	C1
website 2	46	104	100	169,1954	0	87,89198	0	C2
website 3	179	169	102	86,83938	148,0473	145,7716	86,83938	C1
website 4	75	50	50	143,943	79,1012	29,39078	29,39078	C3
website 5	190	100	75	29,08312	146,2088	105,3229	29,08312	C1
website 6	125	80	50	86,85953	96,52461	33,11413	33,11413	C3
website 7	301	105	50	92,55311	259,8577	208,019	92,55311	C1
website 8	99	70	30	116,8442	94,15413	10,04535	10,04535	C3
website 9	107	71	56	106,6985	82,13404	20,2507	20,2507	C3
website 10	175	100	49	33,93123	138,7732	85,63825	33,93123	C1
website 11	90	80	19	125,5143	95,2523	25,29463	25,29463	C3
website 12	75	75	51	136,5735	63,89836	26,41969	26,41969	C3
website 13	100	65	37	116,3006	91,68424	3,942772	3,942772	C3
website 14	189	56	22	60,11304	169,8146	94,22507	60,11304	C1
website 15	80	50	37	140,1903	89,67162	23,56037	23,56037	C3
website 16	87	60	40	129,777	84,95293	11,76281	11,76281	C3
website 17	125	48	36	101,809	116,0733	33,65331	33,65331	C3
website 18	203	75	31	36,54214	173,9281	106,7563	36,54214	C1
website 19	275	75	50	72,50571	236,1821	178,638	72,50571	C1
website 20	103	80	24	111,9289	97,98469	21,2902	21,2902	C3

Hasil perhitungan ditunjukkan pada Tabel 6. *Website 1* termasuk ke dalam *cluster 1* yaitu *cluster* gambar berdasarkan nilai minimumnya adalah di D1 sebesar 70,527. *Website 2* termasuk ke dalam *cluster 2* berdasarkan nilai minimumnya berada di D2 sebesar 0. *Website 3* termasuk ke dalam *cluster 1* juga sama seperti *website 1* karena *website 3* memiliki nilai minimum yang berada di D1 sebesar 86,839. *Website 4* termasuk ke dalam *cluster 3* yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 29,390. *Website 5* termasuk ke dalam *cluster 1* yaitu *cluster* gambar karena memiliki nilai minimum di D1 sebesar 29,083. *Website 6* termasuk ke dalam *cluster 3* yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 33,114. *Website 7* termasuk ke dalam *cluster 1* yaitu *cluster* gambar karena memiliki nilai minimum di D1 sebesar 92,553. *Website 8* termasuk ke dalam *cluster 3* yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 10,045. *Website 9* termasuk ke dalam *cluster 3* juga yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 20,250. *Website 10* termasuk ke dalam *cluster 1* yaitu *cluster* gambar karena memiliki nilai minimum di D1 sebesar 33,931. *Website 11* termasuk ke dalam *cluster 3* yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 25,294. *Website 12* termasuk ke dalam *cluster 3* juga yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 26,419. *Website 13* termasuk ke dalam *cluster 3* yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 3,942. *Website 14* termasuk ke dalam *cluster 1* yaitu *cluster* gambar karena memiliki nilai minimum di D1 sebesar 60,113. *Website 15* termasuk ke dalam *cluster 3* yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 23,560. *Website 16* termasuk ke dalam *cluster 3* juga yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 11,762. *Website 17* termasuk ke dalam *cluster 3* yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 33,653. *Website 18* termasuk ke dalam *cluster 1* yaitu *cluster* gambar karena memiliki nilai minimum di D1 sebesar 36,542. *Website 19* termasuk ke dalam *cluster 1* juga yaitu *cluster* gambar karena memiliki nilai minimum di D1 sebesar 72,505. *Website 20* termasuk ke dalam *cluster 3* yaitu *cluster* video karena memiliki nilai minimum di D3 sebesar 21,290.

SIMPULAN

Dapat disimpulkan bahwa *big data* yang ada dalam *log server web* mengandung banyak informasi yang menghasilkan pengetahuan tentang *cluster* pengakses gambar/image, *cluster* pengakses teks, dan *cluster* pengakses audio/video. Pada pengujian *skenario 1* menghasilkan dimana para pengguna *web* lebih senang mengakses gambar di *web 7,14,18,19*. Sedangkan pada *web 1,3,5,10* para pengguna *web* lebih menyukai mengakses teks. Sedangkan pada *web 2,4,6,8,9,11,12,13,15,16,17,dan 20* para pengguna lebih suka menonton video-video. Hasil pengujian pada *skenario 2* yaitu para pengguna *web* lebih senang mengakses gambar di *web 1,3,5,7,10,14,18 dan 19*. Sedangkan pada *web 2* para pengguna *web* lebih menyukai mengakses teks. Sedangkan pada *web 4,6,8,9,11,12,13,15,16,17,dan 20* para pengguna lebih suka menonton video-video.

PENGHARGAAN

Ucapan terima kasih kepada lembaga LPPM Institut teknologi nasional Malang yang telah memberikan sponsor atau support terhadap berlangsungnya penelitian ini.

REFERENSI

- [1] Fadlan, C., S.Ningsih, & A.P.Windarto.2018.Penerapan Metode Naive Bayes Dalam Klasifikasi kelayakan Keluarga Penerima Beras Rastra.JUTIM.Vol.3 No.1.
- [2] Bedboudi A.,C.Bouras, & M.T.Kimour.2017.Indonesian Journal of Electrical Engineering and Informatics.Vol.5.No.3.ISSN:2089-3272, DOI: 10.1159/ijeei.v5i3.299
- [3] Ramdani, A.L., & H.B.Firmansyah.2018.Clustering Application for UKT Determination using Pillar K-means clustering Algorithm and Flask Web Framework.Indonesian Journal of Electrical Engineering and Informatics.Vol.1.No.2.p-ISSN:2614-3372.e-ISSN:2614-6150
- [4] Gustientiedina, M.H.Adiya, & Y.Desnelita.2019.Penerapan Algoritma K-means untuk Clustering data Obat-Obatan Pada RSUD Pekanbaru.Jurnal nasional Teknologi dan Sistem Informasi.Vol.5.No.1p-ISSN : 2460-3465.e-ISSN : 2476-8812
- [5] Darmansah, & N.W.Wardani.2020.Analisa Penyebab Kerusakan Tanaman Cabai menggunakan Metode K-means.Jurnal Teknik Informatika dan Sistem Informasi.Vol.7.No.2.ISSN:2407-4322.e-ISSN:2503-2933
- [6] Annur,H.2019.Penerapan Data mining menentukan Strategi Penjualan Variasi Mobil menggunakan Metode K-means Clustering.Jurnal Informatika UPGRIS.Vol.5.No.1.p/e ISSN: 2460-4801/2447-6645
- [7] Handoko,K, & L.S.Lesmana.2018.Pengelompokkan Data mining Pada Jumlah Penumpang Di Bandara Hang Nadim.CBIS Journal.Vol.6.No.2.p-ISSN: 2337-8794.e-ISSN: 2621-5292
- [8] Arista,R.R.,R.A.Asmara, &D.Puspitasari.2017.Pengelompokan Kejadian Gempa Bumi menggunakan Fuzzy C-Means Clustering. Jurnal teknologi Informasi dan Terapan.Vol.4No.2.ISSN: 235-838X
- [9] Novianto,R., & L.Geoirmanto. 2019.Penerapan data mining Menggunakan Algoritma K-means Clustering Untuk Menganalisa Bisnis Perusahaan Asuransi. Jurnal Teknik Informatika dan Sistem Informasi.Vo.6.No.1.ISSN: 2407-4322.e-ISSN: 2503-2933
- [10] Han,J, M.Kamber,& J.Pei.2012.Data mining Concepts and Techniques.Third edition.ISBN 978-0-12-381479-1
- [11] Liu.B.Web data mining, Exploring Hyperlinks,Contents and Usage data", 2 nd edition, Springer New York, ISBN:9783642194597,2011