



Penerapan Algoritma K-Means dan Linear Reggression Sederhana Dalam Klasterisasi Grafik Bandwidth

Muhammad Yasir¹, Fried Sinlae², Nurfiyah³

¹Universitas Bhayangkara Jakarta Raya, Bekasi, Indonesia, muhammad.yasir@dsn.ubharajaya.ac.id

²Universitas Bhayangkara Jakarta Raya, Bekasi, Indonesia, fried.sinlae@dsn.ubharajaya.ac.id

³Universitas Bhayangkara Jakarta Raya, Bekasi, Indonesia, nurfiyah@dsn.ubharajaya.ac.id

Corresponding Author: muhammad.yasir@dsn.ubharajaya.ac.id¹

Abstract: *The COVID-19 pandemic has altered activity patterns across all aspects of life today. The enactment of DKI Regulation number 14 of 2020 urging the DKI provincial government to implement Work from Home (WFH) policies to break the chain of COVID-19 transmission has transformed various activities such as schooling, work, commerce, and others to be conducted from home, primarily reliant on internet connections. Internet Service Providers (ISPs) have become crucial in supporting these activities. In monitoring user networks, a stable bandwidth is essential to ensure smooth operations. Analyzing bandwidth traffic is necessary as one approach for Quality of Service (QoS). Data mining utilizing algorithms like K-means and linear regression can assist in generating information for network quality management to cluster bandwidth. This study resulted in identifying low bandwidth clusters centered around 35.88 GB, medium bandwidth clusters centered around 43.15 GB, and high bandwidth clusters centered around 55.31 GB over the past four months from October 2020 to January 2021. Additionally, performance testing of the distance clustering using the Davies Bouldin index yielded a result of 0.240, along with volume traffic predictions for each bandwidth cluster.*

Keyword: *Clustering, K-Means, Simple Linear Regression, Bandwidth Traffic*

Abstrak: Pandemi covid-19 mengubah pola aktifitas di semua aspek kehidupan saat ini, Adanya Perda DKI nomor 14 tahun 2020 mengenai himbauan pemda DKI untuk *Work from Home* (WFH) agar memutuskan rantai covid 19 mengubah semua kegiatan baik aktifitas sekolah, bekerja, berdagang dan lain-lain dilakukan dari rumah dan tentunya menggunakan koneksi internet. *Internet Service Provider* (ISP) menjadi sangat di butuhkan dalam menunjang aktifitas tersebut, dalam monitoring jaringan user, dibutuhkan bandwith yang stabil agar semua berjalan dengan baik. Analisa terhadap suatu trafik bandwith di butuhkan sebagai salah satu cara untuk *Quality of Sevices* (QOS). Data mining dengan menggunakan algoritma *K-means* dan linear regression dapat membantu dalam membuat informasi *network quality management* untuk klasterisasi bandwith. Hasil dari penulisan ini didapatkan klaster bandwith low yang berada di *centroid* 35.88 GB, klaster bandwith medium berada di *centroid* 43.15 GB, dan

klaster bandwidth high di *centroid* 55.31 GB dalam 4 bulan terakhir dari oktober 2020 sampai januari 2021 dan juga dilakukan pengujian *performance distance clustering* menggunakan *davies bouldin index* didapatkan hasil 0.240 serta prediksi volume trafik trafik bandwidth per klaster.

Kata Kunci: Klasterisasi, K-Mean, Simple Linear Regression, Trafik Bandwith

PENDAHULUAN

Pandemik covid-19 mengubah pola aktifitas di semua aspek kehidupan saat ini, Adanya Perda DKI nomor 14 tahun 2020 mengenai himbauan pemda DKI untuk *Work From Home* (WFH) agar memutuskan rantai covid 19 mengubah semua kegiatan, baik aktifitas sekolah, bekerja, berdagang dan lain-lain semua dilakukan dari rumah dan tentunya menggunakan koneksi internet. *Internet Service Provider* (ISP) menjadi sangat di butuhkan dalam menunjang aktifitas *Work form Home* salah satu nya ISP INDONET. *Network Monitoring Center* di suatu ISP memiliki peran yang sangat penting dalam memonitoring jaringan pelanggan 24 jam, ISP INDONET mempunyai beberapa layanan koneksi internet seperti link, Wiring, Wireless, FO, VSAT, ALI CLOUD, SAMSUNG, HSX dan CCTV berbasis IOT serta bank-bank yang melakukan COB, EOD setiap malam dan EOM setiap bulannya. Urgensi nya, apabaila terjadi gangguan yang dikarenakan bandwidth pemakaian yang tinggi (*link peak*) menyebabkan koneksi intermitten/tidak stabil dan sangat kritikal dan berdampak besar ke semua pelanggan.

Penelitian sejenis telah dilakukan oleh Vijayakumar mereka menganalisis aliran trafik jaringan menggunakan teknik *Hierarchical Clustering Stream Network Traffic*. Data yang digunakan adalah data set trafik jaringan dari *Internet Service Provider* (ISP) dan pada tahapan *preprocessing* menggunakan Algoritma *K-means* untuk membentuk *cluster* trafik jaringan. Berdasarkan hasil yang diperoleh, mereka menyimpulkan bahwa perpaduan *Hierarchical Clustering Stream Network Traffic* dengan *K-means* dapat meningkatkan akurasi hasil *cluster* (Vijayakumar, 2010).

Pada tahun 2011, Arya dan Mishra juga melakukan penelitian yang sejenis. Mereka mengevaluasi performa lima Algoritma seperti *J48*, *Random Tree*, *Random Forest*, *Bagging* dan *Boosting Classifiers* untuk klasifikasi *dataset benchmark*. Tingkat akurasi, presisi, dan *recall* digunakan untuk membandingkan performa dari kelima classifiers. Hasil penelitian menunjukkan bahwa multilevel *classifier* lebih akurat dalam proses klasifikasi jaringan internet (Arya & Mishra, 2011).

Özyirmidokuz, dkk. juga melakukan penelitian untuk menemukan pengetahuan yang berguna dari data trafik jaringan. Mereka menganalisis data trafik jaringan 150Mbps yang menghubungkan antara AS dan Jepang menggunakan metode *CART decision tree*. Penelitian ini menganalisis deteksi anomaly dan menerapkan *Kohonen Networks* untuk mengurangi data matriks. Hasil penelitian menunjukkan bahwa metode *CART decision tree* berhasil diterapkan dalam analisis lalu lintas jaringan dan memiliki kinerja yang cukup baik (Kahya-Özyirmidokuz et al., 2012).

Pengelolaan trafik bandwidth penting dilakukan karena dapat membantu *network monitoring center* (NOC) mengalokasi bandwidth kepada pengguna dan aplikasi pendukung berdasarkan prioritas tertentu dan sebagai salah satu cara untuk *Quality of Sevices* (QOS). Dalam fokus penelitian ini adalah dilakukan klasterisasi bandwidth trafik menggunakan algoritma *K-means & Liner Reggresion* yang bertujuan untuk mengklasterisasi pemakaian bandwidth dalam 3 klaster *low*, *medium* dan *high* untuk memberikan informasi ke *Internet service provider* sebagai bagian dari *Quality of services* (QOS) serta prediksi trafik sehingga meminimailsasi koneksi *peak*

KAJIAN PUSTAKA

Klasifikasi Network Trafik

Metode yang pada dasarnya digunakan untuk mengklasifikasikan lalu lintas internet dapat dikategorikan sebagai berikut (Arya & Mishra, 2011):

Port Number Analysis

Nomor Port dalam header TCP dan UDP biasanya sesuai dengan nomor port default. Jika terjadi kecocokan, paket tersebut dapat diidentifikasi sebagai aplikasi P2P, meskipun sistem P2P yang lebih baru dapat menggunakan nomor port secara acak. Sebagai contoh, aplikasi P2P seperti KaZaA, LimeWire, dan sejenisnya mungkin menggunakan nomor port yang tidak terdaftar di IANA (Internet Assigned Numbers Authority). Oleh karena itu, pendekatan analisis berdasarkan port tidak selalu efektif untuk mengklasifikasikan lalu lintas Internet.

Payload Based Analysis

Teknik ini hanya mengidentifikasi lalu lintas P2P yang mendeteksi string unik yang ditemukan dalam payload paket data. Pendekatan deteksi ini membutuhkan proses yang intensif, melakukan inspeksi paket mendalam (yang tidak akan mungkin dilakukan jika privasi diperlukan). Kedua, Beberapa aplikasi P2P memiliki data terenkripsi membuatnya tidak mungkin untuk mendeteksi string unik dalam payload. Analisis payload membutuhkan banyak daya komputasi karena menganalisis muatan penuh paket. Ketiga, teknik ini sering kali membutuhkan peningkatan kapasitas pemrosesan dan penyimpanan.

Algoritma K-means

Algoritma K-means merupakan salah satu *unsupervised learning*, dimana metode kluster analisis non hirarki yang berusaha untuk memetakan objek yang ada kedalam satu atau lebih kluster atau kelompok objek berdasarkan karakteristiknya (Agus Nur Khormarudin, 2016). Berikut tahap algoritma K-means:

1. Pilih secara acak k buah data sebagai pusat cluster.
2. Jarak antara dan pusat cluster dihitung menggunakan Euclidian Distance. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

$$D(i, j) = \sqrt{(x_{1i} + x_{1j})^2 + (x_{2i} + x_{2j})^2 + \dots + (x_{ki} + x_{kj})^2}$$

Dimana:

D (i,j) = Jarak data ke i ke pusat cluster j

Xki = Data ke i pada atribut data ke k

Xkj = Titik pusat ke j pada atribut ke k

3. Data ditempatkan dalam cluster yang terdekat, dihitung dari tengah cluster
4. Pusat cluster baru akan ditentukan bila semua data telah ditetapkan dalam cluster terdekat
5. Proses penentuan pusat cluster dan penempatan data dalam cluster diulangi sampai nilai *centroid* tidak berubah lagi

Linear Regression

Analisis regresi adalah suatu metode statistik yang mengamati hubungan antara variabel terikat Y dan serangkaian variabel bebas X_1, \dots, X_p . Untuk memprediksi variabel Y yang merupakan variabel terikat dari variabel X yang diberikan (Hijriani et al., 2016). Rumus linear regresi sederhana adalah $Y = c + mX$. Dimana Y merupakan Variabel akibat (variabel terikat), X merupakan Variabel penyebab (variabel bebas), m merupakan konstanta dan c merupakan koefisien regresi (kemiringan). Untuk nilai m dan c akan dihitung menggunakan rumus dimana m merupakan banyak data (jumlah X * Y) – (jumlah X) (jumlah Y) / banyak data (jumlah X *

$(X - (\text{jumlah } X)^2 / \text{banyak data}) - c$ merupakan $(\text{jumlah data } y) - (\text{jumlah data } X * X) - (\text{jumlah data } X) / (\text{jumlah data } X * Y) / \text{banyak data} - (\text{jumlah data } X * X) - (\text{jumlah data } X)^2$

Langkah-langkah melakukan perhitungan dalam regresi linear sederhana adalah sebagai berikut:

- a. Identifikasi variabel penyebab dan akibat,
- b. Siapkan data-data yang akan dijadikan variabel terikat (Y) dan bebas (X), untuk variabel terikat (Y) data yang digunakan adalah data mahasiswa yang masuk ke universitas Samudra selama 5 tahun sebelum tahun yang akan prediksi.
- c. Hitung nilai untuk X^2 (variabel X * variabel X), Y^2 (variabel Y * variabel Y), XY (variabel X * variabel Y) dan cari jumlah dari masing-masing variabel tersebut
- d. Lakukanlah perhitungan untuk nilai m dan c dengan mengikuti rumus yang terdapat diatas,
- e. Bentuklah sebuah model yang digunakan untuk persamaan regresi linear sederhana.

Davies Bouldin Index

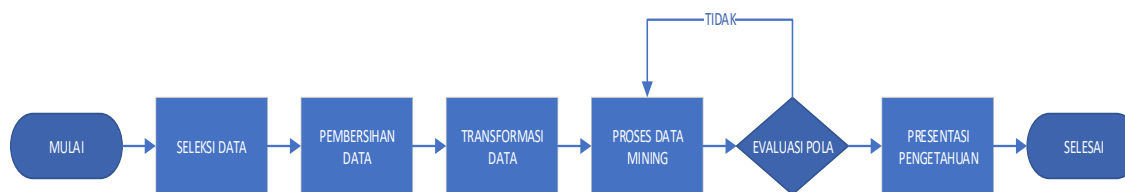
Indeks validitas *Davies-Bouldin* (DB) menghitung rata-rata nilai setiap titik pada himpunan data. Perhitungan nilai setiap titik adalah jumlah nilai compactness yang dibagi dengan jarak antara kedua titik pusat kluster sebagai separation. Jumlah kluster terbaik ditunjukkan dengan nilai DB yang semakin kecil (Davies & Bouldin, 1979) (Khairati et al., 2019)

Rapid Miner

Rapid Miner adalah perangkat lunak *open source* untuk analisis data yang dulunya dikenal sebagai YALE. Dikembangkan pertama kali pada 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di University of Dortmund. Perangkat lunak ini berlisensi AGPL (GNU Affero General Public License) versi 3 dan telah digunakan dalam ribuan aplikasi di lebih dari 40 negara. (Aprilla Dennis, 2013)

METODE

Data mining adalah proses mengekstrak atau menambang pengetahuan dari volume data yang besar atau jumlah data yang luas untuk mengubahnya menjadi informasi yang dapat dikembangkan menjadi pengetahuan baru. (Aprilla Dennis, 2013)



Gambar 1 Tahapan Data Mining

Berikut tahapan Data mining yang diilustrasikan dalam gambar 1 sebagai berikut (Fadli, 2003):

- a. Pembersihan Data: Proses untuk menghilangkan data yang tidak konsisten (*noise*) dan mengintegrasikan data dari berbagai sumber yang terpisah.
- b. Seleksi Data: Proses di mana data-data yang relevan untuk analisis dikembalikan atau dihilangkan dari database.
- c. Transformasi Data: Proses di mana kumpulan data yang relevan diubah atau digabungkan ke dalam format yang tepat sehingga memudahkan untuk dilakukan penambangan data
- d. Proses Mining: Proses di mana metode analisis atau algoritma kecerdasan digunakan untuk mengekstrak pola data.
- e. Evaluasi Pola: Proses untuk mengidentifikasi pola yang benar-benar menarik, mewakili

- pengetahuan berdasarkan beberapa tindakan yang menarik.
- f. Presentasi Pengetahuan: Proses di mana teknik visualisasi dan pengetahuan digunakan untuk menyajikan pengetahuan yang telah ditambang kepada pengguna.

HASIL DAN PEMBAHASAN

A. Data Cleansing / Pembersihan Data

Dalam mendukung penelitian ini, penulis mendapatkan data trafik bandwidth menggunakan tools dari PRTG *Network Monitoring* yang mengambil data trafik dari perangkat Switch port 1 di *server data center* kantor ISP INDONET ke dalam bentuk dataset berformat CSV. Di tahap ini penulis menghilangkan noise data yang tidak konsisten dan data yang diperlukan dalam penelitian. Berikut hasil dari cleansing data:

Date Time	Traffic Total (volume)	Traffic Total (speed)	Traffic In (volume)	Traffic In (speed)	Traffic Out (volume)	Traffic Out (speed)	Downtime	Coverage
10/6/2020 12:00:00 AM - 205,632 KB	468 kbit/s	176,087 KB	401 kbit/s	29,545 KB	67 kbit/s	0%	100%	
10/6/2020 1:00:00 AM - 211,910 KB	482 kbit/s	185,032 KB	421 kbit/s	26,877 KB	61 kbit/s	0%	100%	
10/6/2020 2:00:00 AM - 233,641 KB	532 kbit/s	207,347 KB	472 kbit/s	26,294 KB	60 kbit/s	0%	100%	
10/6/2020 3:00:00 AM - 207,859 KB	473 kbit/s	177,252 KB	403 kbit/s	30,607 KB	70 kbit/s	0%	100%	
10/6/2020 4:00:00 AM - 207,069 KB	471 kbit/s	177,430 KB	404 kbit/s	29,638 KB	67 kbit/s	0%	100%	
10/6/2020 5:00:00 AM - 202,935 KB	462 kbit/s	176,661 KB	402 kbit/s	26,274 KB	60 kbit/s	0%	100%	
10/6/2020 6:00:00 AM - 225,269 KB	513 kbit/s	198,963 KB	453 kbit/s	26,306 KB	60 kbit/s	0%	100%	
10/6/2020 7:00:00 AM - 461,431 KB	1,050 kbit/s	435,766 KB	992 kbit/s	25,666 KB	58 kbit/s	0%	100%	
10/6/2020 8:00:00 AM - 487,431 KB	1,109 kbit/s	460,644 KB	1,048 kbit/s	26,787 KB	61 kbit/s	0%	100%	
10/6/2020 9:00:00 AM - 390,283 KB	888 kbit/s	358,500 KB	816 kbit/s	31,782 KB	72 kbit/s	0%	100%	
10/6/2020 10:00:00 AM - 399,221 KB	908 kbit/s	377,960 KB	860 kbit/s	21,261 KB	48 kbit/s	0%	100%	
10/6/2020 11:00:00 AM - 595,072 KB	1,354 kbit/s	570,356 KB	1,298 kbit/s	24,716 KB	56 kbit/s	0%	100%	
10/6/2020 12:00:00 PM - 463,775 KB	1,055 kbit/s	434,933 KB	990 kbit/s	28,842 KB	66 kbit/s	0%	100%	
10/6/2020 1:00:00 PM - 532,765 KB	1,212 kbit/s	510,990 KB	1,163 kbit/s	21,775 KB	50 kbit/s	0%	100%	
10/6/2020 2:00:00 PM - 364,780 KB	830 kbit/s	329,625 KB	750 kbit/s	35,155 KB	80 kbit/s	0%	100%	
10/6/2020 3:00:00 PM - 401,173 KB	913 kbit/s	378,438 KB	861 kbit/s	22,734 KB	52 kbit/s	0%	100%	
10/6/2020 4:00:00 PM - 307,413 KB	700 kbit/s	278,131 KB	633 kbit/s	29,282 KB	67 kbit/s	0%	100%	
10/6/2020 5:00:00 PM - 396,717 KB	903 kbit/s	372,358 KB	847 kbit/s	24,359 KB	55 kbit/s	0%	100%	
10/6/2020 6:00:00 PM - 426,006 KB	969 kbit/s	402,877 KB	917 kbit/s	23,129 KB	53 kbit/s	0%	100%	
10/6/2020 7:00:00 PM - 1700,366 KB	1,594 kbit/s	673,119 KB	1,532 kbit/s	27,247 KB	62 kbit/s	0%	100%	
10/6/2020 8:00:00 PM - 1227,199 KB	517 kbit/s	204,523 KB	465 kbit/s	22,676 KB	52 kbit/s	0%	100%	
10/6/2020 9:00:00 PM - 199,803 KB	455 kbit/s	178,288 KB	406 kbit/s	21,515 KB	49 kbit/s	0%	100%	

Gambar 2 Tahap Cleansing Data

B. Data Selection / Seleksi Data

Dalam tahap ini dilakukan proses seleksi data. Untuk data yang diseleksi akan ditransformasikan ke format yang sesuai untuk analisis data. Seleksi data menggunakan beberapa kriteria, dalam tahap ini penulis mengelompokkan data trafik bandwidth per minggu.

C. Data Transformation / Transformasi Data

Dalam tahap ini, Proses dimana kumpulan data yang relevan diubah atau disatukan kedalam bentuk format yang sudah tepat untuk dilakukan penambangan. Pada tahap ini, dilakukan normalisasi atribut, Normalization digunakan untuk mentransformasi sebuah atribut numerik diskalakan dalam range yang lebih kecil (Junaedi et al., 2011). Berikut Transformasi data:

TANGGAL	MINGGU	TRAFFIC IN (GB)	TRAFFIC OUT (GB)	TOTAL VOLUME TRAFFIC (GB)
6 OKT - 12 OKT	1	43	4,2	47,2
13 OKT - 19 OKT	2	54,3	4,25	58,55
20 OKT - 26 OKT	3	48,6	4,1	52,7
27 OKT - 1 NOV	4	39,2	3,85	43,05
2 NOV - 8 NOV	5	59,6	3,2	62,8
9 NOV - 15 NOV	6	33,2	3,3	36,5
16 NOV - 22 NOV	7	28,1	3,5	31,6
23 NOV - 29 NOV	8	34,4	3,32	37,6
30 NOV - 6 DES	9	33	3,2	36,2
7 DES - 13 DES	10	27,8	3,1	30,9
14 DES - 20 DES	11	33,6	4,1	37,7
21 DES - 27 DES	12	37,2	3,7	40,9
28 DES - 3 JAN	13	37,3	4,3	41,6
4 JAN - 10 JAN	14	33,1	3,89	36,99
11 JAN - 17 JAN	15	43,1	4,2	47,3
18 JAN - 24 JAN	16	39,2	3,7	42,9
25 JAN - 31 JAN	17	32,6	3,7	36,3
1 FEB - 7 FEB	18	35,5	3,7	39,2

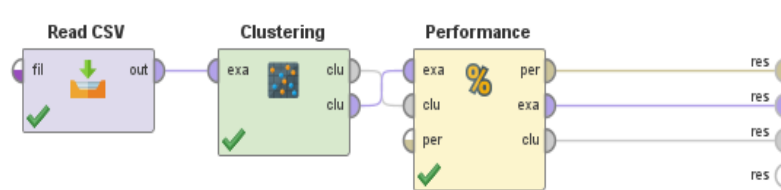
Gambar 3 Tahap Transformasi Data

D. Proses Data Mining

Dalam Tahap ini merupakan proses dimana metode analisis atau algoritma intelejen digunakan untuk mengekstrak pola data. Sebagai berikut:

Klasterisasi bandwidth

Untuk melakukan klasterisasi bandwidth penulis menggunakan metode algoritma K-means. Aplikasi dari rapidminer yang akan mengelompokkan bandwidth menjadi 3 kluster yaitu, kluster bandwidth low, kluster bandwidth medium, dan kluster bandwidth high dengan hasil *performance distance cluster* menggunakan davies bouldin index sebesar 0,240. Tujuannya adalah untuk mengetahui standart dari kebutuhan pemakaian bandwidth ISP indonet per minggu.



Gambar 4 Klasterisasi menggunakan Rapidminer



Gambar 5 Hasil Klasterisasi menggunakan Rapid miner

Linear Regresi Sederhana

Dalam tahap ini dilakukan prediksi trafik volume bandwidth menggunakan metode linear regression. Dari proses klasterisasi k-means terdapat 3 klaster, dimana dari 3 klaster kita lakukan proses peramalan dengan metode linear regression sederhana pada disetiap klaster nya dan yang selanjutnya dilakukan perhitungan error pada prediksi.

Table 1 Data Cluster 0

No	x	Y	Xy	x2
1	1	47.2	47.2	1
2	2	58.55	117.1	4
3	3	52.7	158.1	9
4	4	62.8	251.2	16
Jumlah	10	221.25	573.6	30
Rata - rata	2.5	55.3125		

Sumber Hasil Riset

$$c = (221.25)(30) - (10)(573.6) / 4(30) - (100)$$
$$c = 45.075$$

$$m = 4(573.6) - (10)(221.25) / 4(30) - (100)$$
$$m = 4.095$$

$$Y = c + mX$$

$$Y = 45.075 + 4.095 (4)$$

$$Y = 60.455 \text{ untuk data selanjutnya atau data ke 5 maka } y \text{ (total traffic volume)} = 60.455$$

Table 2 Data Cluster 1

No	x	Y	Xy	x2
1	1	36.5	36.5	1
2	2	31.6	63.2	4
3	3	37.6	112.8	9
4	4	36.2	144.8	16
5	5	30.9	154.5	25
6	6	37.7	226.2	36
7	7	36.99	258.93	49
8	8	36.3	290.4	64
9	9	39.2	352.8	81
Jumlah	45	322.99	1640.13	285
Rata-rata	5	35.887		

Sumber Hasil Riset

$$c = (322.99)(285) - (45)(1640.13) / 9(285) - (2025)$$
$$c = 33.78944444$$

$$m = 9(1640.13) - (45)(322.99) / 9(285) - (2025)$$
$$m = 0.419666667$$

$$Y = c + mX$$

$$Y = 33.789 + 0.419(9)$$

$Y = 37.560$ untuk data selanjutnya atau data ke 10 maka y (total traffic volume) = 37.560

Table 3 Data Cluster 3

No	x	Y	xy	x ²
1	1	43.05	43.05	1
2	2	40.9	81.8	4
3	3	41.6	124.8	9
4	4	47.3	189.2	16
5	5	42.9	214.5	25
Jumlah	15	215.75	653.35	55
Rata-rata	3	43.15		

Sumber Hasil Riset

$$c = (215.75)(55) - (15)(653.35) / 5(55) - (225)$$

$$c = 41.32$$

$$m = 5(653.35) - (15)(215.75) / 5(55) - 225$$

$$m = 0.61$$

$$Y = c + mX$$

$$Y = 41.32 + 0.61(5)$$

$$Y = 44.37 \text{ untuk data selanjutnya atau data ke 6 maka } y \text{ (total traffic volume) } = 44.37$$

E. *Evaluation Pattern/ Evaluasi Pola*

Dalam tahap ini merupakan proses untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan atas beberapa tindakan yang menarik untuk di jadikan *knowledge presentation*.

F. *Knowledge Presentation / Presentasi Pengetahuan*

Dalam Tahap ini, merupakan proses merpresentasikan hasil pengolahan data mining yang sudah diolah. Representasi dari proses mining di atas adalah pengelompokan bandwidth berdasarkan jumlah volume bandwidth per minggu menjadi 3 klaster yaitu, klaster bandwidth low yang berada di *centroid* 35.88 GB, klaster bandwidth medium berada di *centroid* 43.15 GB, dan klaster bandwidth high di *centroid* 55.31 GB dengan hasil *performance distance cluster* menggunakan *davies bouldin index* sebesar 0,240. Tujuannya adalah untuk mengetahui standart dari kebutuhan pemakaian bandwidth ISP indonet per minggu. Kemudian penulis melakukan prediksi menggunakan linear regression, hasil dari olahan perhitungan manual seperti penjelasan di atas

KESIMPULAN

Penelitian ini menjawab pentingnya pengelolaan trafik bandwidth karena dapat membantu *network monitoring center* (NOC) mengetahui alokasi bandwidth kepada pengguna dan aplikasi pendukung berdasarkan prioritas tertentu dan sebagai salah satu cara untuk *Quality of Sevices* (QOS). Dalam peneltian didapatkan hasil pengelompokan bandwidth berdasarkan jumlah volume bandwidth per minggu menjadi 3 klaster yaitu, klaster bandwidth low yang berada di *centroid* 35.88 GB, klaster bandwidth medium berada di *centroid* 43.15 GB, dan klaster bandwidth high di *centroid* 55.31 GB dengan hasil *performance distance cluster* menggunakan *davies bouldin index* sebesar 0,240 serta melakukan prediksi menggunakan linear regression, hasil dari olahan tersebut di dapatkan prediksi volume trafik per klaster

REFERENSI

- Agus Nur Khormarudin. (2016). Teknik Data Mining: Algoritma K-Means Clustering. *Jurnal Ilmu Komputer*, 1–12. <https://ilmukomputer.org/category/datamining/>
- Aprilla Dennis. (2013). Belajar Data Mining dengan RapidMiner. *Innovation and Knowledge Management in Business Globalization: Theory & Practice, Vols 1 and 2*, 5(4), 1–5. http://esjournals.org/journaloftechnology/archive/vol1no6/vol1no6_6.pdf%5Cnhttp://www.airccse.org/journal/nsa/5413nsa02.pdf
- Arya, I. B., & Mishra, R. (2011). *Internet Traffic Classification: An Enhancement in Performance using Classifiers Combination*. 2(2), 663–667.
- Fadli, A. (2003). Konsep Data Mining. *Konsep Data Mining*, 1–9. <http://fadli84.wordpress.com>
- Hijriani, A., Muludi, K., & Andini, E. A. (2016). Implementasi Metode Regresi Linier Sederhana Pada Penyajian Hasil Prediksi Pemakaian Air Bersih Pdam Way Rilau Kota Bandar Lampung Dengan Sistem Informasi Geografis. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 11(2), 37. <https://doi.org/10.30872/jim.v11i2.212>
- Junaedi, H., Budianto, H., Maryati, I., & Melani, Y. (2011). Data Transformation pada Data Mining. *Prosiding Konferensi Nasional Inovasi Dalam Desain Dan Teknologi-IDEaTech*, 7, 93–99.
- Kahya-Özyirmidokuz, E., Gezer, A., & Ciflikli, C. (2012). Characterization of Network Traffic Data: A Data Preprocessing and Data Mining Application. *International Conference on Data Analytics, September*, 18–23.
- Khairati, A. F., Adlina, A. A., Hertono, G. F., & Handari, B. D. (2019). Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA. *Prosiding Seminar Nasional Matematika*, 2, 161–170.
- Vijayakumar, M. (2010). *Clustering of High-Volume Data Streams In Network Traffic*. 8(6), 228–233.