

Segmentasi Pengalaman Pelanggan Pengguna Aplikasi *Youtube Streaming* Pada Daerah Jakarta Selatan Menggunakan Metode *K-Means* (Studi Kasus PT. XL AXIATA Tbk.)

Patria Adhistian

Program Studi Teknik Industri, Fakultas Teknik, Universitas Pamulang,
Tangerang Selatan, Banten
patria.adhistian@yahoo.com

Abstract

YouTube is one of the largest video sharing sites on the internet. More than 1.9 billion users per month visit YouTube. 70% of total YouTube watch time comes from mobile phones. Therefore, as a network service provider, PT XL Axiata Tbk. intends to improve the quality of service and quality of customer experience by targeting high value customers. One way is to track the customer experience of watching videos on the internet, one of which is on YouTube. Segmentation is one way to find and identify the quality of the experience that customers receive. To choose the right quality problem handling, service providers need to look at network service quality parameters, such as throughput, delay, and latency. Segmentation is carried out using a data mining approach, to be precise with the K-means clustering technique. The working method follows the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. Through the data preparation stage, 4 important attributes are obtained for modeling. Determination of the optimal number of clusters using the Elbow Method shows that there are 3 clusters formed, with cluster 2 as the best cluster, and cluster 1 as the most dominant cluster of users. This research can be used for PT XL Axiata Tbk. as input in the quality of network services for customers.

Keywords: *Data mining, CRISP-DM, clustering, K-means, QoS parameters*

Abstrak

YouTube adalah salah satu situs berbagi video terbesar di internet. Lebih dari 1,9 miliar pengguna per bulan mengunjungi YouTube. Sebesar 70% dari total waktu tonton YouTube berasal dari ponsel. Maka dari itu, sebagai penyedia layanan jaringan, PT XL Axiata Tbk. berniat untuk meningkatkan kualitas layanan dan kualitas pengalaman pelanggan dengan menargetkan high value customer. Salah satu caranya adalah melacak pengalaman pelanggan menonton video di internet, salah satunya di YouTube. Segmentasi menjadi salah satu cara dalam mencari dan mengidentifikasi kualitas pengalaman yang pelanggan terima. Untuk memilih penanganan masalah kualitas yang tepat, maka penyedia layanan perlu melihat parameter kualitas layanan jaringan, seperti throughput, delay, dan latency yang ada. Segmentasi dilakukan dengan pendekatan data mining, tepatnya dengan teknik clustering metode K-means. Metode pengerjaan mengikuti metodologi Cross Industry Standard Process for Data Mining (CRISP-DM). Melalui tahap Persiapan Data, didapat 4 atribut yang penting untuk pemodelan. Penentuan jumlah cluster optimal dengan Metode Elbow memberikan hasil bahwa ada 3 cluster yang terbentuk, dengan cluster 2 sebagai cluster terbaik, dan cluster 1 sebagai cluster yang jumlah penggunanya paling dominan. Penelitian ini dapat dimanfaatkan untuk PT XL Axiata Tbk. sebagai masukan dalam kualitas layanan jaringan untuk pelanggan.

Kata kunci: Data mining, CRISP-DM, clustering, K-means, Parameter QoS

1. PENDAHULUAN

Tujuan utama dalam setiap usaha bisnis adalah memaksimalkan profit. Terutama dalam bidang penyedia jasa layanan komunikasi (*Communication Service Provider*). Pada penelitian yang dilakukan oleh J. Spiess et

al.[1], CSP dapat meningkatkan keuntungan dengan meningkatkan Nilai Umur Pelanggan (*Customer Lifetime Value*), dengan meningkatkan kualitas pengalaman pelanggan (*customer experience*). Salah satu cara untuk meningkatkan kualitas pengalaman pelanggan

adalah *Tracking Video Experience*. Pantau *customer experience*, dan mengoptimalkan pengiriman video di seluruh jaringan dari sumber ke *end user*.

Pada referensi [2] disebutkan bahwa lebih dari 2 miliar pengguna mengunjungi *YouTube* setiap bulannya, dan lebih dari 70% waktu tonton *YouTube* berasal dari ponsel. Maka, peneliti mengambil *use case* tentang *tracking video experience* untuk aplikasi *YouTube*. Untuk mengidentifikasi pengalaman pelanggan yang beragam, segmentasi dipilih sebagai cara yang tepat. Sebagai strategi *customer experience* [3], segmentasi menyediakan cara sistematis dalam menjangkau orang yang berbeda dengan pesan yang berbeda yang dirancang khusus untuk mereka, sehingga dapat meningkatkan pelayanan kepada pelanggan secara proaktif.

Pada referensi [4] menyatakan bahwa segmentasi pelanggan diperlukan untuk mengelompokkan pelanggan yang memiliki kesamaan karakteristik. Penggunaan teknik *data mining* diharapkan membantu proses pengambilan keputusan yang tepat, salah satu teknik yang dipakai adalah *clustering*. Teknik ini [5] mencari pengamatan yang berdekatan sehingga pengamatan yang berada pada *cluster* yang sama lebih serupa satu sama lain daripada pengamatan yang berada di *cluster* yang lain. Salah satu metode *clustering* yang populer adalah *K-means*. Pendekatan *K-means* mengklasifikasi data tertentu yang ditetapkan melalui sejumlah *cluster*, dengan mendefinisikan nilai pusat *cluster* (*centroid*), satu untuk setiap *cluster* [5].

Pada referensi [6] menyatakan bahwa *provider* harus dapat bereaksi secara tepat dalam masalah kualitas. Untuk memilih tindakan yang tepat, *provider* perlu tahu bagaimana parameter *Quality of Service* pada kualitas pengalaman pelanggan. *Quality of Service* (QoS) adalah teknik untuk mengelola *throughput*, *delay*, *jitter*, dan *packet loss* untuk aliran paket dalam jaringan [7]. Peneliti juga menambahkan atribut tambahan seperti nomor *cell*, waktu pemakaian, perangkat yang digunakan, *apps* id, dan protokol id.

Peneliti juga membatasi area penelitian, yaitu di Jakarta Selatan saja. Peneliti juga membatasi kegiatan penggunaan *YouTube* hanya pada *streaming* saja. Atribut yang dipakai dalam penelitian ini juga mengambil atribut parameter QoS dan atribut jumlah pengguna. Maka dari itu, untuk mengetahui

hasil segmentasi pengalaman pelanggan di PT XL Axiata Tbk. peneliti mengambil penelitian berjudul, segmentasi pengalaman pelanggan pengguna aplikasi *YouTube streaming* pada daerah Jakarta Selatan menggunakan metode *K-means*.

Penelitian-penelitian terdahulu yang terkait dalam segmentasi pengalaman pelanggan adalah Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik *Clustering* [4], Analisa *Quality of Service* (QoS) Jaringan Internet Kampus (Studi Kasus: UIN Suska Riau) [7], Algoritma *Cluster* Dinamik untuk Optimasi *Cluster* pada Algoritma *K-means* dalam Pemetaan Nasabah Potensial [8], Segmentasi Nasabah Tabungan Menggunakan Model RFM (Recency, Frequency, Monetary) dan *K-means* pada Lembaga Keuangan Mikro [9], *Quality of Experience Key Metrics Framework for Network Mobility User* [10].

Pada penelitian ini juga merujuk pada tabel kecepatan internet yang disarankan oleh *YouTube* [11] seperti pada Tabel 1.

Tabel 1. Tabel Kecepatan Internet yang Disarankan

| Resolusi Video | Kecepatan Berkelanjutan yang Direkomendasikan |
|----------------|---|
| 4K | 20 Mbps |
| HD 1080p | 5 Mbps |
| HD 720p | 2,5 Mbps |
| SD 480p | 1,1 Mbps |
| SD 360p | 0,7 Mbps |

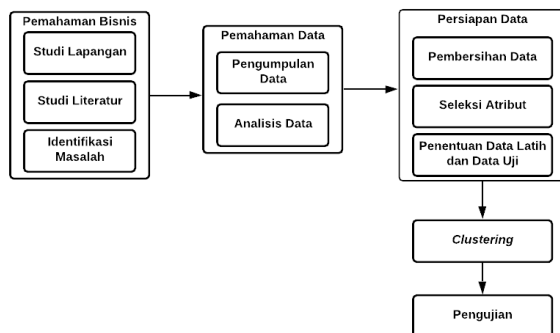
Pada referensi [12], *delay* terbagi menjadi empat kategori seperti pada Tabel 2.

Tabel 2. Tabel Kategori *Delay*

| Kategori | Delay |
|--------------|-----------------|
| Sangat Bagus | <150ms |
| Bagus | 150ms s/d 300ms |
| Sedang | 300ms s/d 450ms |
| Buruk | >450ms |

2. METODE PENELITIAN

Tahapan penelitian dibagi ke dalam beberapa tahap seperti Gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan penelitian mengikuti metodologi Cross Industry Standard Process for Data Mining (CRISP-DM), dengan langkah-langkah: Pemahaman Bisnis, Pemahaman Data, Persiapan Data, *Clustering*, dan Pengujian.

2.1. Pemahaman Bisnis

Tahap ini terbagi menjadi tiga kegiatan yang semuanya bersama-sama membangun latar belakang dari penelitian ini:

a. Studi Lapangan

Studi lapangan dilakukan mengenai pengalaman pelanggan tentang apa sebaiknya yang dapat diangkat sebagai permasalahan.

b. Studi Literatur

Peneliti mencari penelitian-penelitian sebelumnya tentang topik-topik yang masih berkaitan dengan segmentasi pelanggan, *clustering*, dan parameter kualitas layanan jaringan, dan pengalaman pelanggan.

c. Identifikasi Masalah

Peneliti membuat latar belakang juga mengidentifikasi masalah yang ditemukan.

2.2. Pemahaman Data

Tahap ini terbagi menjadi dua kegiatan dalam usaha untuk memahami data yang diterima:

a. Pengumpulan Data

Pengumpulan data dilakukan dengan menerima data yang diberikan perusahaan yang sesuai dengan arahan dari pihak perusahaan.

b. Analisis Data

Data dianalisis untuk dapat dipahami tentang atribut, tipe data, dan hal lain yang dapat didapatkan.

2.3. Persiapan Data

Tahap ini terbagi menjadi tiga tahapan untuk menyiapkan data yang akan dimodelkan yaitu:

a. Pembersihan Data

Pada tahap ini data akan dibersihkan dari data-data kosong, dan data-data tidak konsisten

b. Seleksi Atribut

Pada tahap ini akan dicari atribut mana yang akan dipakai untuk membuat model segmentasi. Tahap ini dilakukan sebelum tahap Penentuan Data Latih dan Data Uji agar data uji mempunyai atribut yang sama dengan data latih yang digunakan untuk membuat model segmentasi.

c. Penentuan Data Latih dan Data Uji

Penentuan data latih dan data uji dilakukan untuk memisahkan dataset dengan pembagian yang sudah ditentukan untuk proses pembuatan model segmentasi dan proses pengujian.

2.4. Clustering

Pada tahap ini, dimulai proses memodelkan data untuk mengetahui informasi apa yang dapat diambil dari data. Jika model sudah didapat maka diuji dengan data uji.

2.5. Pengujian

Pada tahap ini, hasil dari segmentasi diuji dengan metode *Sum of Squared Error* (SSE) untuk mengetahui jumlah K yang paling optimal.

3. HASIL DAN PEMBAHASAN

3.1. Pemahaman Data

Dalam pemahaman data, terdapat dua bagian yaitu pengumpulan data dan analisis data.

3.1.1. Pengumpulan Data

Pada pengumpulan data, peneliti menerima data sekunder dengan jumlah baris 238.775 dari perusahaan XL Axiata Tbk. Data mempunyai sepuluh atribut dengan keterangan seperti pada Tabel 3.

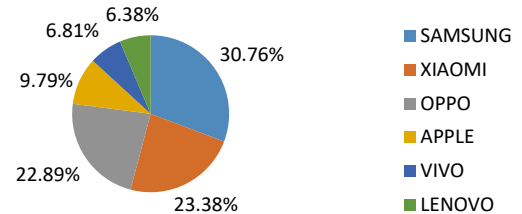
Tabel 3. Atribut Dataset dan Keterangan

| Nama Atribut | Keterangan |
|----------------|--|
| date_id | Tanggal pelanggan mengakses <i>YouTube streaming</i> |
| cgisai | Berisi nomor cell |

| | | |
|------------------------|--------------------|---|
| | identity | yang menunjukkan area pelanggan yang mengakses <i>YouTube</i> |
| duration | Lamanya | waktu pelanggan mengakses <i>YouTube</i> |
| device | Merek perangkat | yang digunakan pelanggan |
| thruputttpskbps | Berisi throughput | kilo bit per detik yang dialami pelanggan |
| tcp_latency | Berisi latensi | yang terjadi saat pelanggan mengakses <i>YouTube</i> |
| delay | Berapa lama delay | yang terjadi saay pengguna menonton video |
| prot_type | Nomor identifikasi | untuk aplikasi |
| prot_category | Nomor identifikasi | untuk kegiatan <i>streaming</i> |
| numberofuser | Jumlah pelanggan | yang mengakses <i>YouTube streaming</i> |

atribut *device*, merek perangkat yang digunakan terlalu banyak maka peneliti membatasi hanya menampilkan merek yang digunakan lebih dari 25.000 orang.

Merek Perangkat yang Paling Banyak Digunakan Pelanggan dalam Mengakses *YouTube*



*Gambar 3. Grafik Merek Perangkat yang Paling Banyak Digunakan dalam Mengakses *YouTube**

Pada Gambar 3, dapat dilihat bahwa merek yang paling banyak digunakan adalah Samsung dengan persentase sebesar 30,76%. Sedangkan merek yang hampir mendekati batas jumlah pengguna adalah Lenovo dengan persentase 6,38%.

3.1.2. Analisis Data

Pada tahapan analisis data, peneliti mencoba menggali informasi yang ada di dataset.



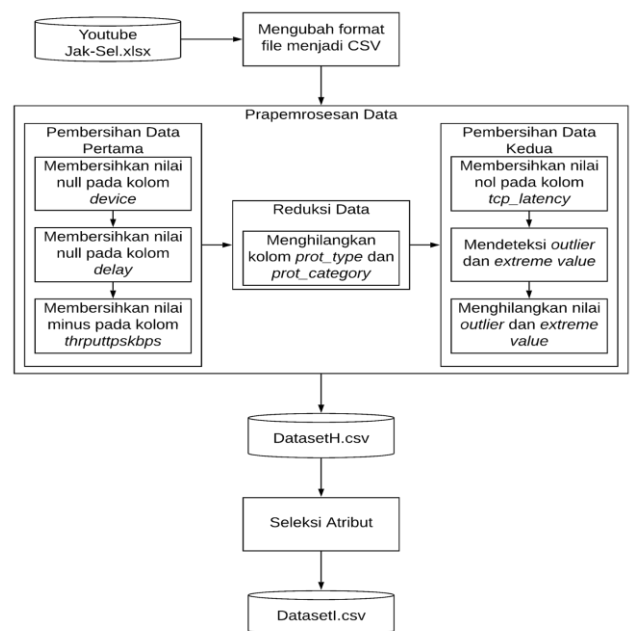
*Gambar 2. Grafik Jumlah Pelanggan yang Mengakses *YouTube* Selama Seminggu*

Dapat dilihat pada Gambar 2, jumlah pengguna terendah berada pada tanggal 16 Juni 2019 sedangkan jumlah pelanggan tertinggi berada pada tanggal 20 Juni 2016.

Analisis data juga dilakukan untuk melihat merek perangkat yang paling banyak digunakan untuk mengakses *YouTube streaming*. Namun, karena pada

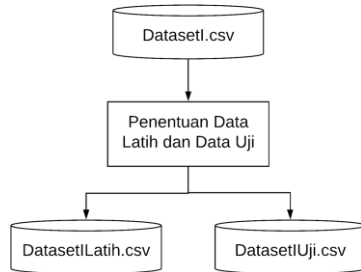
3.2. Persiapan Data

Pada Gambar 4, persiapan data dibagi menjadi lima tahap. Pembersihan data pertama, reduksi data, permbersihan data kedua, seleksi fitur, dan penentuan data latih dan data uji.



Gambar 4. Tahapan Prapemrosesan Data

Pada dataset dilakukan tahap prapemrosesan data, dan langkah pembersihan data dilakukan dua kali juga reduksi data dilakukan setelah pembersihan data pertama. Setelah prapemrosesan baru kemudian dataset diseleksi atributnya.



Gambar 5. Tahap Penentuan Data Latih dan Data Uji

Lalu dataset yang telah diseleksi atributnya masuk ke tahap penentuan data latih dan data uji seperti pada Gambar 5. Pada proses penentuan data latih dan data uji, menghasilkan dua dataset yaitu data latih dan data uji. Peneliti membagi data latih dan data uji menggunakan rasio perbandingan 80:20[13].

Tabel 4. Keseluruhan Tahap Persiapan Data

| Nama Dataset | Keterangan | Jumlah Baris | Jumlah Kolom |
|----------------------|--|--------------|--------------|
| YouTube Jak-Sel.xlsx | Dataset yang diterima. | 238.774 | 10 |
| DatasetA.csv | Format dataset diubah menjadi CSV (Comma Separated Value). | 238.774 | 10 |
| DatasetB.csv | Hasil penghilangan nilai “(null)” pada kolom delay. | 238.10 | 10 |
| DatasetC.csv | Hasil penghilangan nilai “null” pada kolom delay. | 224.06 | 10 |
| DatasetD.csv | Hasil penghilangan nilai minus pada kolom <i>thrupthttpskbps</i> | 224.05 | 10 |
| DatasetE.csv | Hasil penghilangan kolom <i>prot_type</i> dan kolom <i>prot_category</i> | 224.05 | 8 |
| DatasetE1.csv | Hasil pembagian DatasetE. Nilai 0 pada kolom <i>tcp_latency</i> | 215.51 | 8 |

dihilangkan.

DatasetE2.csv Hasil pembagian DatasetE. Nilai 0 pada kolom *tcp_latency* diganti hasil pembulatan rata-rata seluruh nilai pada kolom *tcp_latency*.

Proses Pertama

DatasetF.csv Hasil proses deteksi nilai outlier dan extreme value sebagai perbandingan, 7 disimpan sebagai file baru.

DatasetG.csv Hasil penghilangan nilai “Yes” pada kedua kolom terakhir.

Proses Kedua

DatasetF.csv Hasil proses deteksi nilai outlier dan extreme value sebagai perbandingan, 5 disimpan sebagai file baru.

DatasetG.csv Hasil proses deteksi nilai outlier dan extreme value

DatasetH.csv Hasil penghilangan nilai “Yes” pada kedua kolom terakhir.

Proses Seleksi Atribut & Penentuan Data Latih dan Data Uji

DatasetI.csv Hasil seleksi atribut DatasetH Proses Kedua.

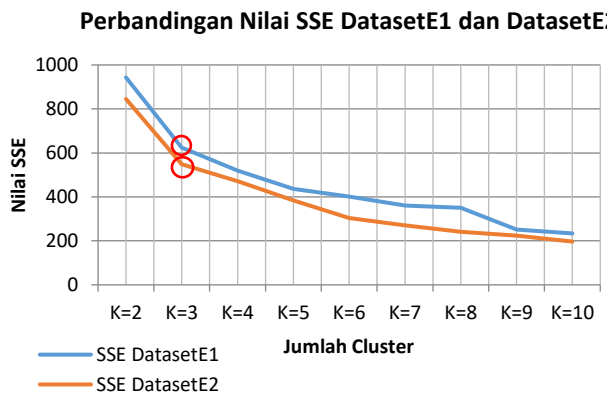
DatasetI1.csv Hasil pengurangan 20% dari total baris DatasetI.csv.

DatasetI2.csv Hasil pengurangan 80% dari total baris DatasetI.csv.

Pada Tabel 4, proses persiapan dataset dimulai dari mengubah format dataset YouTube Jak-Sel.xlsx ke format DatasetA.csv, sampai dengan hasil penghilangan kolom *prot_type* dan *prot_category* yaitu DatasetE.csv. Pada langkah penghilangan nilai 0 di kolom *tcp_latency*, peneliti memutuskan untuk menghilangkan atau mengganti dengan nilai rata-rata [14]. Dataset yang nilai 0 di kolom *tcp_latency* dihilangkan disimpan sebagai file baru dengan nama

DatasetE1.csv. Sedangkan dataset yang nilai 0 di kolom *tcp_latency* diganti rata-rata disimpan sebagai file baru dengan nama DatasetE2.csv.

Kemudian kedua dataset diuji coba menggunakan *K-means* dan dicari jumlah *cluster* optimalnya. Penentuan *cluster* optimal ditentukan dengan nilai *Sum of Squared Error* dan metode *Elbow*. *Sum of Squared Error* (SSE) menunjukkan variansi *cluster*, semakin kecil nilainya semakin baik *cluster* yang terbentuk [15]. Sedangkan metode *Elbow* adalah teknik untuk mencari jumlah *cluster* optimal dengan cara melihat penurunan drastis di titik tertentu dalam grafik nilai SSE tiap jumlah *cluster*. Penurunan yang drastis itu membentuk sudut, maka nilai itu menjadi nilai *k* atau jumlah *cluster* terbaik [15]. Gambar 6 adalah grafik Perbandingan Nilai SSE DatasetE1 dan DatasetE2.

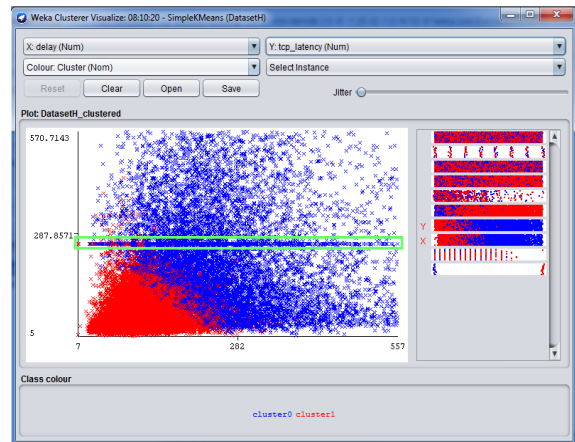


Gambar 6. Grafik Perbandingan Nilai SSE DatasetE1 dan DatasetE2

Dapat dilihat pada gambar 6 penurunan tajam sehingga membentuk sudut terjadi pada K=3 di kedua grafik. Sehingga, pada kedua dataset, jumlah *cluster* optimal adalah 3. Kemudian, peneliti mencari dataset yang nilai SSE-nya lebih kecil. DatasetE2 mempunyai SSE lebih kecil daripada DatasetE1 pada K=3 sehingga untuk tahap selanjutnya, peneliti menggunakan DatasetE2. Namun, DatasetE1 tetap disimpan.

DatasetE2 kemudian masuk ke dalam Proses Pertama. DatasetE2 disimpan sebagai file baru, dideteksi nilai *outlier* dan *extreme value*-nya, dan dihapus nilai *outlier* dan *extreme value*-nya. Dataset yang sudah bersih dari nilai *outlier* dan *extreme value* kemudian disimpan sebagai file baru, DatasetH. DatasetH dari Proses Pertama kemudian diuji coba kembali dengan *K-means* untuk diperiksa anomali data. Jika terdapat anomali data, maka DatasetH

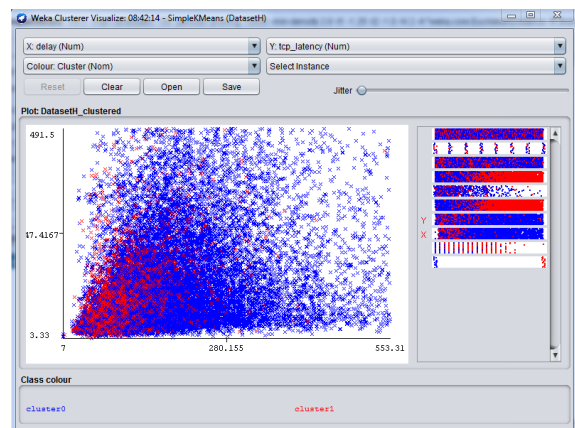
Proses Pertama tidak dapat dipakai pada proses Seleksi Atribut. Gambar 7 adalah visualisasi hasil uji coba *clustering* DatasetH Proses Pertama untuk K=2.



Gambar 7. Visualisasi Hasil Clustering DatasetH Proses Pertama untuk K=2

Dapat dilihat pada Gambar 7, DatasetH Proses Pertama mempunyai anomali data pada hasil uji coba. Maka, DatasetH tidak dapat digunakan untuk Seleksi Atribut.

Kemudian, DatasetE1 yang sudah disimpan masuk ke dalam Proses Kedua. DatasetE1 disimpan sebagai file baru, dideteksi nilai *outlier* dan *extreme value*-nya, dihilangkan nilai *outlier* dan *extreme value*-nya. Dataset hasil penghilangan nilai *outlier* dan *extreme value* disimpan sebagai file baru dengan nama DatasetH. DatasetH Proses Kedua kemudian diuji coba kembali dengan *K-means* untuk melihat anomali data. Jika terdapat anomali data, maka DatasetH Proses Kedua tidak dapat digunakan pada proses Seleksi Atribut. Gambar 8 adalah visualisasi hasil uji coba *K-means* DatasetH Proses Kedua.



Gambar 8. Visualisasi Hasil Clustering DatasetH Proses Kedua

Dapat dilihat pada Gambar 8, bahwa DatasetH Proses Kedua tidak terdapat anomali data, sehingga DatasetH Proses Kedua dapat digunakan pada proses Seleksi Atribut dan setelahnya dapat masuk ke proses Penentuan Data Latih dan Data Uji.

DatasetH yang tidak ada anomali data kemudian diseleksi atributnya. Atribut yang dipilih adalah parameter QoS dan atribut *numberofuser*. Atribut parameter QoS yaitu *thruputhttpskbps*, *delay*, dan *tcp_latency* sebagai pengganti *jitter*. Atribut *numberofuser* untuk menunjukkan jumlah pelanggan. Setelah dataset diseleksi atribut, dataset kemudian disimpan sebagai file baru dengan nama DatasetI.csv. DatasetI.csv kemudian dibagi untuk kebutuhan data latih dan data uji dengan rasio pembagian 80:20. Data latih disimpan sebagai file baru dengan nama DatasetILatih.arff dan data uji disimpan sebagai file baru dengan nama DatasetIUji.arff.

3.3. 1Clustering

Pada tahap ini DatasetI yang sudah dipisah untuk kebutuhan data latih dan data uji akan digunakan dalam proses clustering dan proses pengujian. DatasetILatih.arff dibuka untuk kemudian digunakan pada proses clustering. Metode yang dipakai adalah *K-means* dengan jumlah cluster K=2 sampai K=10. Langkah-langkah metode *K-means* adalah seperti Tabel 5.

Tabel 5. Langkah-langkah dalam Algoritma K-means[5]

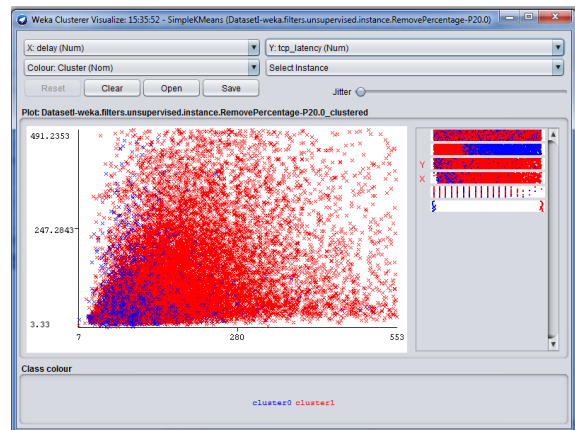
| | Delay |
|---|--|
| 1 | Pilih titik K sebagai nilai awal sebuah <i>centroid</i> |
| 2 | Ulang |
| 3 | Bentuk K <i>cluster</i> dengan menempatkan tiap titik ke <i>centroid</i> |
| 4 | Ulangi perhitungan <i>centroid</i> tiap <i>cluster</i> |
| 5 | Sampai <i>centroid</i> tidak berubah |

Persamaan untuk menghitung jarak data ke pusat *cluster* (*centroid*) adalah *Euclidean Distance*. *Euclidean Distance* adalah persamaan yang sering digunakan untuk mengukur jarak/kedekatan antara titik data pada *centroid* [5].

Persamaan *Euclidean Distance* dapat dilihat pada Persamaan 1.

$$D_{L_2}(x_2, x_1) = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \tag{1}$$

Gambar 9 adalah visualisasi hasil clustering dengan K=2 dengan X axis adalah *delay* dan Y axis adalah *tcp_latency*.



Gambar 9. Visualisasi Hasil Clustering K=2

3.4. Pengujian

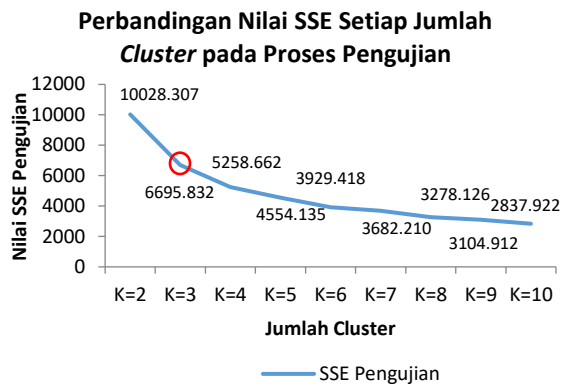
Setelah menyimpan seluruh model tiap *cluster*, maka ditentukan jumlah *cluster* optimal dengan metode *Elbow* berdasarkan nilai SSE. Untuk menentukan jumlah cluster optimal, ditentukan dengan metode *Elbow*.

Pengujian dilakukan dengan model training, sehingga SSE yang muncul adalah SSE Pengujian untuk mencari tahu jumlah *cluster* optimal. Untuk proses pengujian, peneliti sudah memuat model yang disimpan sebelumnya dan mengevaluasi dengan dataset uji yang sudah ada. Nilai SSE Pengujian tiap jumlah K dibandingkan, dan dicari titik yang mengalami penurunan tajam. Tabel 6 berisi nilai SSE Pengujian tiap jumlah K.

Tabel 6. Nilai SSE Pengujian Tiap Jumlah K

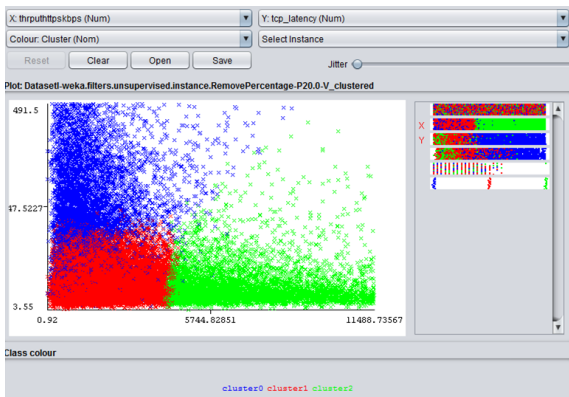
| Jumlah K | Nilai SSE Pengujian |
|----------|---------------------|
| K=2 | 10028.307 |
| K=3 | 6695.832 |
| K=4 | 5258.662 |
| K=5 | 4554.135 |
| K=6 | 3929.418 |
| K=7 | 3682.210 |
| K=8 | 3278.126 |
| K=9 | 3104.912 |
| K=10 | 2837.922 |

Gambar 10 adalah grafik nilai SSE Pengujian.



Gambar 10. Grafik Nilai SSE Pengujian

Dapat dilihat dari Gambar 10 bahwa penurunan yang tajam berada di K=3. Sehingga K=3 menjadi jumlah K optimal. Gambar 11 adalah visualisasi hasil clustering K=3.

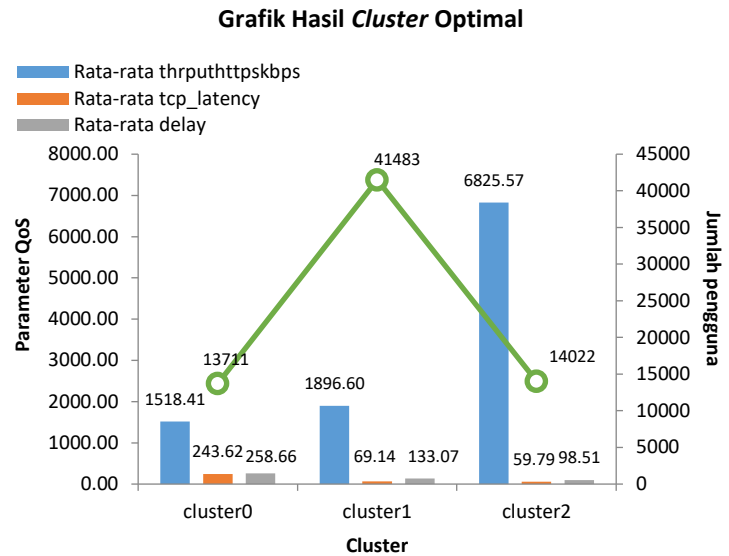


Gambar 11. Visualisasi Hasil Clustering

Pada Gambar 11 dapat dilihat untuk axis X adalah *thrputhtpskbps*, dan axis Y adalah *tcp_latency*. Cluster 0 ditandai dengan warna biru, cluster 1 ditandai dengan warna merah, dan cluster 2 ditandai dengan warna hijau. Pada Gambar 8, dapat dilihat bahwa cluster 2 mempunyai nilai *tcp_latency* yang rendah, namun *thrputhtpskbps*-nya tinggi dengan terletak di bawah dan berada paling kanan dibandingkan cluster 0 dan cluster 1. Cluster 1 mempunyai nilai *thrputhtpskbps* dan nilai *tcp_latency* yang rendah, dengan terletak berada di paling kiri dan di bawah cluster 1.

Gambar 12 adalah grafik hasil jumlah cluster optimal. Untuk parameter QoS diambil nilai rata-rata tiap cluster, sedangkan *numberofusers* dijumlah dalam masing-masing cluster. Kemudian, tabel dibuat grafik dengan axis ganda. Untuk parameter QoS menggunakan diagram batang, sementara untuk

menunjukkan jumlah pengguna menggunakan diagram garis.



Gambar 9. Grafik Hasil Cluster Optimal

Pada K=3 terbentuk tiga cluster, yaitu cluster 0, cluster 1, dan cluster 2. Cluster 2 dikategorikan sebagai cluster terbaik untuk parameter QoS dengan sebab mempunyai rata-rata nilai *throughput* yang tinggi dan rata-rata nilai *latency* dan *delay* yang rendah. Cluster 1 dan cluster 0 mempunyai selisih rata-rata nilai *throughput* yang tidak terlalu besar, namun untuk rata-rata nilai *latency* dan *delay* cluster 0 lebih besar dibandingkan cluster 1. Maka cluster 1 dapat dikategorikan sebagai terbaik kedua.

Berdasarkan kecepatan internet yang direkomendasikan YouTube, cluster 2 masuk ke dalam kategori HD 720p, sedangkan cluster 0 dan cluster 1 masuk ke dalam kategori SD 480p. Berdasarkan kategori *delay*, cluster 2 masuk ke dalam kategori Sangat Bagus, sedangkan cluster 0 dan cluster 1 masuk ke dalam kategori Bagus.

Jika dilihat pada jumlah pengguna, cluster 1 mempunyai jumlah pengguna yang paling banyak sebesar 41.483 pengguna, sedangkan cluster 2 mempunyai 14.022 pengguna, dan cluster 0 mempunyai 13.711 pengguna.

4. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik kesimpulan yaitu bahwa, terbentuk tiga cluster yaitu cluster 0, cluster 1, dan cluster 2. Cluster 2 adalah cluster terbaik pertama dengan rata-rata nilai *throughput*

tertinggi, dan rata-rata nilai *delay* dan *latency* terendah.

Berdasarkan rekomendasi kecepatan internet dari *YouTube*, *cluster 2* masuk kategori HD 720p sementara *cluster 0* dan *cluster 1* masuk ke kategori SD 480p. Berdasarkan kategori *delay*, *cluster 2* masuk ke dalam kategori Sangat Bagus, sementara *cluster 0* dan *cluster 1* masuk kategori Bagus. Berdasarkan jumlah pengguna, *cluster 1* menempati urutan terbanyak, sementara *cluster 2* dan *cluster 0* mempunyai selisih jumlah pengguna yang tidak terlalu besar.

Dari analisa hasil dapat disimpulkan bahwa, kualitas pengalaman pelanggan XL Axiata dalam streaming *YouTube* video pada daerah Jakarta Selatan masih dalam kategori standar. Meskipun tidak buruk, namun jumlah pengguna yang masuk ke dalam *cluster 2* masih tidak jauh berbeda dengan *cluster 0* menunjukkan bahwa kualitas yang pelanggan terima tidak buruk namun perlu peningkatan pelayanan jaringan.

Saran bagi penelitian selanjutnya bahwa penelitian ini perlu dikembangkan lagi dengan atribut yang lebih lengkap, seperti *Mean Opinion Score* (MOS) atau jangka waktu riwayat data yang lebih lama, seperti data pelanggan sebulan sebelumnya. Penelitian selanjutnya juga diharapkan untuk mencoba algoritma selain *K-means*, seperti *K-medoids*, atau *Fuzzy K-means*. Dapat juga menggunakan *tools* yang lain seperti *Rapid Miner*, atau membuat model dengan bahasa pemrograman R, dan *Python*. Pada penelitian selanjutnya diharapkan untuk dapat dilihat kepadatan *cluster-cluster* berdasarkan lokasi tertentu.

DAFTAR PUSTAKA

2009, Developer Zen. [Online]. Available: <https://developerzen.com/data-mining-handling-missing-values-the-database-bd2241882e72>

2019, Bantuan YouTube. [Online] Available: https://support.google.com/youtube/answer/3037019?hl=id&ref_topic=9257580

2019, Lori Carr and Associates. [Online]. Available: <http://loricarrassociates.com/customer-segmentation-a-very-delicious-customer-experience-strategy/>

2019, YouTube for Press. [Online]. Available: <https://www.youtube.com/yt/about/press/>

A. Andriani, "Klasifikasi Berbasis Algoritma C4.5 untuk Deteksi Kenaikan *Case Fatality Rate*

Demam Berdarah", Seminar Nasional Informatika Medis, 2017, pp. 70-75.

- A. Sangsari, Isnawaty, and L. Aksara, Analisis QoS (*Quality of Service*) pada Layanan Video Streaming yang Menggunakan Protokol RTMP (*Real Time Messaging Protocol*), "*semanTIK*", Vol. 2, pp. 177-188.
- A. Savitri, F. Bachtiar, and N. Setiawan, Segmentasi Pelanggan Menggunakan Metode K-means Clustering Berdasarkan Model RFM pada Klinik Kecantikan (Studi Kasus: Belle Crown Malang), "*Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIHK)*", Vol. 2, pp. 2957-2966.
- B. E. Adiana, I. Soesanti, and A. E. Permasari, Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering, "*JUTEP*", Vol. 2, pp. 23-32, 2018.
- I. Iskandar and A. Hidayat, Analisa Quality of Service (QoS) Jaringan Internet Kampus (Studi Kasus: UIN Suska Riau), "*Jurnal Core IT*", Vol. 1, pp. 67-76, 2015.
- J. Spiess, Y. T'Joens, and R. Dragnea, Using Big Data to Improve Customer Experience and Business Performance, "*Bell Labs Technical Journal*", Vol. 18, pp. 3-17, 2014.
- M. Fiedler, T. Hossfield, and P. Tran-Gia, A Generic Quantitative Relationship between Quality of Experience and Quality of Service, "*IEEE Network*", Vol. 24, pp. 36-41, 2010.
- P. N. Tan, M. Steinbach, and V. Kumar, "*Introduction to Data Mining*", Second Edition, M. Goldstein and K. Harutunian, Boston, Pearson Education, Inc, 2006.
- R. Noor, and S. Khosandroo, Quality of Experience Key Metrics Framework for Network Mobility User, "*International Journal of Physical Sciences*", Vol. 6, pp. 6521-6528, 2011.
- T. Hardiani, S. Sulisty, and R. Hartanto, "Segmentasi Nasabah Tabungan Menggunakan Model RFM (*Recency, Frequency, Monetary*) dan *K-Means* pada Lembaga Keuangan Mikro", Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMANTIK), 2017.
- Widiarina and R. S. Wahono, Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-means dalam Pemetaan Nasabah Potensial, "*Journal of Intelligent Systems*", Vol. 1, pp. 33-36, 2015