

---

---

## OPTIMASI ALGORITMA K-MEANS TERHADAP DATA OUTLIER MENGUNAKAN ROBUST DALAM SEGMENTASI PELANGGAN BIZNET MEDAN

Dwina Pri Indini<sup>1</sup>, Solly Ariza<sup>2</sup>, Zulham Sitorus<sup>3</sup>

Universitas Pembangunan Panca Budi, Medan

e-mail: <sup>1</sup>dwinapriindini03@gmail.com

**Abstract:** *The massive growth of internet penetration in Indonesia has created competitive competition among internet service providers (ISPs), including Biznet Medan. However, the highly diverse and heterogeneous characteristics of customers make one-size-fits-all marketing strategies ineffective and inefficient. Processing large customer data for segmentation is often hampered by the presence of outliers, such as extreme income variations, which can disrupt the clustering accuracy of conventional algorithms. This study aims to optimize the K-Means algorithm in segmenting Biznet Medan's target market to maintain accuracy despite outliers, allowing the company to design more targeted marketing strategies. To overcome K-Means' sensitivity to outliers, this study applies the Robust Scaling technique. This solution uses the Median and Interquartile Range (IQR) values instead of the mean, so that the cluster center points (centroids) remain stable and are not distorted by extreme data values. The research flow includes data collection of prospective Biznet Medan customers, data pre-processing (cleaning and scaling), feature selection (income, distance, and network requirements), and implementation of the K-Means model. Cluster quality evaluation is carried out by comparing the results of using StandardScaler and RobustScaler through the Silhouette Score and Davies Bouldin Index (DBI) metrics. The implementation successfully formed two main segments: Cluster 0 (Enterprise/Premium Customers) with high income characteristics and large bandwidth requirements (>200 Mbps), and Cluster 1 (Casual/Economical Customers) who dominate the market with lower network requirements (<50 Mbps). Optimization with Robust showed significant improvements in the evaluation metrics, where the Silhouette Score increased from 0.3133 to 0.6023, and the DBI value improved from 1.1421 to 0.8007. The implementation of Robust Scaling is effectively able to handle outlier problems in the K-Means algorithm, resulting in a more robust, efficient, and representative cluster structure. These results contribute to the processing of customer data for more objective strategic decision-making in ISP market expansion.*

**Keyword:** *k-means; clustering; outlier; robust scaling; customer segmentation.*

**Abstrak:** Pertumbuhan internet di Indonesia yang masif telah menciptakan persaingan kompetitif di antara penyedia jasa internet (ISP), termasuk Biznet Medan. Namun, karakteristik pelanggan yang sangat beragam dan bersifat heterogen membuat strategi pemasaran satu untuk semua (*one-size-fits-all*) menjadi tidak efektif dan tidak efisien. Data calon pelanggan yang besar sering ditemui data yang hilang, perbedaan nilai data yang ekstrem sehingga pengolahan data calon pelanggan untuk segmentasi sering kali terkendala oleh adanya data pencilan (*outlier*) yang dapat mengganggu akurasi pengelompokan pada algoritma konvensional. Penelitian ini bertujuan untuk mengoptimasi algoritma *K-Means* dalam melakukan segmentasi target pasar Biznet Medan agar tetap akurat meskipun terdapat data *outlier*, sehingga perusahaan dapat merancang strategi pemasaran yang lebih tepat sasaran. Untuk mengatasi sensitivitas *K-Means* terhadap pencilan, penelitian ini menerapkan teknik *Robust Scaling*. Solusi ini menggunakan nilai Median dan *Interquartile Range* (IQR) sebagai pengganti rata-rata, sehingga titik pusat klaster (*centroid*) tetap stabil dan tidak terdistorsi oleh nilai data yang

ekstrem. Alur penelitian meliputi pengumpulan data calon pelanggan Biznet Medan, pra-pemrosesan data (pembersihan dan *scaling*), pemilihan fitur (penghasilan, jarak, dan kebutuhan jaringan), serta implementasi model *K-Means*. Evaluasi kualitas kluster dilakukan dengan membandingkan hasil penggunaan *StandardScaler* dan *RobustScaler* melalui metrik *Silhouette Score* dan *Davies Bouldin Index (DBI)*. Implementasi berhasil membentuk dua segmen utama: *Cluster 0* (Pelanggan *Enterprise/Premium*) dengan karakteristik penghasilan tinggi dan kebutuhan *bandwidth* besar (>200 Mbps), serta *Cluster 1* (Pelanggan *Casual/Hemat*) yang mendominasi pasar dengan kebutuhan jaringan lebih rendah (<50 Mbps). Optimasi dengan *Robust* menunjukkan peningkatan signifikan pada metrik evaluasi, di mana *Silhouette Score* meningkat dari 0,3133 menjadi 0,6023, dan nilai DBI membaik dari 1,1421 menjadi 0,8007. Penerapan *Robust Scaling* secara efektif mampu menangani masalah *outlier* pada algoritma *K-Means*, menghasilkan struktur kluster yang lebih kuat, efisien, dan representatif. Hasil ini memberikan kontribusi pada pengolahan data pelanggan untuk pengambilan keputusan strategis yang lebih objektif dalam ekspansi pasar ISP.

**Kata kunci:** *K-Means*; klusterisasi; data pencilan; *Robust Scaling*; Segmentasi Pelanggan.

## PENDAHULUAN

Internet merupakan jaringan komunikasi yang bersifat global, sesuai dengan namanya “*interconnected network*” jaringan ini berfungsi untuk menghubungkan jutaan manusia yang menggunakan perangkat teknologi dan komunikasi. Pada masa transformasi digital yang sudah sangat integral ini, peran internet bukan lagi dianggap sebagai salah satu gaya hidup, akan tetapi telah menjadi kebutuhan primer *utilities* yang harus dipenuhi oleh sebagian besar masyarakat agar kegiatan sehari – harinya dapat berjalan dengan lancar (Megiati et al., 2024).

Pertumbuhan internet di Indonesia tergolong masif dimana hal tersebut dipicu oleh tuntutan teknologi yang mengharuskan penggunaan internet disegala sektor ekonomi dan hidup masyarakat serta hampir meratanya akses internet di seluruh lapisan masyarakat (Yulianti et al., 2024). Penggunaan internet tersebut meningkatkan penetrasi internet di Indonesia dari tahun ke tahun. Berdasarkan data dari APJII (Asosiasi Penyelenggara Jasa Internet Indonesia), tingkat penetrasi internet per tahun 2025 ini mengalami peningkatan dari tahun sebelumnya yakni meningkat menjadi

sebesar 80,66 % dengan jumlah pengguna internet di Indonesia sebesar 229 juta jiwa. (Asosiasi Penyelenggara Jasa Internet Indonesia - Survei, n.d.).

Besarnya angka perkembangan internet di Indonesia dan tingginya pengaruh internet terhadap konektivitas digital tidak hanya memicu meluasnya pengguna internet namun juga menciptakan persaingan yang sangat kompetitif diantara pada *Internet Service Provide (ISP)* di Indonesia, salah satunya adalah Biznet.

Biznet (PT Supra Primatama Nusantara) adalah perusahaan penyedia infrastruktur digital terintegrasi yang anda dengan jaringan kabel fiber optik mutakhir. Biznet didirikan pada 1 Oktober 2000 oleh Adi Kusma di Jakarta. Biznet menyediakan layanan internet *broadband* berkualitas tinggi yang pada awalnya untuk korporasi. Tahun ke Tahun biznet semakin berkembang dan mulai melakukan ekspansi ke rumah warga pada tahun 2012 dan sudah mencakup banyak kota di Indonesia (*Tentang Kami - Biznet Networks*, n.d.).

Akan tetapi saat ini kompetitor biznet cukup besar dan Banyaknya penyedia layanan ISP di Indonesia saling berebut pangsa pasar yang semakin tersegmentasi. Usaha dalam

mempertahankan loyalitas pelanggan secara efisien menuntut biznet tidak hanya mengandalkan keunggulan infrastruktur tetapi juga diperlukan ketajaman strategi agar ekspansi target pasar biznet semakin luas dan loyal serta potensi pendapatan perusahaan meningkat.

Namun, segmentasi pasar pengguna internet di Indonesia sangat beragam dan ditentukan oleh kelas ekonomi masing – masing. Tentunya tiap segmen masyarakat memiliki karakteristik yang berbeda – beda, maka strategi *one-size-fits-all* (satu strategi untuk semua) tidak lagi efektif dan cenderung menciptakan inefisiensi biaya. Pendekatan terhadap pelanggan juga tentunya perlu disesuaikan berdasarkan karakteristik dari mereka. Hal inilah yang perlu dipahami lebih dalam bagi biznet dimana biznet harus mengidentifikasi dan segmentasi kelompok pelanggan yang paling potensial ditengah keberagaman karakteristik sosial dan teknis.

Namun, melakukan segmentasi secara konvensional menjadi sulit dilakukan mengingat variabel yang memengaruhi keputusan pelanggan sangatlah kompleks mulai dari daya beli secara ekonomi hingga kendala teknis seperti jarak lokasi rumah ke titik penyedia.

Data pelanggan yang masif belum mampu dikelola dengan baik untuk mengetahui kelompok target pasar biznet menjadi informasi strategis berdasarkan karakteristik data calon pelanggan. Untuk itu, perlunya pendekatan *machine learning* untuk menganalisis data pelanggan dan mengetahui kelompok target pasar yang tersegmentasi sebagai calon pelanggan casual maupun pelanggan enterprise. Untuk mengetahui sebaran data target pasar dan pola tersembunyi yang tidak tertangkap oleh analisis statistik sederhana, Segmentasi target pasar biznet dapat dianalisa menggunakan teknik *clustering* dengan memanfaatkan algoritma *K-Means Clustering*.

*K-Means* dipilih karena

kemampuannya memroses data dalam jumlah besar secara objektif untuk menentukan segmen pelanggan paling ideal. Kombinasi kesederhanaan, kecepatan dan efektivitasnya juga unggul dibandingkan algoritma cluster yang lainnya.

Penelitian terdahulu oleh (Sitorus et al., 2024) menyatakan bahwa algoritma *K-Means Clustering* mampu menghasilkan hasil clustering yang cukup presisi dikarenakan proses kerja dari algoritma K-Means mengidentifikasi pola kesamaan dari data dan pada penelitian tersebut dinyatakan bahwa k-means lebih unggul daripada algoritma konvensional yang tidak melakukan pengelompokan.

Kemudian pada penelitian (Harjoyudanto & Zulvia, 2023) algoritma *k-means clustering* membantu terciptanya strategi pemasaran yang sesuai dengan target pasar, algoritma *k-means* mampu memetakan pelanggan potensial. Selanjutnya penelitian oleh (Julian, 2025) penelitian tersebut menggunakan *k-means* untuk segmentasi pelanggan pada jaringan telekomunikasi, penerapan *k-means* terbukti efektif dalam membantu perusahaan telekomunikasi memahami pelanggan dan mengembangkan strategi marketing berbasis data yang lebih efisien dan pada penelitian (Sutiono et al., 2025) memaparkan bahwa metode k-means terbukti mampu menjadi komponen peranti yang tepat dalam pengambilan keputusan manajerial perusahaan.

Berdasarkan pemaparan dari penelitian terdahulu algoritma *k-means* terbukti cocok dalam melakukan segmentasi target pemasaran dan menghasilkan kelompok pelanggan yang sesuai dengan strategi pemasaran perusahaan.

Akan tetapi, Permasalahan muncul ketika dalam pengumpulan dan analisis data sering kali ditemui data yang hilang, data yang rentang nilainya jauh dari rentang normal (data pencilan), serta algoritma *k-means* yang memiliki tingkat sensitivitas yang cukup tinggi terhadap data pencilan (*outlier*) yang disebabkan oleh *k-means* menggunakan rata – rata

dari semua titik data sehingga bila terdapat nilai yang ekstrem akan menarik centroid ke arahnya sehingga titik centroid tidak berada pada kumpulan data sebenarnya dan kemudian *k-means* tidak memiliki kemampuan untuk mengabaikan *noise* yang menyebabkan setiap titik data ikut dimasukkan ke dalam cluster tertentu (N Tahta Phudjashakty et al., 2026).

Hal tersebut menjadi urgensi dalam penelitian ini karena *outlier* dapat mengurangi kinerja algoritma dan mengaburkan pola menjadi bias, menghasilkan pemetaan data yang tidak akurat seperti *cluster* yang harusnya terpisah jadi menyatu, *cluster kecil* menjadi menyimpang karena centroidnya tertarik menjauh maka jumlah *k* yang nantinya dihasilkan juga tidak sesuai. Sehingga diperlukan optimalisasi yang *robust* untuk menangani data outlier pada algoritma *k-means* (Khairul Abdi Sinuraya, Andri Nata, 2025).

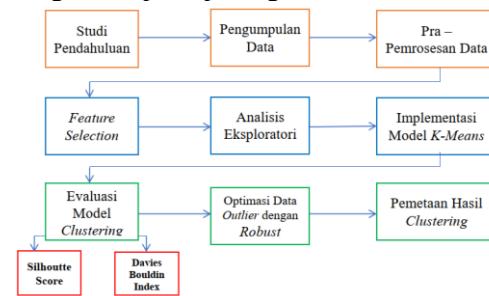
*Robust scaler* merupakan teknik pengelompokan data yang efisien dalam mereduksi *outlier* dengan mengidentifikasi data yang bersifat homogen namun tidak terpengaruh oleh data yang bernilai ekstrem. *Robust scaler* dirancang khusus untuk menangani dataset yang memiliki data *outlier* atau distribusi data yang tidak normal sehingga algoritma tidak terdistorsi oleh nilai ekstrim (Nugraha et al., 2024).

Untuk itu penelitian ini bertujuan untuk meng-optimalisasi algoritma *k-means* terhadap data pencilan (*outlier*) menggunakan *robust* dalam data memetakan target pasar biznet sehingga diharapkan dapat memberikan insight dengan akurasi yang baik untuk perencanaan strategi yang tepat dalam ekspansi pemasaran biznet ke masyarakat dan mampu menyesuaikan karakteristik target pasar yang heterogen.

## METODE

Dalam penelitian ini memiliki alur penelitian yang digambarkan dalam

diagram seperti pada gambar 1:



**Gambar 1 Tahapan Penelitian**

### Studi Pendahuluan

Identifikasi Masalah: segmentasi target pasar perusahaan internet service provider biznet untuk mengetahui pendekatan dan strategi seperti apa yang tepat untuk dilaksanakan dalam menarik daya beli dan loyalitas pelanggan sehingga perlunya pendekatan machine learning untuk klasterisasi target pasar. Akan tetapi, dalam pengumpulan data ditemukan data pencilan yang menyebabkan klasterisasi target pasar kurang akurat sehingga diperlukan penanganan *outlier* untuk mengoptimalkan kinerja *k-means clustering* dalam memetakan *cluster* target pasar biznet.

### Pengumpulan Data

Data dalam penelitian ini adalah data calon pelanggan yang diperoleh dari perusahaan Biznet Medan. Data mentah cukup kompleks dengan properti yang bervariasi dan jumlah yang cukup besar.

### Pra – Pemrosesan Data

Tahap ini wajib dilakukan sebelum memulai pengolahan data dengan *machine learning*. Pra – pemrosesan data membersihkan data dengan menghapus duplikasi, menangani nilai yang hilang (Nan) serta memastikan format data dan satuannya konsisten (Putri Azhiah et al., 2025). Pada penelitian ini data dengan variabel penghasilan masih bersifat objek karena ada lambang rupiah (Rp. ), Jarak terdapat satuan Meter (M) dan kebutuhan jaringan masih terdapat satuan Mbps sehingga dilakukan cleaning data untuk memastikan data tersebut berubah

satunya menjadi integer dan data siap digunakan. Kemudian dilakukan *scaling* yakni proses transformasi rentang nilai fitur ke skala yang sama besar. Dalam *k-means*, *scaling* dilakukan karena *k-means* mengambil jarak euclidean, sehingga dalam penelitian ini dilakukan *scaling* pada fitur penghasilan, jarak dan kebutuhan jaringan yang memiliki rentang nilai dari puluhan, ratusan hingga jutaan. *Scaling* menggunakan *StandardScaler* atau *MinMaxScaler*.

### Feature Selection

Tahapan ini melakukan teknik analisis untuk memilih fitur – fitur yang paling berpengaruh terhadap kebutuhan *internet* dari calon pelanggan Biznet. Fitur yang dipilih adalah penghasilan, jarak rumah dengan ISP, dan kebutuhan jaringan calon pelanggan. Fitur tersebut dinilai memiliki korelasi yang cukup kuat terhadap algoritma dalam penelitian ini. Setelah dilakukan analisis fitur maka ditetapkan 3 fitur yang akan diolah yakni fitur penghasilan, jarak dan kebutuhan jaringan.

### Analisis Eksploratori

Tahapan ini menganalisis visualisasi dari data yang akan digunakan. Visualisasi data seperti box plot atau histogram tujuannya tersebut untuk mendeteksi anomaly dan *outlier* serta menemukan pola yang ada pada fitur dataset (Hakim et al., 2025).

### Implementasi Model K-Means

*K-Means* merupakan algoritma *unsupervised learning* dimana tidak memiliki label/target akan tetapi algoritmanya bertujuan untuk mengelompokkan data dan menemukan pola tersembunyi (Sitorus et al., 2025).

### Evaluasi Model Clustering

Untuk mengetahui seberapa baik algoritma *k-means* bekerja dan cocok dengan data maka digunakan 2 metode evaluasi yakni:

1. *Silhouette Coefficient*, mengukur kepadatan rata – rata jarak antara

satu titik data dengan titik yang lainnya didalam kluster yang sama (*cohesion*) dan seberapa jauh sebuah kluster terpisah dengan cluster yang lain (*separation*) (Hasan, 2024). Tujuannya *silhouette coefficient* mengukur seberapa dekat jarak suatu titik dengan kluster tetangganya dibandingkan titik kluster aslinya. Interpretasi nilainya dikatakan baik jika  $0.6 - 0.7 =$  struktur nilai layak/masuk akal,  $0.8 - 1.0 =$  struktur kluster sangat kuat yang artinya semakin mendekati angka 1 kluster yang dihasilkan semakin baik (Rahmawati et al., 2024).

2. *Davies Bouldin Index* (DBI), memberikan validasi dari sudut pandang keseluruhan model apakah model pengelompokkan sudah rapi dan efisien. Interpretasi nilainya dikatakan semakin baik apabila nilainya semakin mendekati angka 0 (Kilo et al., 2025).

### Optimasi Data Outlier Dengan Robust

Penelitian ini memiliki data pencilan namun tidak ingin menghapus data tertentu, untuk itu diperlukan sebuah teknik yang kuat untuk mengatasi *outlier* dalam hal ini *robust* cocok digunakan karena *robust* tahan terhadap data pencilan. Dengan menggunakan *RobustScaler* dalam pra-premrosesan data maka pusat data akan berpindah ke arah median (nilai tengah) bukan lagi rata – rata (Nugraha et al., 2024). Jika terdapat perbedaan skala data yang cukup ekstrem, dengan menggunakan median sebagai pusat data maka tidak akan mempengaruhi data yang memiliki nilai paling besar atau paling kecil karena yang diambil adalah nilai tengahnya (Khoirunnisa & Ramadhan, 2023). *Robust* diformulasikan dalam rumus berikut:

$$X_{\text{scaled}} = \frac{x - \text{Median}}{\text{IQR}}$$

Keterangan:

IQR = Kuartil 3 – Kuartil 1.

**Pemetaan Hasil Clustering**

Pemetaan dilakukan untuk menunjukkan visualisasi data dan sebaran kelompok data menggunakan *scatter plot*. Dengan demikian sebaran data dapat dianalisis lebih dalam untuk mengetahui pola data.

**HASIL DAN PEMBAHASAN**

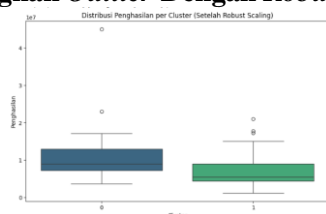
**Dataset Penelitian**

Dataset dalam penelitian ini diperoleh dari Biznet medan yang telah melalui tahapan pra-pemrosesan data dan *feature scaling* datanya adalah sebagai berikut:

**Table 1 Dataset Penelitian**

No	Penghasilan	Jarak	Kebutuhan Jaringan
1	5.200.000	150	50
2	3.500.000	250	20
3	15.000.000	100	100
4	7.400.000	150	250
5	4.400.000	450	350
6	12.000.000	250	150
7	7.700.000	500	50
8	8.400.000	250	20
9	2.300.000	500	20
10	9.000.000	250	20
.....	.....	.....	.....
100	9.100.000	250	50

**Analisis Eksploratori Setelah Penangan Outlier Dengan Robust**

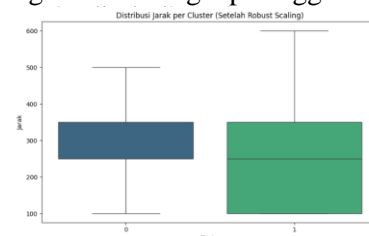


**Gambar 1 Box Plot Penghasilan Dengan Robust**

Berdasarkan analisis dari box plot diatas, dapat terlihat Penerapan Robust Scaling mentransformasi distribusi data yang sebelumnya mengalami kompresi (penekanan) akibat nilai pencilan ekstrem (*extreme outliers*) menjadi struktur data yang lebih terstratifikasi secara ilmiah.

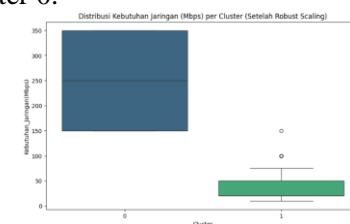
Pada kondisi sebelum Robust, skala grafik terdistorsi sehingga "kotak" pada *box plot* terlihat gepeng dan tumpang tindih (*overlap*) karena dipaksa mengakomodasi nilai maksimum yang sangat tinggi, namun setelah transformasi menggunakan Median dan Interquartile Range (IQR), model berhasil memitigasi pengaruh pencilan tersebut sehingga setiap cluster memiliki pemisahan median yang jelas dan berjenjang. Hal ini membuktikan bahwa algoritma K-Means kini bekerja lebih akurat dalam mengidentifikasi batas-batas segmen pelanggan berdasarkan variansi data yang lebih proporsional, tanpa menghilangkan keberadaan *outlier* asli sebagai representasi data yang tetap valid secara statistik.

Jumlah titik *outlier* bertambah bukan karena datanya berubah, tetapi karena Robust Scaling mempersempit batas "normal" agar hanya mencakup mayoritas data yang paling padat. Hal ini membuat model K-Means jauh lebih objektif karena ia tidak lagi dipaksa menganggap pelanggan berpenghasilan menengah-atas sebagai pelanggan "biasa".



**Gambar 2 Box Plot Jarak Dengan Robust**

Pada box plot jarak, untuk pelanggan enterprise median berada pada kisaran data 300 M dan untuk pelanggan casual median bervariasi dengan median yang mirip atau sedikit lebih tinggi dari kluster 0.



**Gambar 3 Box Plot Kebutuhan Jaringan Dengan Robust**

Sementara box plot data kebutuhan jaringan pelanggan enterprise menunjukkan kebutuhan jaringan yang jauh lebih tinggi. Mediannya berada di atas 200 Mbps, dan ada outlier yang mencapai angka yang sangat tinggi (di atas 300 Mbps), mengindikasikan segmen ini memang membutuhkan jaringan yang sangat besar. Pada pelanggan casual kebutuhan jaringan lebih rendah dengan median dibawah 50 Mbps.

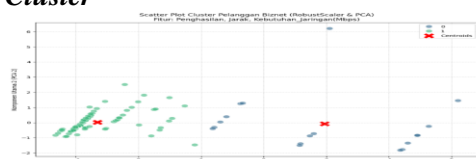
Secara umum, adanya outlier pada ketiga plot menunjukkan bahwa meskipun RobustScaler digunakan (yang dirancang untuk menangani outlier dengan baik), masih ada beberapa data point yang berada jauh dari nilai tengah (median) dan jangkauan interkuartil (IQR) dari cluster masing-masing. Akan tetapi yang perlu digari bawah ialah RobustScaler membantu mengurangi dampak ekstrem dari outlier ini terhadap proses clustering dibandingkan scaler lainnya.

**Implementasi K-Means**

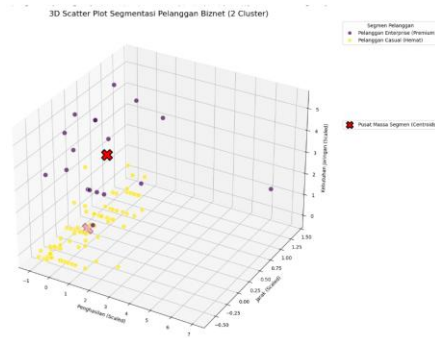
- Menentukan Jumlah *k* Optimal  
 Penelitian ini menetapkan jumlah *k* berdasarkan tujuan analisis bisnis yakni mengelompokkan data berdasarkan target pasar / pelanggan yang casual (hemat) dan pelanggan yang premium. Dimana *cluster 0* yakni calon pelanggan Enterprise (Premium) dan *cluster 1* yakni calon pelanggan Casual (hemat).
- Inisialisasi Centroid Awal  
 Setelah dilakukan penanganan outlier dengan robust pada pra-pemrosesan data, maka diperoleh titik centroid awal pada angkat:

Cluster	Penghasilan	Jarak	Kebutuhan Jaringan
0	0.936667	0.088889	3.755102
1	0.063605	0.012195	0.234943

**Mengelompokkan Data kedalam Cluster**



**Gambar 4 Scatter Plot dengan PCA**



**Gambar 5 Scatter Plot K-Means 3D Dengan Robust**

Dilihat dari kedua gambar scatter plot diatas, Visualisasi menggunakan 3 fitur yakni penghasilan, jarak dan kebutuhan jaringan. Untuk cluster 0 yakni pelanggan enterprise (premium ) diidentifikasi dengan titik berwarna ungu yang mendominasi pada sumbu z pada fitur kebutuhan jaringan ini menunjukkan bahwa pelanggan premium menggunakan bandwidth dengan skala besar. Dari segi penghasilan pelanggan premium memiliki rentang penghasilan dari area menengah hingga ke tinggi. Titik centroid yang bertanda silang merah berada jauh dari titik asal, mencerminkan nilai rata-rata penggunaan jaringan yang signifikan.

Untuk cluster 1 yakni pelanggan casual (hemat) yang diidentifikasi dengan titik kuning/hijau mayoritas pelanggan ini menggunakan jaringan dengan skala bandwidth kecil hingga sedang. Fitur penghasilan pada klaster ini terpusat pada area kiri bawah yakni pada angka penghasilan Rp.1.000.000 sampai Rp.3.000.000 (penghasilan rendah hingga menengah). Jika dilihat dari segi volume titik yang tampil, pelanggan casual lebih mendominasi target pasar biznet dimana jumlah pelanggan casual lebih besar, namun nilai kontribusi per variabel lebih kecil.

**Evaluasi Model K-Means**

Untuk melihat seberapa tepat dan akurat algoritma *k-means* dalam mengelompokkan data berdasarkan karakteristiknya maka dilakukan evaluasi nilai dengan 2 teknik pengujian yakni *silhouetter coeficient* dan *Davies Bouldin*

Index. Hasilnya diperoleh nilai sebelum dan sesudah ditangani outlier dengan robust.

**Sebelum Dioptimasi Robust (StandardScaler)  $k=2$**

Tabel 2 hasil nilai evaluasi

<i>Silhouette Score</i>	<i>Davies Bouldin Index</i>
0.3133	1.1421

Dari tabel hasil nilai evaluasi diatas, dapat didapatkan informasi bahwa hasil clustering  $k$ -means menggunakan *standardscaler* yang artinya belum dioptimalkan menggunakan *robust* nilai evaluasinya kurang tepat yakni *silhouette score*nya berada jauh dibawah 0.5 yang merupakan interpretasi standar dari nilai *silhouette* dimana kelompok data yang dihasilkan memiliki rentang jarak yang jauh, serta data tidak sama walaupun dalam satu kelompok. Kemudian nilai *davies bouldin index* yang menjauhi angka 0 dimana nilai performa dari  $k$ -means dikatakan kurang baik interpretasi nilainya karena dinilai kluster yang dihasilkan saling tumpang tindih dan tidak sama nilainya dalam satu kluster.

**Setelah Dioptimasi Robust (RobustScaler)  $k=2$**

Tabel 3 hasil nilai evaluasi

<i>Silhouette Score</i>	<i>Davies Bouldin Index</i>
0.6023	0.8007

Setelah dilakukan optimasi algoritma  $k$ -means dengan *robust* diperoleh nilai evaluasi yang jauh lebih baik dibandingkan sebelum menggunakan *robust*. Nilai *Silhouette Score* diatas 0.60 yang mengindikasikan bahwa interpretasi tersebut mendekati angka 1 artinya  $k$ -means setelah dioptimalkan dengan *robust* berhasil mengklusterisasi cukup kuat, data terpisah ke kelompok klasternya dengan baik dan objek klusterisasinya sesuai dengan karakteristik data yang diidentifikasi serta data dalam sebuah kluster sama dan terpisah dengan

baik dari kluster lainnya. Kemudian nilai DBInya 0.800 yang artinya sudah berada dibawah rentang 1 yang berarti bahwa hasil kluster baik dan cukup menunjukkan kelompok terpisah dengan jelas dan pemisahannya wajar tidak saling tumpang tindih.

**Pemetaan Hasil Clustering**

Hasil data dari cluster 0 : Pelanggan Enterprise (Premium)

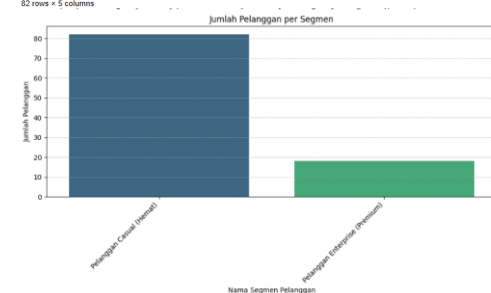
```

--- detail untuk Cluster #: Pelanggan Enterprise (Premium) ---
Nama_Pelanggan Penghasilan Jarak Kebutuhan_Jaringan(Mbps) Nama Segmen
3 LANCE RERE JULIANSIA SIHOMBING 7400000 150 250 Pelanggan Enterprise (Premium)
4 POLTAK PANORANGAN SIBARAN 4400000 450 350 Pelanggan Enterprise (Premium)
5 ESTERIDA TAMBUNAN 12000000 250 150 Pelanggan Enterprise (Premium)
26 PITUA SIAGIAN 4500000 100 250 Pelanggan Enterprise (Premium)
27 RIATA ANASTASIA MANUALU 7200000 150 350 Pelanggan Enterprise (Premium)
30 VERA LOLITA TARIGAN 8000000 350 250 Pelanggan Enterprise (Premium)
31 RANI FARIDA SINAGA 13230000 400 350 Pelanggan Enterprise (Premium)
32 JEFRIANDO SINAGA 17100000 100 150 Pelanggan Enterprise (Premium)
63 ANNISA YUANTIA ANGGRENI 3700000 250 250 Pelanggan Enterprise (Premium)
54 ASHELVA PRAHASDITA SARI 10000000 250 350 Pelanggan Enterprise (Premium)
55 ASRI SULSTIYOHADI 10000000 250 150 Pelanggan Enterprise (Premium)
56 ASRIKA 10000000 250 350 Pelanggan Enterprise (Premium)
57 ASSET JAKARIA SEMBIRING 8000000 250 150 Pelanggan Enterprise (Premium)
63 DAVID GINTING 4500000 500 150 Pelanggan Enterprise (Premium)
64 DAVID SIMON PANDAPOTAN S 2300000 350 350 Pelanggan Enterprise (Premium)
65 DEAR MARINA PURBA 7500000 250 150 Pelanggan Enterprise (Premium)
71 DIRMAN HARAPAN HUTAGALUNG 17000000 350 150 Pelanggan Enterprise (Premium)
93 EVITA SOLA GRACIA 4400000 250 350 Pelanggan Enterprise (Premium)
    
```

Hasil data dari cluster 1 : Pelanggan Casual (Hemat)

```

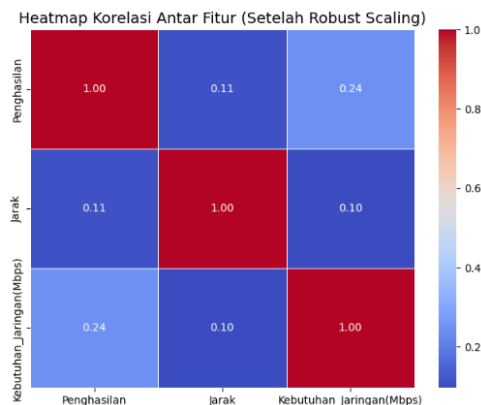
--- detail untuk Cluster 1: Pelanggan Casual (Hemat) ---
Nama_Pelanggan Penghasilan Jarak Kebutuhan_Jaringan(Mbps) Nama Segmen
0 RAFIDA EVELINA DAMANIK 5200000 150 50 Pelanggan Casual (Hemat)
1 MELUTIA NAULY 3500000 250 20 Pelanggan Casual (Hemat)
2 AHMAD JUHRI TINDAON 15000000 100 100 Pelanggan Casual (Hemat)
6 ALICIA ULINA SUSLOWATI 7700000 150 50 Pelanggan Casual (Hemat)
7 YEMIMA SEMBIRING 8400000 500 20 Pelanggan Casual (Hemat)
95 ALFONZO ARUGA BARUS 7700000 500 20 Pelanggan Casual (Hemat)
96 AYRTON WILLIAM SEBASTIAN H 8400000 100 20 Pelanggan Casual (Hemat)
97 YADIDA SIBARANI 2300000 150 10 Pelanggan Casual (Hemat)
98 P PARLINGGOMAN SILAEN 9000000 350 100 Pelanggan Casual (Hemat)
99 MANOTAS BENEDIKTUS SIMAMORA 9100000 250 50 Pelanggan Casual (Hemat)
    
```



Gambar 5 Bar Chart Segmentasi Target Pasar

Visualisasi diatas merepresentasikan banyaknya jumlah pelanggan per kluster yang berhasil tersegmentasi. Dimana bar chart yang berwarna biru mensimbolisasikan pelanggan casual (hemat) dengan besar data hampir 80% dari dataset dan bar chart berwarna hijau merupakan jumlah besar data pelanggan enterprise yang

kisaran datanya kurang lebih hampir 20 data



**Gambar 6 Heatmap Korelasi Fitur**

Berdasarkan heatmap korelasi setelah *Robust Scaling*:

1. **Penghasilan dan Kebutuhan Jaringan (Mbps)** menunjukkan korelasi positif terkuat (sekitar 0.24). Ini berarti pelanggan dengan penghasilan lebih tinggi cenderung memiliki kebutuhan jaringan yang lebih besar, meskipun hubungannya tidak terlalu kuat.
2. **Penghasilan dan Jarak** memiliki korelasi positif yang sangat lemah (sekitar 0.11).
3. **Jarak dan Kebutuhan Jaringan (Mbps)** juga memiliki korelasi positif yang sangat lemah (sekitar 0.10).

Secara keseluruhan, korelasi antar ketiga fitur ini relatif rendah setelah scaling, menunjukkan bahwa masing-masing fitur memberikan informasi yang cukup independen untuk proses clustering. Ini bagus, karena fitur-fitur tersebut tidak terlalu redundan satu sama lain.

Implementasi *k-means* setelah optimasi menghasilkan 2 kluster : *cluster 0* (pelanggan enterprise/premium) dengan karakteristik penghasilan tinggi, kebutuhan jaringan besar (>200 Mbps) dan jarak yang relatif dekat. Adapun *cluster 1* (Pelanggan casual/hemat) dengan penghasilan rendah hingga menengah, kebutuhan jaringan yang relative kecil (<50 Mbps) namun

pelanggan cluster ini memiliki dominasi yang cukup banyak sebagai pelanggan biznet. Evaluasi model menunjukkan peningkatan signifikan: silhouette score dari 0.3133 (belum optimal) menjadi 0.6023 (optimal kluster cukup kuat) dan DBI dari 1.1421 (buruk) menjadi 0.8007 (lebih efisien). Korelasi heatmap mengonfirmasi independensi fitur setelah scaling sehingga klusterisasi lebih representative. Dengan adanya optimasi outlier terhadap *k-means* dengan *robust* mampu mengatasi centroid sehingga tidak lagi terdistorsi nilainya oleh *outlier*.

## SIMPULAN

Optimasi algoritma *k-means* menggunakan *Robust Scaling* secara efektif mampu mengatasi masalah outlier dalam segmentasi target pasar Biznet Medan, kluster yang dihasilkan akurat dan terpisah jelas antar segmen. *Robust* mengganti penggunaan rata – rata dengan Median dan Interquartile range, sehingga centroid tetap stabil dan representatif terhadap mayoritas data akan tetapi centroidnya tidak terpengaruh oleh data ekstrem. Peningkatan metrik evaluasi (*Silhouette Score* > 0.6 dan DBI < 1) membuktikan bahwa klusterisasi *k-means* setelah dioptimalkan dengan *robust* memiliki struktur yang lebih kuat dan efisien, hal ini memungkinkan biznet untuk menggunakan data tersebut untuk merancang strategi pemasaran yang tepat sasaran untuk pelanggan enterprise (premium) dan pelanggan casual (hemat). Hasil ini memberikan kontribusi signifikan pada pengetahuan pengolahan data dengan *outlier*.

## DAFTAR PUSTAKA

- Asosiasi Penyelenggara Jasa Internet Indonesia - Survei. (n.d.). Retrieved April 10, 2026, from <https://survei.apjii.or.id/>
- Hakim, B., Kaunang, F. J., Susanto, C., Salim, J., & Indradjaja, R. (2025).

- Implementasi Machine Learning Dalam Pengelompokan Musik Menggunakan Algoritma K-Means Clustering. *IDEALIS: InDonEsiA Journal Information System*, 8(1), 74–83.  
<https://doi.org/10.36080/idealis.v8i1.3357>
- Harjoyudanto, Y. R., & Zulvia, P. (2023). Klusterisasi Pelanggan Potensial dengan Penggunaan Algoritma K-Means Dalam Mencari Segementasi Pasar (Studi Kasus GUTS Barber and Shop). *Konferensi Nasional Ilmu ...*, 160–166.  
<http://knia.stialanbandung.ac.id/index.php/knia/article/view/959>
- Hasan, Y. (2024). Pengukuran Silhouette Score Dan Davies-Bouldin Index Pada Hasil Cluster K-Means Dan Dbscan. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3S1).  
<https://doi.org/10.23960/jitet.v12i3s1.5001>
- Julian, A. (2025). Segmentasi Pelanggan Menggunakan Algoritma K-Means Pada Jaringan Telekomunikasi Untuk Optimalisasi Strategi Pemasaran. *JTT (Jurnal Teknologi Terpadu)*, 13(1), 94–100.  
<https://doi.org/10.32487/jtt.v13i1.2436>
- Khairul Abdi Sinuraya, Andri Nata, N. A. (2025). *Penanganan Data Outlier Menggunakan Robust Cluster Dalam Pemetaan Dataset Stunting*. 8(4), 5504–5510.
- Khoirunnisa, A., & Ramadhan, N. G. (2023). Improving malaria prediction with ensemble learning and robust scaler: An integrated approach for enhanced accuracy. *Jurnal Infotel*, 15(4), 326–334.  
<https://doi.org/10.20895/infotel.v15i4.1056>
- Kilo, N. A., Katili, M. R., & Hasan, I. K. (2025). Perbandingan Metode K-Means dan K-Medoids Dengan Validitas Davies-Bouldin Indeks, Dunn Indeks dan Indeks Connectivity Pada Pengelompokan Masyarakat Penerima Bantuan Langsung Tunai. *Research in the Mathematical and Natural Sciences*, 4(1), 8–15.  
<https://doi.org/10.55657/rmns.v4i1.190>
- Megiati, Y. E., Komari Pratiwi, N., Nurdiansyah, D., Yusuf, S., & Fauzi, T. R. (2024). Kapas : Kumpulan Artikel Pengabdian Masyarakat Bijak Bermedia Sosial sebagai Bentuk Pemanfaatan Internet Sehat. *Kapas : Kumpulan Artikel Pengabdian Masyarakat*, 2(3), 332–340.
- N Tahta Phudjashakty, Hasbi Firmansyah, Wahyu Asriyani, & Ali Sofyan. (2026). Segmentasi Pelanggan Grosir Menggunakan K-Means: Analisis Outlier dan Ketidakseimbangan Data. *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan*, 4(3), 15993–16002.  
<https://doi.org/10.31004/jerkin.v4i3.4771>
- Nugraha, W., Sabaruddin, R., & Murni, S. (2024). Teknik Scaling Menggunakan Robust Scaler Untuk Mengatasi Outlier Data Pada Model Prediksi Serangan Jantung. *Techno.Com*, 23(2), 319–327.  
<https://doi.org/10.62411/tc.v23i2.10463>
- Putri Azhiah, Jasmir, & Purnama Benni. (2025). Klasifikasi Kelayakan Penerima Beasiswa Menggunakan Naive Bayes dengan Optimasi Atribut Berbasis K-Means. *Institut Riset Dan Publikasi Indonesia*, 5(October), 1450–1462.
- Rahmawati, T., Wilandari, Y., & Kartikasari, P. (2024). Analisis Perbandingan Silhouette Coefficient Dan Metode Elbow Pada Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Ipm Dengan K-Medoids. *Jurnal Gaussian*, 13(1), 13–24.  
<https://doi.org/10.14710/j.gauss.13.1.13-24>
- Sartika, D., Iqbal, M., & Sitorus, Z. (2025). *Analisis Algoritma Exponensial Smoothing dan K-*

- Means untuk Optimalisasi Penerimaan Mahasiswa Baru di Universitas Haji Sumatera Utara.* 07(02), 1–10.
- Sitorus, Z., Iqbal, M., Putera Utama Siahaan, A., & Wahyuni, S. (2025). Analysis of The Sales Potential of BUMDES Products Using the K-Means Clustering Algorithm. *International Journal of Computer Sciences and Mathematics Engineering*, 4(1).
- Sitorus, Z., Syahputra, I., Indra Angkat, C., & Sartika, D. (2024). Implementation of K-Means Clustering for Inventory Projection. *International Journal of Science, Technology & Management*, 5(3), 673–678. <https://doi.org/10.46729/ijstm.v5i3.8>
- Sutiono, S., Wijaya, R. F., & Novelan, M. S. (2025). Analisis Metode K-Means Clustering Untuk Pengelompokan Jadwal Mengajar Guru Di SMKN 1 Stabat Dengan Visualisasi Data. *INTECOMS: Journal of Information Technology and Computer Science*, 8(2), 596–602. <https://doi.org/10.31539/intecom.v8i2.15154>
- Tentang Kami - Biznet Networks.* (n.d.). Retrieved April 22, 2026, from <https://www.biznetnetworks.com/company/about-us>
- Yulianti, G., Chaidir, M., & Pramono, A. S. (2024). Peran Teknologi Keuangan (Fintech) Dalam Mendorong Pertumbuhan Ekonomi Dan Inklusi Keuangan Di Indonesia: Tantangan Dan Peluang. *Citizen: Jurnal Ilmiah Multidisiplin Indonesia*, 4(4), 349–355. <https://doi.org/10.53866/jimi.v4i4.649>