

Pemetaan Kemiskinan Digital Kabupaten/Kota di Sumatera Utara Menggunakan *Ward Hierarchical Clustering*

Digital Poverty Mapping of Regencies/Cities in North Sumatra Using Ward Hierarchical Clustering

*Ari Mulyadi Aberson Silalahi Sidebang

*Politeknik Statistika STIS, E-mail: 222111919@stis.ac.id

ABSTRAK

Pemetaan suatu fenomena merupakan proses mentransformasi data menjadi visualisasi yang mudah dipahami, yang dapat memudahkan pemangku kebijakan dalam mengidentifikasi masalah, menentukan prioritas, dan merumuskan strategi secara efisien dan tepat. Di era digital, teknologi memegang peranan penting dalam mencapai tujuan SDGs, dimana mendorong kemajuan negara dalam mempercepat pembangunan yang berkelanjutan. Kemiskinan digital merujuk pada ketidakmampuan individu atau kelompok dalam mengakses dan memanfaatkan teknologi informasi dan komunikasi (TIK). Berdasarkan Indeks Pembangunan Teknologi Informasi dan Komunikasi (IP-TIK), Sumatera Utara masih berada dalam kategori sedang dan memerlukan perhatian serius dari berbagai pemangku kepentingan, terutama pemerintah. Kurangnya penelitian yang memetakan kemiskinan digital dengan pendekatan *machine learning* menjadi salah satu alasan penelitian ini dilakukan. Metode *Ward Hierarchical Clustering* dipilih karena keunggulannya, seperti ketahanan terhadap pencilon, serta berdasarkan hasil validasi yang ada menunjukkan metode ini sesuai dengan data yang digunakan. Hasil evaluasi tendensi menunjukkan skor sebesar 0,6527 yang mengindikasikan bahwa data cenderung mengelompok dan cocok untuk dilakukan klasterisasi. Berdasarkan *silhouette index*, data dikelompokkan menjadi tiga kluster. Hasil klasterisasi dengan *Ward Hierarchical Clustering*, yaitu kluster 1 (ketimpangan digital tinggi) yang terdiri dari 5 kabupaten/kota, kluster 2 (ketimpangan digital sedang) yang terdiri dari 16 kabupaten/kota, dan kluster 3 (ketimpangan digital rendah) yang terdiri dari 12 kabupaten/kota.

Kata kunci: Kemiskinan Digital, Klasterisasi, Sumatera Utara

ABSTRACT

Mapping a phenomenon is a process of transforming data into easily understood visualizations, which can assist policymakers in identifying issues, setting priorities, and formulating strategies efficiently and accurately. In the digital era, technology plays a crucial role in achieving the SDGs, promoting national progress toward accelerating sustainable development. Digital poverty refers to the inability of individuals or groups to access and utilize information and communication technologies (ICT). Based on the Information and Communication Technology Development Index (ICT-DI), North Sumatra is still categorized as moderate and requires serious attention from various stakeholders, particularly the government. The lack of research that maps digital poverty using a machine learning approach is one of the reasons for conducting this study. The Ward Hierarchical Clustering method was chosen due to its advantages, such as its resistance to outliers, and validation results show that this method is appropriate for the data used. The evaluation of the tendency resulted in a score of 0.6527, indicating that the data tends to cluster and is suitable for clustering analysis. Based on the silhouette index, the data was grouped into three clusters. The clustering results with Ward Hierarchical Clustering yielded cluster 1 (high digital inequality) consisting of 5 districts/cities, cluster 2 (moderate digital inequality) consisting of 16 districts/cities, and cluster 3 (low digital inequality) consisting of 12 districts/cities.

Keywords: Digital Poverty, Clustering, North Sumatra

I. PENDAHULUAN

Kementerian Komunikasi dan Informatika (Kominfo) sebagai focal point pemerintah dalam mengimplementasikan percepatan transformasi digital ke seluruh wilayah Indonesia, menyadari bahwa percepatan pembangunan infrastruktur digital menjadi hal yang urgensi saat ini. Hal ini tercermin melalui Rencana Strategis Kementerian Komunikasi dan Informatika tahun 2020-2024 yang menyatakan bahwa fokus utama Kominfo untuk melakukan percepatan transformasi digital nasional [1]. Dalam laporan riset yang diterbitkan oleh International Institute for Management Development (IMD) pada tahun 2023, Indonesia masih berada dibawah negara-negara di kawasan ASEAN seperti Singapura, Malaysia, dan Thailand dalam hal kesiapan digital. Laporan ini menyoroti bahwa meskipun Indonesia mengalami perkembangan positif dalam beberapa aspek digital, namun Indonesia menghadapi tantangan dalam hal infrastruktur dan Sumber Daya Manusia (SDM) dalam bidang teknologi [2].

Sedangkan berdasarkan Indeks Pembangunan Teknologi Informasi dan Komunikasi (IP-TIK) oleh Badan Pusat Statistik (BPS) yang dikembangkan oleh International Telecommunication Union (ITU) dengan nama ICT Development Index (ICT-DI), menyatakan bahwa Indonesia menunjukkan perkembangan positif dalam lima tahun terakhir dimana terjadi peningkatan nilai IP-TIK dari 5,07 pada tahun 2018 menjadi 5,85 pada tahun 2022. Meskipun terjadi peningkatan tiap tahunnya, Indonesia masih jauh dari skor sepuluh. Begitu juga dengan halnya IP-TIK per provinsi, Sumatera Utara masih berada dalam kategori sedang dan masih perlu menjadi perhatian yang serius oleh stakeholder terutama pemerintah [3].

Sumatera Utara menghadapi tantangan besar dalam mengatasi kemiskinan digital. Meskipun ada kemajuan dalam beberapa aspek digital, banyak daerah di provinsi ini yang masih tertinggal dalam hal infrastruktur teknologi dan pemanfaatan TIK. Kemiskinan digital adalah ketidakmampuan individu dalam menggunakan teknologi informasi, baik itu karena tidak ada akses maupun karena tidak ada kecakapan digital. Dalam forum ekonomi dunia, Thierry Geiger sebagai pakar ekonomi senior mengatakan bahwa teknologi sebagai faktor penting dalam pembangunan nasional yang dapat menjadi solusi dalam menyelesaikan masalah-masalah lain.

Penelitian terdahulu telah melakukan pemetaan kemiskinan digital kabupaten/kota di Jawa Timur dengan metode spasial, namun sejauh ini belum ada yang memetakan kemiskinan digital dengan pendekatan *machine learning*. Penggunaan *machine learning* dalam pemetaan kemiskinan digital menawarkan beberapa keunggulan dibandingkan metode konvensional. Metode spasial memang memberikan gambaran yang cukup baik mengenai distribusi kemiskinan digital berdasarkan variabel-variabel geografis dan demografis. Namun, *machine learning* mampu menangani data dalam jumlah besar dan berbagai jenis data, seperti data tekstual dari laporan atau berita, data numerik dari survei, hingga data spasial dari citra satelit.

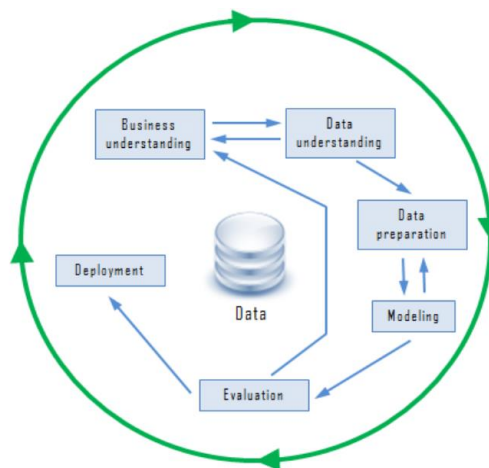
Selain itu, *machine learning* dapat menemukan pola dan tren yang tidak dapat dilihat manusia. Dengan demikian, pendekatan ini dapat memberikan sudut pandang lain yang dapat menjadi bahan pertimbangan stakeholder dalam memetakan kemiskinan digital.

Dalam penelitian ini, peneliti menerapkan analisis kluster untuk memetakan kemiskinan digital kabupaten/kota di Sumatera Utara. Analisis kluster merupakan pendekatan yang digunakan untuk mengidentifikasi kesamaan dalam data dan mengelompokkan data serupa ke dalam kelompok-kelompok atau kluster. Metode *Ward Hierarchical Clustering* dipilih peneliti karena beberapa keunggulannya, seperti efektif pada dataset yang kompleks, dan ketahanannya terhadap pencilan. Artinya, peneliti memungkinkan penggunaan data dengan pencilan tanpa harus mengeluarkannya untuk memastikan representasi data yang lebih komprehensif. Metode ini juga dapat menangani berbagai tipe data, baik numerik maupun kategorik, sehingga sangat cocok untuk analisis data yang kompleks seperti dalam pemetaan kemiskinan digital [4].

Berdasarkan hal-hal di atas, peneliti akan melakukan analisis kluster menggunakan metode *Ward Hierarchical Clustering* untuk mengelompokkan kemiskinan digital kabupaten/kota di Sumatera Utara dengan menggunakan data TIK dari publikasi Statistik Kesejahteraan Rakyat Sumatera Utara 2023 yang bersumber dari hasil Survei Sosial Ekonomi Nasional (Susenans). Penelitian ini mengadopsi kerangka kerja CRISP-DM untuk memastikan proses analisis data yang sistematis dan terstruktur.

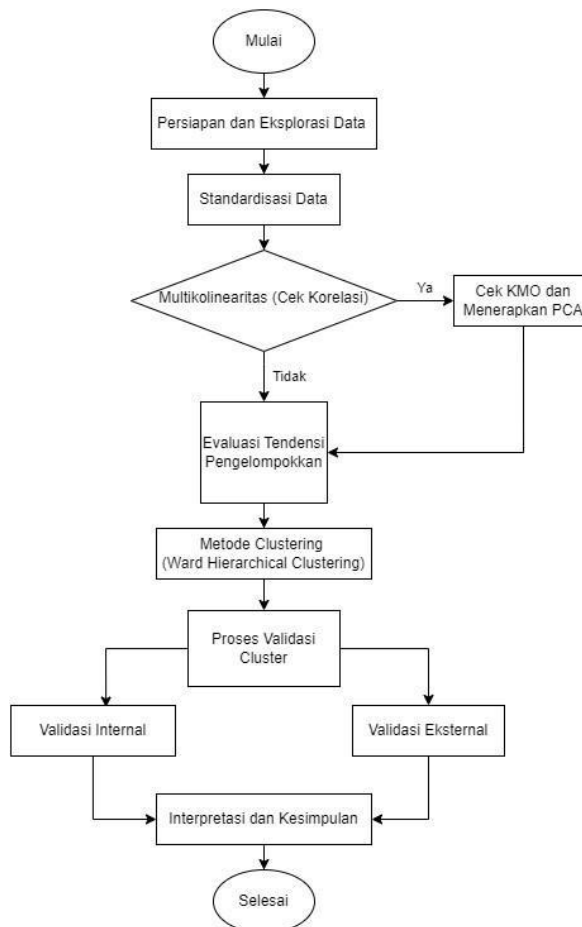
II. METODOLOGI

Penelitian ini menggunakan data sekunder yang diperoleh dari Statistik Kesejahteraan Rakyat Sumatera Utara 2023. Terdapat 6 variabel yang digunakan dalam pengolahan data ini berdasarkan variabel TIK dalam publikasi tersebut. Penelitian ini akan diolah menggunakan perangkat lunak RStudio dan mengikuti kerangka kerja CRISP-DM sebagai landasan metodologi. Model proses CRISP-DM terdiri dari enam tahapan utama, seperti *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.



Gambar 1. Tahapan CRIS-DM [5]

Tahap pertama, *business understanding*, merupakan tahapan terpenting karena di sini akan ditetapkan tujuan dan persyaratan proyek secara jelas, serta merancang strategi untuk mencapai tujuan bisnis dengan menggunakan data yang tersedia. Tahap kedua, *data understanding*, dilakukan untuk memeriksa dan mengidentifikasi masalah dalam data, dengan membuat ringkasan dan visualisasi yang membantu pemahaman mendalam terhadap dataset. Masalah seperti nilai yang *missing*, *outlier*, atau distribusi data yang tidak normal diidentifikasi. Selanjutnya, dalam tahap *data preparation*, data dibersihkan, ditransformasi bila diperlukan sehingga siap untuk dianalisis lebih lanjut. Pada tahap *modeling*, metode statistika dan pembelajaran mesin digunakan untuk membangun model prediktif atau deskriptif. Berbagai teknik data mining seperti klasifikasi, penilaian, perankingan, pengelompokan, penemuan relasi, dan karakterisasi diterapkan sesuai kebutuhan. Selanjutnya, tahap *evaluation*, hasil pemodelan akan dievaluasi berdasarkan validasi yang digunakan sebagai pertimbangan untuk menentukan metode terbaik. Terakhir, tahap *deployment*, melibatkan perencanaan untuk menggunakan model dalam lingkungan operasional, termasuk konversi skor keputusan dan integrasi keputusan dalam sistem yang relevan.



Gambar 2. Diagram Alur Penelitian

Alur penelitian ini dimulai dengan tahap persiapan dan eksplorasi data, dimana data dikumpulkan dan dieksplorasi untuk memahami karakteristik dasar serta pola-pola yang mungkin ada dalam data tersebut. Setelah itu, dilakukan standardisasi data untuk memastikan bahwa semua variabel memiliki skala yang sama. Langkah selanjutnya adalah memeriksa multikolinearitas antar variabel dengan mengecek korelasi. Jika ditemukan korelasi yang tinggi antar variabel, maka diperlukan *Principal Component Analysis* (PCA) untuk mengatasi multikolinearitas yang ada. Asumsi yang harus terpenuhi untuk melakukan *Principal Component Analysis* (PCA) adalah uji *Kaiser Meyer Olkin* (KMO) untuk memastikan kelayakan data. PCA berguna untuk mereduksi dimensi data dengan mempertahankan variabel yang ada sehingga dapat mempertahankan informasi yang ada. Setelah itu, evaluasi tendensi pengelompokan dilakukan untuk mengecek apakah data teridentifikasi mengelompok atau tidak. Jika tidak ditemukan yang korelasi tinggi antar variabel, PCA tidak diperlukan, dan proses dapat dilanjutkan ke tahap pemodelan menggunakan metode *Ward Hierarchical Clustering*. Metode *Ward Hierarchical Clustering* digunakan untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik digital, dimana hasil kluster kemudian dievaluasi menggunakan validasi internal untuk melihat seberapa baik pemodelan yang dihasilkan. Dalam penelitian ini, validasi eksternal tidak digunakan peneliti karena tidak sedang melakukan pengujian hasil pemodelan terhadap data asli.

Setelah validasi dilakukan, interpretasi hasil kluster diperlukan untuk menjelaskan karakteristik setiap kluster berdasar profil tertentu sehingga kesimpulan dapat diambil. Langkah terakhir adalah pemetaan hasil kluster dengan menggunakan perangkat lunak QGIS, yang memberikan visualisasi geografis dari hasil analisis kluster dan membantu dalam memahami distribusi kemiskinan digital di wilayah yang diteliti secara efisien.

2.1 Analisis Kluster

Analisis kluster adalah metode yang digunakan untuk mengelompokkan objek-objek ke dalam kelompok-kelompok atau kluster berdasarkan kesamaan karakteristik. Tujuan utama dari analisis kluster untuk meminimalkan variasi dalam satu kluster dan memaksimalkan variasi antar kluster. Analisis kluster memiliki berbagai manfaat, antara lain mengidentifikasi pola atau struktur dalam data yang tidak terlihat dengan teknik analisis data tradisional, mengelompokkan konsumen ke dalam segmen-segmen pasar berdasarkan karakteristik demografis atau perilaku pembelian, mengurangi kompleksitas data dengan mengelompokkan variabel-variabel yang serupa, serta membuat data yang kompleks menjadi lebih mudah dimengerti dan dianalisis [6].

2.2 Analisis Faktor (PCA)

Uji asumsi perlu dilakukan sebelum menentukan model terbaik dalam suatu analisis. Salah satu uji yang penting adalah uji multikolinearitas. Uji multikolinearitas bertujuan untuk mengetahui ada atau tidaknya korelasi antar variabel bebas dalam suatu penelitian [7]. Ketika terjadi multikolinearitas, salah satu cara yang dapat dilakukan adalah dengan analisis faktor. Analisis faktor bertujuan untuk mereduksi dimensi data dengan mempertahankan variabel yang ada sehingga dapat mempertahankan informasi yang ada. Ini sangat berguna dalam analisis multivariat karena dapat mengatasi variabel yang kompleks dengan mentransformasikannya menjadi faktor utama. Proses analisis faktor meliputi penentuan variabel yang akan dianalisis, *Bartlett's Test of Sphericity*, *Kaiser-Meyer-Olkin (KMO)*, *Measure of Sampling Adequacy (MSA)*, ekstraksi faktor, serta penentuan jumlah faktor utama yang akan digunakan [8].

2.3 Ward Hierarchical Clustering

Ward Hierarchical Clustering merupakan teknik pengelompokan yang bertujuan untuk meminimalkan jumlah kuadrat dari jarak antara titik-titik data dan pusat kluster. Salah satu keunggulan utama metode ini adalah keandalannya terhadap *noise* dan *outlier*. Dibandingkan dengan metode pengelompokan berbasis jarak lainnya, metode *Ward Hierarchical Clustering* lebih robust terhadap *outlier*. Hal ini disebabkan oleh pendekatannya yang mempertimbangkan semua titik data dalam kluster saat menghitung varians, sehingga pengaruh dari *outlier* dapat diminimalkan. Selain itu, metode ini juga sangat efektif dalam menghasilkan kluster yang seimbang dalam hal ukuran, karena berfokus pada meminimalkan varians antar kluster. Proses pengelompokannya yang iteratif dan hirarkis memungkinkan identifikasi struktur data yang lebih dalam, dan hasilnya mudah divisualisasikan dalam bentuk dendrogram. Selain itu, metode ini memungkinkan analisis yang lebih fleksibel dan adaptif terhadap karakteristik data. [9].

Proses pengelompokan dalam metode ini dilakukan secara hierarkis dengan pendekatan *bottom-up*, dimulai dengan setiap titik data sebagai kluster individual. Pada setiap langkah, dua kluster yang menghasilkan peningkatan minimum dalam jumlah kuadrat jarak akan digabungkan. Proses ini berlanjut hingga semua titik data digabungkan menjadi satu kluster besar atau hingga mencapai jumlah kluster yang diinginkan. Hasil dari metode ini sering divisualisasikan menggunakan dendrogram, yaitu diagram pohon yang menunjukkan proses penggabungan kluster secara hierarkis. Dendrogram ini sangat membantu dalam menentukan jumlah optimal kluster dengan mengidentifikasi titik dimana penambahan kluster tidak lagi memberikan pengurangan yang signifikan dalam varian.

III. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Pemerintah Provinsi (Pemprov) Sumatera Utara (Sumut) telah menunjukkan komitmen yang kuat untuk mendorong transformasi digital di berbagai sektor, tidak hanya terbatas pada lingkungan pemerintahan, tetapi juga mencakup sektor-sektor lain yang berpengaruh terhadap pembangunan ekonomi dan kesejahteraan masyarakat. Pemprov Sumatera Utara berkomitmen untuk mengintegrasikan seluruh data dari berbagai instansi, termasuk sektor pendidikan, kesehatan, dan ekonomi, agar menjadi satu wadah yang terpadu. Hal ini bertujuan untuk meningkatkan kualitas layanan publik secara keseluruhan serta mendorong inovasi digital di berbagai bidang. Pemprov Sumut juga aktif berkolaborasi dengan sektor swasta, akademisi, dan masyarakat untuk memperkuat infrastruktur teknologi, yang menjadi fondasi bagi transformasi digital yang lebih luas [10].

Upaya ini tidak hanya sekedar modernisasi, tetapi juga langkah strategis untuk memastikan bahwa semua sektor dapat beroperasi dengan efisiensi yang lebih tinggi dan memiliki akses terhadap data yang akurat dan real-time. Dengan sistem digital yang andal, informasi yang diperlukan untuk pengambilan keputusan dapat diakses dengan lebih mudah dan cepat, yang pada gilirannya mendukung proses pengambilan keputusan yang lebih tepat dan efektif di semua sektor. Transformasi digital di Sumatera Utara juga diharapkan dapat memberikan dampak positif bagi pembangunan ekonomi, dengan membuka peluang baru dalam bidang ekonomi digital, meningkatkan daya saing Usaha Mikro Kecil dan Menengah (UMKM), serta mendorong pertumbuhan lapangan kerja di sektor teknologi. Selain itu, inovasi digital diharapkan dapat meningkatkan kualitas hidup masyarakat melalui peningkatan layanan di bidang kesehatan, pendidikan, dan infrastruktur publik.

Dengan dukungan dari Dinas Komunikasi dan Informatika (Kominfo) Sumut, transformasi ini diharapkan dapat mempercepat proses digitalisasi di seluruh kabupaten/kota di Sumut, tidak hanya di lingkungan pemerintahan tetapi juga di seluruh lapisan masyarakat. Melalui inisiatif ini, Sumut diharapkan dapat mengatasi tantangan-tantangan digital yang ada, sekaligus memanfaatkan peluang-peluang baru yang muncul dari kemajuan teknologi, sehingga membawa manfaat yang luas bagi semua sektor.

Dalam proses transformasi digital, penting untuk memastikan bahwa semua daerah mengalami transformasi secara merata. Ketimpangan dalam akses dan pemanfaatan teknologi informasi di antara wilayah-wilayah dapat menghambat tujuan transformasi digital yang inklusif. Oleh karena itu, pemetaan kondisi digital di Sumatera Utara menjadi fokus dalam penelitian ini untuk mengidentifikasi daerah-daerah yang masih tertinggal dalam penerapan teknologi. Penelitian ini bertujuan untuk mengidentifikasi dan memetakan kabupaten/kota di Sumatera Utara ke dalam beberapa kategori berdasarkan tingkat kemiskinan digital, menggunakan metode *Ward Hierarchical Clustering*. Melalui pemetaan ini, diharapkan dapat

terlihat daerah mana saja yang memerlukan perhatian lebih dalam pengembangan infrastruktur digital, sehingga transformasi digital dapat berjalan secara merata di seluruh wilayah.

3.2 Data Understanding

Terdapat 33 kabupaten/kota di Sumatera Utara yang dianalisis menggunakan berbagai variabel TIK pada Statistik Kesejahteraan Rakyat Sumatera Utara 2023. Variabel-variabel ini digunakan untuk memetakan seluruh kabupaten/kota di Sumatera Utara ke dalam kategori berdasarkan kemiskinan digital. Dataset ini memiliki jumlah record sebanyak 33 kabupaten/kota dan 6 variabel. Dimana variabel yang digunakan, antara lain

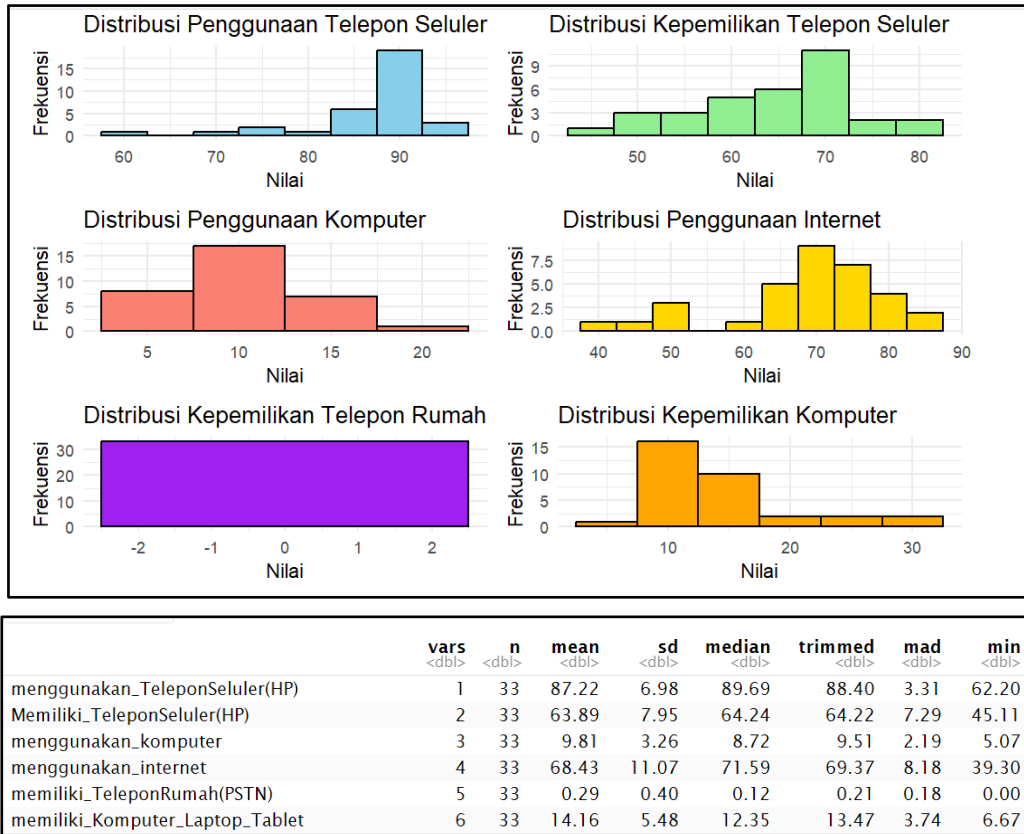
- a) Persentase penduduk berumur 5 tahun ke atas yang menggunakan telepon seluler (hp)/nirkabel untuk keperluan komunikasi dalam 3 bulan terakhir;
- b) Persentase penduduk berumur 5 tahun ke atas yang memiliki/menguasai telepon seluler (hp)/nirkabel dalam 3 bulan terakhir;
- c) Persentase penduduk berumur 5 tahun ke atas yang menggunakan komputer (PC/desktop, laptop/notebook, tablet) dalam 3 bulan terakhir;
- d) Persentase penduduk berumur 5 tahun ke atas pernah menggunakan internet (termasuk Facebook, Twitter, Youtube, Instagram, Whatsapp);
- e) Persentase rumah tangga yang memiliki telepon rumah/PSTN; dan
- f) Persentase rumah tangga yang memiliki komputer/laptop/tablet.

Semua variabel memiliki tipe data numerik, dengan skala yang bervariasi antar variabel. Hal ini terlihat dari perbedaan signifikan dalam nilai maksimum dan minimum di setiap variabel, yang mengindikasikan adanya potensi nilai pencilan.

3.3 Data Preparation

Berikut tahapan dalam *preprocessing* dengan menggunakan RStudio.

- a) Observasi Dataset



Gambar 3. Statistik Deskriptif

Berdasarkan gambar di atas, distribusi penggunaan telepon seluler dan internet memiliki rata-rata yang tinggi, menunjukkan bahwa akses terhadap teknologi komunikasi cukup merata. Namun, terdapat perbedaan signifikan dalam kepemilikan perangkat seperti komputer dan telepon rumah, yang menunjukkan akses teknologi masih terbatas di beberapa daerah. Data ini memberikan wawasan awal tentang sebaran penggunaan teknologi, yang penting untuk memahami pola-pola penggunaan dan potensi kesenjangan akses. Selain itu, statistik deskriptif seperti *mean* dan standar deviasi membantu mengidentifikasi variasi antarwilayah, yang nantinya bisa digunakan untuk analisis lebih dalam.

b) Mendeteksi *Missing Value* dari Variabel yang Digunakan

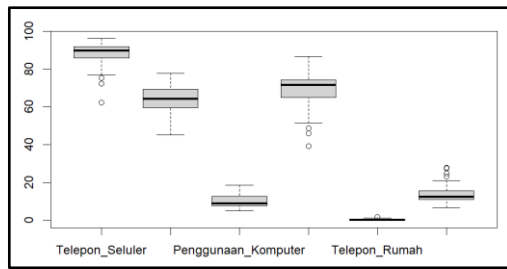
FALSE
198

Gambar 4. Cek *Missing Value*

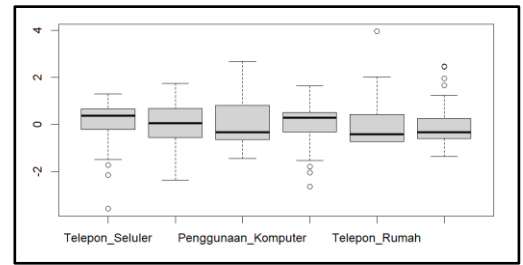
Hasil *FALSE* di atas menunjukkan bahwa tidak ada data yang *missing*. Sedangkan, nilai 198 menunjukkan jumlah total entri yang valid (tidak *missing*) dalam kolom yang diperiksa.

c) Normalisasi/Standardisasi

Sebelum Standardisasi



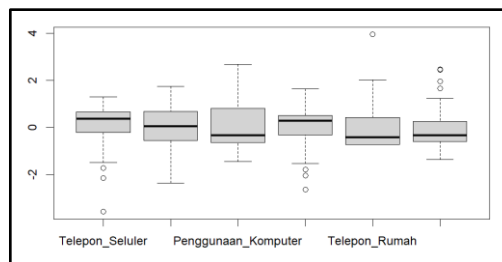
Setelah Standardisasi



Gambar 5. Standardisasi Data

Normalisasi/Standardisasi dataset (*scaling*) digunakan untuk menyamakan skala atau satuan dari setiap atribut yang akan digunakan dalam analisis. Normalisasi dilakukan dengan cara transformasi z-score. Setelah distandarisasi, terlihat pada grafik di atas bahwa setiap variabel sudah berada pada skala yang sama.

d) Mengecek Pencilan

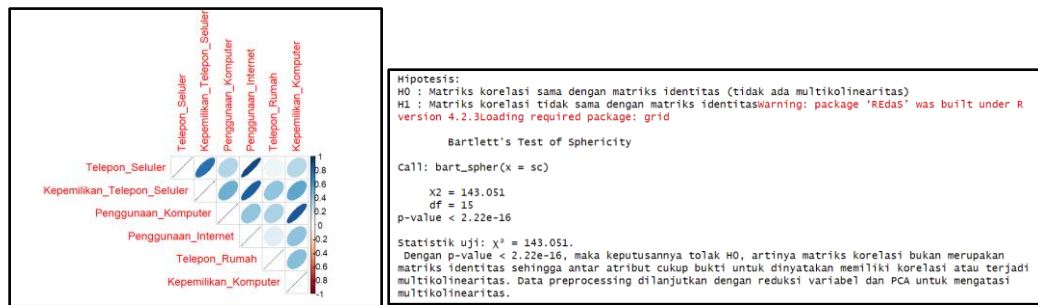


Gambar 6. Mengecek Pencilan

Peneliti di sini mempertahankan pencilan dalam dataset karena hal ini memungkinkan untuk menjaga integritas data asli. Menghapus atau mentransformasi pencilan dikhawatirkan dapat mengubah variasi yang sebenarnya dan menghilangkan informasi penting. Oleh karena itu, untuk memastikan analisis yang tidak bias, diperlukan penggunaan metode kluster yang robust atau tahan terhadap pencilan.

e) Mengecek Korelasi

Setelah memastikan data tidak mengandung *missing value* dan *outlier*, langkah selanjutnya adalah memastikan data tidak mengandung multikolinearitas untuk memeriksa korelasi antar atribut yang akan digunakan dalam klusterisasi. Deteksi multikolinearitas dapat dilakukan dengan melihat plot korelasi antar atribut dan melakukan uji Bartlett.



Gambar 7. Mengecek Korelasi

Dari plot korelasi, terlihat adanya indikasi multikolinearitas, di mana beberapa variabel menunjukkan korelasi yang cukup kuat. Hal ini menandakan adanya hubungan antar atribut yang dapat mempengaruhi hasil analisis. Untuk memperkuat indikasi ini, dilakukan uji Bartlett, yang menghasilkan nilai statistik $\chi^2 = 143.051$ dengan p-value < 2.22e-16. Dengan hasil tersebut, tola H0, yang berarti matriks korelasi bukan merupakan matriks identitas, yang artinya terdapat multikolinearitas. Untuk mengatasi masalah ini, digunakan *Principal Component Analysis* (PCA), yang berfungsi mengubah sekumpulan variabel yang saling berkorelasi menjadi sekumpulan variabel baru yang tidak berkorelasi, yang disebut komponen utama. PCA membantu mereduksi dimensi dataset tanpa kehilangan informasi penting, sehingga analisis *clustering* dapat dilakukan secara lebih efisien dan akurat tanpa dipengaruhi multikolinearitas.

f) *Kaiser Meyer Olkin* (KMO)

Sebelum menerapkan PCA, terlebih dahulu mereduksi atribut yang tidak memenuhi *Kaiser Meyer Olkin* (KMO). Atribut yang tidak memenuhi yakni atribut yang memiliki *Measure of Sampling Adequacy* (MSA) kurang dari 0.5 dapat direduksi (dibuang) dan tidak diikutkan dalam analisis lanjutan.

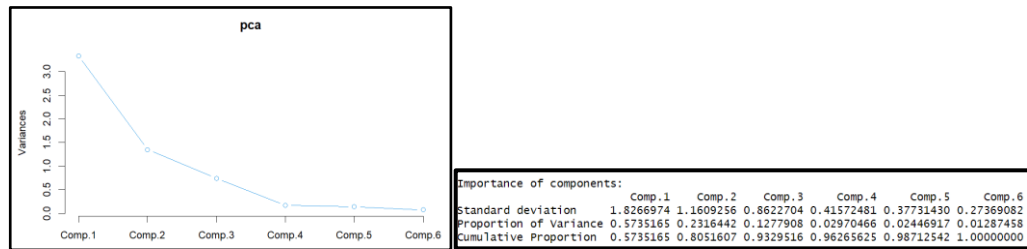
```

Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = sc)
Overall MSA = 0.71
MSA for each item =
      Telepon_Seluler  Kepemilikan_Telepon_Seluler      Penggunaan_Komputer
      0.72              0.82                          0.67
Penggunaan_Internet      Telepon_Rumah      Kepemilikan_Komputer
      0.69              0.61                          0.66
    
```

Gambar 8. KMO

Terlihat tidak ada nilai MSA yang kurang dari 0.5. Oleh karena itu, tidak ada atribut yang harus dikeluarkan serta setiap atribut dapat dilanjutkan pada PCA untuk mengatasi multikolinearitas yang terjadi.

g) *Principal Component Analysis* (PCA)



Gambar 9. PCA

Pada scree plot di atas, terlihat kurva mulai melandai pada komponen utama yang kedua. Kemudian, diperkuat dengan proporsi kumulatif dari PCA pada komponen kedua sudah mencapai minimal 80%. Sehingga kedua komponen tersebut sudah dapat mewakili semua atribut yang digunakan dalam penelitian.

h) Evaluasi Tendensi Pengelompokkan

```

shopkins_stat
[1] 0.6526795

$plot
    
```

Gambar 10. Evaluasi Tendensi Pengelompokkan

Hasil *Hopkins Statistic* sebesar 0.6526, di mana nilai semakin mendekati 1. Artinya, dataset cenderung untuk mengelompok sehingga cocok untuk diklaster.

3.4 Modelling

Peneliti menerapkan berbagai metode *clustering* yang robust terhadap *outlier* untuk memastikan hasil analisis tetap akurat tanpa terlalu dipengaruhi oleh data ekstrem. Metode yang diterapkan, antara lain PAM dan CLARA, yang berbasis medoid dan lebih stabil terhadap pencilaan. Selain itu, peneliti juga menggunakan berbagai metode *Hierarchical Clustering*, seperti *Ward*, *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Centroid Linkage*, dan *Median Linkage*, yang masing-masing memiliki keunggulan dalam menangani struktur data yang berbeda. Metode probabilistik seperti *Gaussian Mixture Models* (GMM) juga diterapkan untuk menangani distribusi data yang kompleks. Berdasarkan validasi internal, metode *Ward* menjadi yang paling baik dalam menghasilkan klaster sesuai data yang ada.

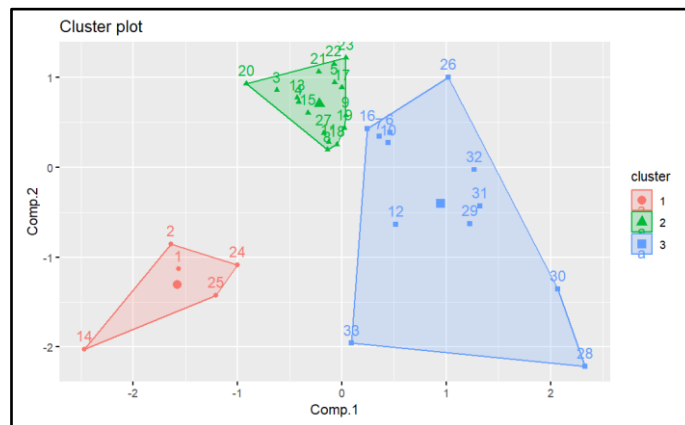
3.4 Evaluation

	Score <dbl>	Method <chr>	Clusters <chr>
Connectivity	4.4690	hierarchical	2
Dunn	0.4089	hierarchical	6
Silhouette	0.5457	hierarchical	3

Gambar 11. Validasi Internal

Berdasarkan hasil validasi internal di atas, terdapat beberapa hasil kluster yang berbeda. Peneliti memutuskan untuk mengacu pada *silhouette index* karena keunggulannya mengukur keseimbangan antara kedekatan titik dalam kluster dan jarak antar kluster. *Silhouette index* memberikan gambaran yang jelas tentang seberapa baik titik data berada dalam kluster yang sesuai dibandingkan dengan kluster lain. Sehingga peneliti akan membagi wilayah kabupaten/kota ke dalam tiga kategori berdasarkan kemiskinan digital.

3.5 Deployment



Gambar 12. Clustering Plot

Tabel 1. Hasil Klasterisasi

Hasil Klaster	Kabupaten/Kota
Klaster 1	Nias Selatan, Mandailing Natal, Nias, Nias Barat, Nias Utara
Klaster 2	Padang Lawas Utara, Tapanuli Selatan, Gunungsitoli, Langkat, Tapanuli Tengah, Humbang Hasundutan, Serdang Bedagai, Tanjungbalai, Asahan, Karo, Padang Lawas, Labuhanbatu Selatan, Tapanuli Utara, Batu Bara, Simalungun, Samosir, Labuanbatu Utara
Klaster 3	Pakpak Bharat, Deli Serdang, Labuhan Batu, Dairi, Toba, Tebing Tinggi, Binjai, Sibolga, Padangsidimpuan, Medan, Pematangsiantar

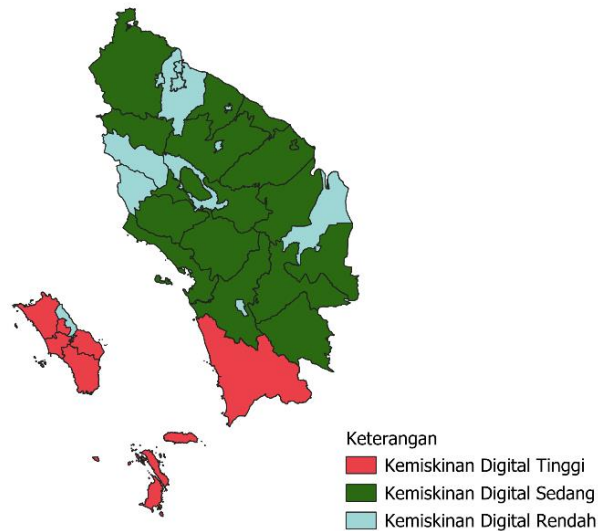
Untuk mengetahui kategori dari setiap klaster digunakan nilai rata-rata pada setiap variabel. Nilai rata-rata ini dijadikan sebagai dasar dalam pengkategorian klaster. Berikut disajikan nilai rata-rata setiap variabel di masing-masing klaster yang dapat dilihat pada Tabel 2.

Tabel 2. Nilai Rata-Rata Setiap Variabel

Variabel	Klaster 1	Klaster 2	Klaster 3
menggunakan_TeleponSeluler(HP)	73,83	88,84	90,64
memiliki_TeleponSeluler(HP)	50,05	63,72	69,89
menggunakan_komputer	8,43	7,64	13,28
menggunakan_internet	46,77	69,97	75,39
memiliki_TeleponRumah(PSTN)	0,26	0,15	0,48
memiliki_Komputer_Laptop_Tablet	11,67	11,28	19,05

Dari hasil *profiling* di atas, untuk mengetahui kategori dari setiap klaster digunakan nilai rata-rata pada setiap variabel. Klaster 1 menunjukkan nilai rata-rata rendah di hampir semua variabel, termasuk penggunaan telepon seluler, internet, dan komputer, serta memiliki nilai rendah untuk telepon rumah. Klaster 2 memiliki nilai rata-rata sedang untuk penggunaan telepon seluler dan internet, serta nilai yang moderat untuk penggunaan komputer dan telepon rumah. Sementara itu, klaster 3 memiliki nilai rata-rata tertinggi di semua variabel, menunjukkan penggunaan telepon seluler, internet, dan komputer yang tinggi, serta akses yang lebih baik terhadap telepon rumah. Oleh karena itu, klaster 1 dikategorikan sebagai kemiskinan digital yang tinggi, klaster 2 sebagai kemiskinan digital sedang, dan klaster 3 sebagai kemiskinan digital yang rendah.

IV. KESIMPULAN



Gambar 13. Visualisasi Pemetaan Hasil Klasterisasi dengan QGIS

Penelitian ini bertujuan untuk memetakan kemiskinan digital di kabupaten/kota di Sumatera Utara dengan menggunakan metode *Ward Hierarchical Clustering*. Pemetaan ini dilakukan dengan *machine learning* berbasis *clustering*. Sebelum penerapan model, evaluasi tendensi dilakukan dengan uji Hopkins, diperoleh sebesar 0,6526795 yang menunjukkan bahwa data cenderung mengelompok dan cocok dilakukan klasterisasi. Kemudian, berdasarkan evaluasi menurut validasi yang diimplementasikan ke beberapa metode klaster yang robust terhadap pencilan, ditemukan bahwa metode *Ward Hierarchical Clustering* adalah yang paling terbaik untuk data ini.

Hasil validasi internal menunjukkan adanya beberapa hasil klaster yang berbeda. Namun, peneliti memutuskan untuk menggunakan *silhouette index* sebagai indikator utama, karena metrik ini secara efektif mengukur keseimbangan antara kedekatan titik dalam klaster dan jarak antar klaster. *silhouette index* memberikan gambaran yang jelas mengenai seberapa baik titik data terkelompok dalam klaster yang sesuai dibandingkan dengan klaster lainnya. Berdasarkan analisis tersebut, wilayah kabupaten/kota dibagi ke dalam tiga kategori berdasarkan tingkat kemiskinan digital, antara lain

a) Klaster 1 (Kemiskinan Digital Tinggi)

Mencakup kabupaten/kota dengan tingkat penggunaan teknologi yang sangat rendah. Variabel seperti penggunaan telepon seluler, internet, dan komputer berada pada tingkat rendah, menunjukkan adanya kemiskinan digital yang signifikan dan memerlukan perhatian khusus dalam pembangunan infrastruktur teknologi. Terdapat 5 kabupaten/kota yang termasuk dalam kategori ini, antara lain Nias Selatan, Mandailing Natal, Nias, Nias Barat, Nias Utara.

b) Klaster 2 (Kemiskinan Digital Sedang)

Mencakup kabupaten/kota dengan tingkat penggunaan teknologi yang moderat. Meskipun ada kemajuan, terdapat gap dalam akses dan pemanfaatan teknologi jika dibandingkan dengan klaster yang lebih maju. Daerah-daerah ini menunjukkan kebutuhan untuk meningkatkan akses dan keterampilan digital. Terdapat 16 kabupaten/kota yang termasuk dalam kategori ini, antara lain Padang Lawas Utara, Tapanuli Selatan, Gunungsitoli, Langkat, Tapanuli Tengah, Humbang Hasundutan, Serdang Bedagai, Tanjungbalai, Asahan, Karo, Padang Lawas, Labuhanbatu Selatan, Tapanuli Utara, Batu Bara, Simalungun, Samosir, Labuanbatu Utara.

c) Klaster 3 (Kemiskinan Digital Rendah)

Mencakup kabupaten/kota dengan tingkat penggunaan teknologi yang relatif tinggi. Di klaster ini, variabel-variabel seperti penggunaan telepon seluler, internet, dan komputer menunjukkan nilai rata-rata yang tinggi, menandakan akses yang baik terhadap teknologi dan keterampilan digital yang lebih baik. Terdapat 12 kabupaten/kota yang termasuk dalam kategori ini, antara lain Pakpak Bharat, Deli Serdang, Labuhan Batu, Dairi, Toba, Tebing Tinggi, Binjai, Sibolga, Padangsidimpuan, Medan, Pematangsiantar.

DAFTAR PUSTAKA

- Kominfo. (2021). Rencana Strategis Kementerian Kominfo 2020 - 2024 untuk Percepatan Transformasi Digital Nasional. *Kementerian Komunikasi dan Informatika Republik Indonesia*. <https://www.kominfo.go.id/content/detail/35108/rencana-strategis-kementerian-kominfo-2020-2024-untuk-percepatan-transformasi-digital-nasional/0/pengumuman>. Diakses pada tanggal 14 Agustus 2024.
- Kominfo. (2024). Daya Saing Digital Indonesia. *Kementerian Komunikasi dan Informatika Republik Indonesia*. <https://www.kominfo.go.id/content/detail/54774/daya-saing-digital-indonesia/0/artikel>. Diakses pada tanggal 14 Agustus 2024.
- BPS. (2023). Indeks Pembangunan Teknologi Informasi dan Komunikasi 2022. *Badan Pusat Statistik Republik Indonesia*. <https://www.bps.go.id/id/publication/2023/09/29/cfa3a7c9e8b2397799ec6bb3/indeks-pembangunan-teknologi-informasi-dan-komunikasi-2022.html>. Diakses pada tanggal 15 Agustus 2024.
- Xu, R., & Wunsch, D. (2008). *Clustering*. John Wiley & Sons.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guides.
- Simamora, B. (2005). *Analisis Multivariat Pemasaran*. Edisi 1. Jakarta: PT Gramedia Pustaka.

- Ghozali, I. (2018). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25*. Semarang: Badan Penerbit Universitas Diponegoro.
- Subagja, & Disan. (2020). Analisis Motivasi Karyawan Terhadap Keselamatan Dan Kesehatan Kerja Menggunakan Analisis Faktor Di PT. Perkebunan Nusantara VIII. Skripsi, Universitas Komputer Indonesia.
- Samin, D. (2020). Hierarchical Clustering Algorithm in Machine Learning. *Medium*. Diakses dari <https://medium.com/@dhivyasamin/hierarchical-clustering-algorithm-in-machine-learning-bd9f97ca6701>.
- Pemerintah Provinsi Sumatera Utara. (2024). Dinas Kominfo Sumut Terus Percepat Transformasi Digital Pemprov Sumut. *Pemerintah Provinsi Sumatera Utara*. <https://sumutprov.go.id/artikel/artikel/dinas-kominfo-sumut-terus-percepat-transformasi-digital-pemprov-sumut>. Diakses pada tanggal 15 Agustus 2024.