

Implementasi K-Nearest Neighbor Untuk Prediksi Karakteristik Penghuni Citraland Gama City

Implementation Of The K-Nearest Neighbor For Predicting The Characteristics Of Residents Of Citraland Gama City

Agustian Gunawan^a, Mohammad Irfan Fahmi^{b*}

PUI Inovasi Teknologi Ilmu Komputer, Universitas Prima Indonesia Medan^{a,b}
agustiangunawan7694@gmail.com^a, mohammadirfanfahmi@unprimdn.ac.id^b

Abstract

Data management of residents at Citraland Gama City still faces various issues, such as suboptimal utilization of resident data, manual data analysis processes, a high risk of errors in data grouping, and the absence of a system capable of automatically predicting resident characteristics. This condition makes it difficult for management to accurately identify patterns and characteristics of residents, resulting in less effective decision-making regarding facility provision and services. In addition, the lack of implementation of data mining-based classification methods, such as K-Nearest Neighbor (KNN), makes data processing less efficient amid the increasing number of residents. Therefore, the application of the KNN method is needed to help predict and classify resident characteristics in a faster, more systematic, and more accurate way, thereby supporting more optimal residential area management. This study uses the K-Nearest Neighbor (KNN) method to predict the characteristics of residents of Citraland Gama City based on data such as age, occupation, income, number of family members, and housing type. The research stages include data collection, preprocessing, KNN implementation, and accuracy testing. The analysis is carried out through Euclidean Distance calculations, determining the value of K, and evaluating prediction results to determine the level of accuracy of the method in classifying resident characteristics. Based on the results of the study, the implementation of the K-Nearest Neighbor (KNN) method for predicting the characteristics of residents of Citraland Gama City in Medan City has been proven to help the process of classifying resident data more quickly and systematically. The KNN method can group resident characteristics based on variables such as age, occupation, income level, number of family members, and housing type. The test results show that the KNN method produces an accuracy level of 79%, indicating that this method is quite good for predicting resident characteristics.

Keywords: K-Nearest Neighbor, Resident Characteristics

Abstrak

Pengelolaan data penghuni di Citraland Gama City masih menghadapi berbagai permasalahan, seperti belum optimalnya pemanfaatan data penghuni, proses analisis yang masih dilakukan secara manual, tingginya risiko kesalahan dalam pengelompokan data, serta belum adanya sistem yang mampu memprediksi karakteristik penghuni secara otomatis. Kondisi ini menyebabkan pihak pengelola kesulitan mengidentifikasi pola dan karakteristik penghuni secara akurat sehingga pengambilan keputusan terkait penyediaan fasilitas dan pelayanan menjadi kurang efektif. Selain itu, belum diterapkannya metode klasifikasi berbasis data mining, seperti K-Nearest Neighbor (KNN), membuat proses pengolahan data menjadi kurang efisien di tengah meningkatnya jumlah penghuni. Oleh karena itu, diperlukan penerapan metode KNN untuk membantu proses prediksi dan klasifikasi karakteristik penghuni secara lebih cepat, sistematis, dan akurat sehingga dapat mendukung pengelolaan kawasan perumahan yang lebih optimal. Penelitian ini menggunakan metode K-Nearest Neighbor (KNN) untuk memprediksi karakteristik penghuni Citraland Gama City berdasarkan data usia, pekerjaan, pendapatan, jumlah anggota keluarga, dan tipe hunian. Tahapan penelitian meliputi pengumpulan data, preprocessing, implementasi KNN, serta pengujian akurasi. Analisis dilakukan melalui perhitungan jarak Euclidean Distance, penentuan nilai K, dan evaluasi hasil prediksi untuk mengetahui tingkat ketepatan metode dalam mengklasifikasikan karakteristik penghuni. Berdasarkan hasil penelitian, implementasi metode K-Nearest Neighbor (KNN) untuk prediksi karakteristik penghuni Citraland Gama City Kota Medan terbukti mampu membantu proses klasifikasi data penghuni secara lebih cepat dan sistematis. Metode KNN dapat mengelompokkan karakteristik penghuni berdasarkan variabel usia, pekerjaan, tingkat pendapatan, jumlah anggota keluarga, dan tipe hunian. Hasil pengujian menunjukkan

bahwa metode KNN menghasilkan tingkat akurasi sebesar 79%, sehingga metode ini cukup baik digunakan untuk memprediksi karakteristik penghuni.

Kata Kunci: K-Nearest Neighbor, Karakteristik Penghuni

1. Pendahuluan

Perkembangan sektor properti dan kawasan perumahan di Kota Medan mengalami peningkatan yang cukup signifikan seiring dengan bertambahnya kebutuhan masyarakat terhadap hunian yang nyaman, modern, dan strategis. Banyak perusahaan pengembang berlomba-lomba menghadirkan kawasan perumahan dengan fasilitas yang lengkap guna menarik minat masyarakat. Menurut Limbong, Simarmata, dan Iskandar (2022), pemanfaatan data yang tepat dapat memberikan informasi yang bernilai dalam mendukung berbagai kebutuhan pengambilan keputusan. Salah satu kawasan perumahan yang berkembang di Kota Medan adalah Citraland Gama City yang menjadi salah satu kawasan hunian dengan jumlah penghuni yang terus meningkat setiap tahunnya. Bertambahnya jumlah penghuni tersebut secara tidak langsung menyebabkan jumlah data penghuni yang dimiliki oleh pihak pengelola maupun pengembang juga semakin banyak dan kompleks.

Data penghuni yang dimiliki sebenarnya memiliki potensi besar untuk dijadikan sebagai sumber informasi dalam mendukung proses pengambilan keputusan. Data tersebut meliputi usia penghuni, pekerjaan, tingkat pendapatan, jumlah anggota keluarga, tipe rumah, serta berbagai informasi lain yang dapat digunakan untuk mengetahui karakteristik penghuni secara lebih mendalam. Namun pada kenyataannya, data-data tersebut belum dimanfaatkan secara optimal dan sebagian besar hanya dijadikan sebagai arsip administrasi semata. Kondisi ini menyebabkan data penghuni yang jumlahnya terus bertambah tidak mampu memberikan informasi yang bermanfaat bagi pihak pengelola kawasan perumahan.

Permasalahan yang terjadi saat ini adalah pihak pengelola masih mengalami kesulitan dalam mengidentifikasi dan mengelompokkan karakteristik penghuni berdasarkan data yang dimiliki. Pengelompokan penghuni masih dilakukan secara sederhana melalui pengamatan langsung atau pencatatan manual tanpa menggunakan metode pengolahan data tertentu. Akibatnya, proses identifikasi karakteristik penghuni menjadi tidak efektif karena membutuhkan waktu yang lama dan hasil yang diperoleh seringkali kurang akurat. Kondisi ini semakin sulit ketika jumlah penghuni terus meningkat sehingga volume data yang harus dianalisis menjadi semakin besar.

Selain itu, proses analisis data penghuni yang dilakukan secara manual juga rentan terhadap kesalahan dalam pengelompokan data. Kesalahan tersebut dapat terjadi karena adanya data yang tidak konsisten, kesalahan pencatatan, maupun keterbatasan kemampuan manusia dalam menganalisis data dalam jumlah besar. Akibatnya, informasi yang dihasilkan tidak mampu menggambarkan kondisi karakteristik penghuni secara tepat. Hal ini menyebabkan pihak pengelola kesulitan dalam menentukan kelompok penghuni tertentu berdasarkan pola karakteristik yang dimiliki.

Permasalahan lainnya adalah belum adanya sistem yang mampu melakukan prediksi karakteristik penghuni secara otomatis berdasarkan data penghuni sebelumnya. Selama ini, proses penentuan karakteristik penghuni hanya dilakukan berdasarkan asumsi umum atau pengamatan subjektif tanpa menggunakan metode analisis data yang terstruktur. Kondisi tersebut menyebabkan hasil identifikasi karakteristik penghuni seringkali tidak memiliki tingkat akurasi yang jelas. Apabila

proses ini terus dilakukan secara manual, maka kemungkinan terjadinya kesalahan analisis akan semakin besar seiring bertambahnya jumlah data penghuni.

Kurangnya pemanfaatan teknologi dalam pengolahan data penghuni juga menjadi salah satu masalah yang cukup penting. Di era perkembangan teknologi saat ini, banyak bidang telah memanfaatkan sistem berbasis data mining dan machine learning untuk membantu proses pengambilan keputusan secara lebih cepat dan akurat. Namun pada kenyataannya, penerapan teknologi tersebut dalam pengelolaan data penghuni kawasan perumahan masih sangat minim. Sebagian besar data penghuni masih dikelola menggunakan cara konvensional sehingga proses analisis data menjadi kurang efektif dan sulit menghasilkan informasi yang akurat.

Data penghuni yang tidak diolah dengan baik menyebabkan pihak pengelola sulit mengetahui pola-pola tertentu yang sebenarnya dapat dijadikan sebagai dasar dalam pengambilan keputusan. Misalnya, pihak pengelola kesulitan mengetahui kelompok penghuni berdasarkan tingkat ekonomi, jenis pekerjaan dominan, atau karakteristik keluarga yang tinggal di kawasan tersebut. Ketidakmampuan dalam mengetahui pola karakteristik penghuni menyebabkan berbagai keputusan yang diambil seringkali tidak sesuai dengan kondisi penghuni yang sebenarnya.

Selain itu, ketidakakuratan dalam mengidentifikasi karakteristik penghuni juga dapat berdampak pada kurang optimalnya penyediaan fasilitas dan pelayanan di lingkungan perumahan. Pengelola kawasan menjadi kesulitan menentukan fasilitas apa yang paling dibutuhkan oleh penghuni karena tidak memiliki informasi karakteristik penghuni yang jelas. Akibatnya, fasilitas yang disediakan terkadang tidak sesuai dengan kebutuhan mayoritas penghuni sehingga tingkat kepuasan penghuni terhadap lingkungan perumahan menjadi kurang optimal.

Permasalahan lainnya adalah belum adanya metode klasifikasi data yang diterapkan untuk membantu proses pengelompokan karakteristik penghuni secara sistematis. Padahal, data penghuni yang tersedia dapat dianalisis menggunakan metode tertentu untuk menghasilkan informasi yang lebih akurat. Tanpa adanya metode klasifikasi, data penghuni hanya menjadi kumpulan data biasa yang sulit dimanfaatkan sebagai sumber informasi penting. Hal ini menyebabkan proses pengelolaan data menjadi kurang maksimal dan tidak mampu mendukung pengambilan keputusan secara efektif.

Dalam bidang data mining, terdapat berbagai metode klasifikasi yang dapat digunakan untuk membantu proses prediksi dan pengelompokan data, salah satunya adalah metode K-Nearest Neighbor (KNN). Menurut Siregar (2022), klasifikasi merupakan suatu proses pengelompokan data berdasarkan karakteristik tertentu sehingga dapat digunakan untuk membantu pengambilan keputusan. Metode K-Nearest Neighbor merupakan metode klasifikasi yang bekerja berdasarkan kedekatan antar data sehingga mampu menentukan kategori suatu data baru berdasarkan data sebelumnya yang memiliki karakteristik serupa. Namun pada kenyataannya, metode KNN belum diterapkan dalam proses prediksi karakteristik penghuni di Citraland Gama City. Padahal metode tersebut dapat membantu pihak pengelola dalam mengolah data penghuni secara lebih cepat dan sistematis.

Tidak diterapkannya metode klasifikasi seperti KNN menyebabkan proses analisis data penghuni masih dilakukan secara manual dan kurang efisien. Menurut Windarto, Wanto, dan Hartama (2023), penerapan data mining sangat diperlukan ketika jumlah data yang diolah semakin besar karena mampu membantu proses pengelompokan dan analisis data secara lebih efektif. Kondisi ini menyebabkan proses

analisis membutuhkan waktu yang lama dan hasil yang diperoleh seringkali tidak konsisten. Selain itu, tanpa adanya sistem prediksi, pihak pengelola juga mengalami kesulitan dalam menentukan pola karakteristik penghuni berdasarkan data yang tersedia.

Di sisi lain, perkembangan teknologi informasi sebenarnya telah menyediakan berbagai solusi untuk membantu proses pengolahan data dalam jumlah besar. Pemanfaatan metode data mining seperti K-Nearest Neighbor dapat membantu proses klasifikasi data penghuni secara otomatis berdasarkan pola kemiripan tertentu. Dengan menggunakan metode tersebut, proses prediksi karakteristik penghuni dapat dilakukan secara lebih cepat, akurat, dan terstruktur dibandingkan dengan metode manual. Menurut Prasetyo (2022), data mining merupakan proses untuk menemukan pola dan informasi penting dari sekumpulan data sehingga dapat dimanfaatkan dalam mendukung pengambilan keputusan. Namun kurangnya penerapan metode tersebut menyebabkan pengelolaan data penghuni di Citraland Gama City masih belum optimal.

Berdasarkan berbagai permasalahan tersebut, dapat diketahui bahwa pengelolaan data penghuni di Citraland Gama City masih memiliki berbagai kendala, khususnya dalam proses identifikasi dan prediksi karakteristik penghuni. Banyaknya data penghuni yang belum diolah secara optimal, proses analisis yang masih manual, tingginya risiko kesalahan pengelompokan data, serta belum adanya penerapan metode klasifikasi menjadi masalah utama yang perlu diselesaikan. Oleh karena itu, diperlukan suatu penelitian yang mampu mengimplementasikan metode K-Nearest Neighbor untuk memprediksi karakteristik penghuni Citraland Gama City berdasarkan data yang tersedia sehingga proses klasifikasi data dapat dilakukan secara lebih efektif, sistematis, dan akurat.

2. Metode

Penelitian ini menggunakan pendekatan data mining dengan metode klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN) untuk memprediksi karakteristik penghuni Citraland Gama City berdasarkan data yang tersedia. Penelitian ini termasuk dalam penelitian eksperimen komputasional karena berfokus pada penerapan dan pengujian suatu algoritma untuk menghasilkan klasifikasi serta mengukur tingkat akurasi. Pendekatan yang digunakan adalah deskriptif kuantitatif, yaitu menggambarkan kondisi data penghuni sekaligus mengolah data dalam bentuk numerik agar dapat dianalisis menggunakan metode KNN. Variabel yang digunakan dalam penelitian ini meliputi usia, pekerjaan, tingkat pendapatan, jumlah anggota keluarga, dan tipe hunian.

Penelitian dilaksanakan di Citraland Gama City Kota Medan sebagai objek penelitian karena memiliki data penghuni yang beragam dan sesuai untuk dianalisis menggunakan metode data mining. Proses penelitian dilakukan melalui beberapa tahapan yang dimulai dari identifikasi masalah, pengumpulan data, pengolahan data, implementasi algoritma, hingga evaluasi hasil. Data yang dikumpulkan berupa data penghuni yang mencakup usia, pekerjaan, pendapatan, jumlah anggota keluarga, dan tipe hunian. Data tersebut kemudian dipersiapkan melalui proses pembersihan data untuk menghilangkan data yang tidak lengkap atau tidak valid sehingga dapat digunakan dalam proses analisis.

Setelah proses pembersihan data selesai dilakukan, tahap berikutnya adalah preprocessing data yang meliputi transformasi dan normalisasi data. Transformasi dilakukan dengan mengubah data kategorikal menjadi data numerik agar dapat

diproses oleh algoritma KNN. Selanjutnya dilakukan normalisasi untuk menyamakan skala setiap variabel sehingga tidak terjadi dominasi nilai tertentu dalam proses perhitungan jarak. Tahap ini sangat penting karena metode KNN bekerja berdasarkan kedekatan antar data sehingga kualitas data yang digunakan akan mempengaruhi hasil klasifikasi.

Data yang telah diproses kemudian dibagi menjadi data training dan data testing. Data training digunakan sebagai dasar pembentukan pola klasifikasi, sedangkan data testing digunakan untuk menguji kemampuan model dalam memprediksi data baru. Pada tahap implementasi, algoritma K-Nearest Neighbor diterapkan dengan menentukan nilai K sebagai jumlah tetangga terdekat yang digunakan dalam proses klasifikasi. Jarak antara data testing dan data training dihitung menggunakan metode Euclidean Distance untuk mengetahui tingkat kemiripan antar data. Berdasarkan hasil perhitungan jarak tersebut, sistem akan menentukan kelas atau karakteristik penghuni berdasarkan mayoritas data tetangga terdekat.

Untuk mendukung pelaksanaan penelitian, digunakan perangkat keras berupa laptop atau komputer sebagai media pengolahan data serta perangkat pendukung lainnya. Perangkat lunak yang digunakan meliputi sistem operasi Windows, Microsoft Excel untuk pengolahan awal data, dan aplikasi data mining untuk implementasi algoritma KNN. Data penghuni yang digunakan dalam penelitian ini menjadi bahan utama dalam proses klasifikasi dan prediksi karakteristik penghuni.

Tahap terakhir dalam penelitian adalah evaluasi hasil klasifikasi yang diperoleh dari algoritma KNN. Evaluasi dilakukan dengan membandingkan hasil prediksi dengan data aktual untuk mengetahui tingkat ketepatan model yang dibangun. Hasil evaluasi digunakan untuk mengukur kemampuan metode K-Nearest Neighbor dalam memprediksi karakteristik penghuni Citraland Gama City secara akurat dan konsisten. Dengan demikian, penelitian ini diharapkan dapat memberikan solusi dalam membantu pengelola kawasan perumahan mengelompokkan dan memahami karakteristik penghuni berdasarkan data yang dimiliki secara lebih efektif dan sistematis.

3. Hasil Dan Pembahasan

Analisis Data Awal

Analisis data awal dilakukan untuk mengetahui gambaran umum karakteristik penghuni sebelum data diproses menggunakan metode KNN. Analisis ini bertujuan untuk mengidentifikasi pola distribusi data sehingga dapat memberikan pemahaman awal mengenai kondisi penghuni Citraland Gama City.

Tabel 1. Distribusi Karakteristik Penghuni

No	Karakteristik Penghuni	Jumlah	Persentase
1	Keluarga Muda	35	29%
2	Keluarga Berkembang	48	40%
3	Keluarga Mapan	37	31%
Total	120	100%	

Berdasarkan Tabel di atas dapat diketahui bahwa kelompok Keluarga Berkembang merupakan kelompok yang paling dominan dengan jumlah 48 penghuni atau sebesar 40% dari total data penelitian. Kelompok ini umumnya terdiri dari penghuni yang berada pada usia produktif, memiliki pekerjaan tetap, serta sedang berada pada fase peningkatan kesejahteraan keluarga. Kelompok Keluarga Mapan berada pada posisi kedua dengan jumlah 37 penghuni atau sebesar 31%. Penghuni

pada kategori ini umumnya memiliki tingkat pendapatan yang lebih tinggi dan menempati rumah dengan tipe yang lebih besar dibandingkan kategori lainnya. Sementara itu, kelompok Keluarga Muda berjumlah 35 penghuni atau sebesar 29%. Kelompok ini terdiri dari pasangan muda atau keluarga yang baru membangun kehidupan rumah tangga sehingga jumlah anggota keluarga relatif lebih sedikit dibandingkan kelompok lainnya. Hasil distribusi tersebut menunjukkan bahwa sebagian besar penghuni Citraland Gama City berada pada kategori keluarga berkembang dan keluarga mapan. Kondisi ini mengindikasikan bahwa kawasan perumahan tersebut didominasi oleh masyarakat dengan tingkat ekonomi menengah hingga menengah atas.

Tabel 2. Distribusi Usia Penghuni

No	Kelompok Usia	Jumlah	Persentase
1	20–30 Tahun	20	17%
2	31–40 Tahun	42	35%
3	41–50 Tahun	36	30%
4	>50 Tahun	22	18%
Total	120	100%	

Berdasarkan Tabel di atas dapat diketahui bahwa kelompok usia 31–40 tahun merupakan kelompok usia terbanyak dengan jumlah 42 penghuni atau sebesar 35%. Kelompok usia ini termasuk usia produktif yang umumnya telah memiliki pekerjaan tetap, pendapatan yang stabil, serta kebutuhan hunian yang lebih tinggi. Kelompok usia 41–50 tahun menempati posisi kedua dengan jumlah 36 penghuni atau sebesar 30%. Penghuni pada kelompok ini sebagian besar telah memiliki kondisi ekonomi yang relatif mapan dan cenderung menetap dalam jangka waktu yang lebih lama. Sementara itu, kelompok usia di atas 50 tahun berjumlah 22 penghuni atau sebesar 18%, sedangkan kelompok usia 20–30 tahun berjumlah 20 penghuni atau sebesar 17%. Berdasarkan hasil tersebut dapat disimpulkan bahwa mayoritas penghuni Citraland Gama City berada pada usia produktif. Kondisi ini menunjukkan bahwa kawasan perumahan tersebut memiliki tingkat aktivitas ekonomi yang cukup tinggi karena sebagian besar penghuninya masih aktif bekerja dan menjalankan aktivitas profesional.

Distribusi Pekerjaan Penghuni

Pekerjaan merupakan salah satu variabel penting dalam penelitian ini karena dapat memberikan gambaran mengenai kondisi sosial ekonomi penghuni Citraland Gama City. Jenis pekerjaan memiliki hubungan yang cukup erat dengan tingkat pendapatan, kemampuan ekonomi, serta tipe hunian yang dimiliki oleh penghuni. Oleh karena itu, analisis terhadap pekerjaan penghuni perlu dilakukan untuk mengetahui karakteristik umum masyarakat yang tinggal di kawasan perumahan tersebut.

Tabel 3. Distribusi Pekerjaan Penghuni

No	Pekerjaan	Jumlah	Persentase
1	Pegawai Swasta	45	38%
2	Wiraswasta	34	28%
3	Profesional	25	21%
4	PNS	16	13%
Total	120	100%	

Berdasarkan Tabel di atas dapat diketahui bahwa pekerjaan yang paling dominan adalah Pegawai Swasta dengan jumlah 45 penghuni atau sebesar 38% dari total data penelitian. Jumlah tersebut menunjukkan bahwa sebagian besar penghuni Citraland Gama City bekerja pada sektor swasta yang meliputi berbagai bidang usaha dan industri. Kelompok Wiraswasta menempati posisi kedua dengan jumlah 34 penghuni atau sebesar 28%. Tingginya jumlah penghuni yang bekerja sebagai wiraswasta menunjukkan bahwa kawasan perumahan ini juga dihuni oleh masyarakat yang memiliki usaha mandiri dan aktivitas bisnis yang cukup aktif. Selanjutnya kelompok Profesional berjumlah 25 penghuni atau sebesar 21%. Kelompok ini terdiri dari berbagai profesi seperti dokter, pengacara, konsultan, akuntan, arsitek, dan profesi lainnya yang membutuhkan keahlian khusus. Keberadaan kelompok profesional menunjukkan bahwa kawasan Citraland Gama City memiliki daya tarik bagi masyarakat dengan tingkat pendidikan dan kompetensi yang relatif tinggi. Sementara itu, kelompok Pegawai Negeri Sipil (PNS) berjumlah 16 penghuni atau sebesar 13%. Meskipun jumlahnya tidak sebanyak kelompok lainnya, keberadaan penghuni yang berstatus PNS tetap memberikan kontribusi terhadap keberagaman karakteristik penghuni di kawasan tersebut. Berdasarkan hasil distribusi pekerjaan dapat disimpulkan bahwa mayoritas penghuni Citraland Gama City berasal dari kalangan pekerja sektor swasta dan pelaku usaha. Kondisi ini menunjukkan bahwa kawasan perumahan tersebut banyak diminati oleh masyarakat yang memiliki aktivitas ekonomi produktif dan tingkat mobilitas yang cukup tinggi.

Distribusi Tingkat Pendapatan Penghuni

Tingkat pendapatan merupakan salah satu indikator yang dapat digunakan untuk mengetahui kondisi ekonomi penghuni. Variabel ini sangat penting karena memiliki pengaruh terhadap kemampuan penghuni dalam memilih tipe hunian, memenuhi kebutuhan keluarga, serta menentukan pola kehidupan sehari-hari.

Tabel 4. Distribusi Tingkat Pendapatan Penghuni

No	Tingkat Pendapatan	Jumlah	Persentase
1	< Rp5.000.000	18	15%
2	Rp5.000.000 – Rp10.000.000	39	33%
3	Rp10.000.001 – Rp15.000.000	37	31%
4	> Rp15.000.000	26	21%
Total	120	100%	

Berdasarkan Tabel di atas dapat diketahui bahwa kelompok pendapatan yang paling dominan adalah penghuni dengan pendapatan antara Rp5.000.000 sampai Rp10.000.000 per bulan yaitu sebanyak 39 penghuni atau sebesar 33%. Kelompok dengan pendapatan antara Rp10.000.001 sampai Rp15.000.000 berada pada posisi kedua dengan jumlah 37 penghuni atau sebesar 31%. Hal ini menunjukkan bahwa sebagian besar penghuni berada pada kategori ekonomi menengah hingga menengah atas. Selanjutnya penghuni dengan pendapatan di atas Rp15.000.000 berjumlah 26 orang atau sebesar 21%. Kelompok ini menunjukkan adanya penghuni dengan kondisi ekonomi yang relatif tinggi sehingga mampu menempati hunian dengan tipe yang lebih besar dan fasilitas yang lebih lengkap. Sementara itu, penghuni dengan pendapatan kurang dari Rp5.000.000 berjumlah 18 orang atau sebesar 15%. Jumlah tersebut merupakan kelompok terkecil dalam penelitian ini.

Hasil analisis menunjukkan bahwa sebagian besar penghuni Citraland Gama City memiliki tingkat pendapatan yang relatif baik. Kondisi ini menunjukkan bahwa kawasan tersebut termasuk kawasan hunian yang diminati oleh masyarakat dengan kemampuan ekonomi menengah hingga tinggi.

Distribusi Jumlah Anggota Keluarga

Jumlah anggota keluarga menjadi salah satu faktor yang mempengaruhi kebutuhan hunian. Semakin banyak anggota keluarga maka semakin besar pula kebutuhan ruang dan fasilitas yang diperlukan dalam sebuah rumah.

Tabel 5. Distribusi Jumlah Anggota Keluarga

No	Jumlah Anggota Keluarga	Jumlah	Persentase
1	1-2 Orang	18	15%
2	3-4 Orang	54	45%
3	5-6 Orang	35	29%
4	>6 Orang	13	11%
Total	120	100%	

Berdasarkan Tabel di atas dapat diketahui bahwa jumlah anggota keluarga yang paling dominan adalah kelompok 3-4 orang dengan jumlah 54 penghuni atau sebesar 45%. Kelompok tersebut umumnya terdiri dari keluarga inti yang terdiri atas ayah, ibu, dan satu hingga dua anak. Kondisi ini menunjukkan bahwa sebagian besar penghuni berada pada fase keluarga berkembang. Kelompok dengan jumlah anggota keluarga 5-6 orang berjumlah 35 penghuni atau sebesar 29%. Kelompok ini menunjukkan adanya keluarga yang memiliki jumlah anak lebih banyak atau tinggal bersama anggota keluarga lainnya. Sementara itu, kelompok 1-2 orang berjumlah 18 penghuni atau sebesar 15%, sedangkan kelompok dengan anggota keluarga lebih dari 6 orang berjumlah 13 penghuni atau sebesar 11%.

Distribusi Tipe Hunian

Tipe hunian merupakan salah satu indikator yang dapat mencerminkan kondisi ekonomi dan kebutuhan ruang penghuni. Semakin besar tipe rumah yang dimiliki maka umumnya semakin tinggi pula kemampuan ekonomi penghuni tersebut.

Tabel 6. Distribusi Tipe Hunian

No	Tipe Hunian	Jumlah	Persentase
1	Tipe Komersil	22	18%
2	Tipe Villa	41	34%
3	Tipe Rumah Mewah	35	29%
4	Tipe Town House	22	18%
Total	120	100%	

Berdasarkan Tabel di atas dapat diketahui bahwa tipe hunian yang paling banyak ditempati adalah Tipe Villa dengan jumlah 41 penghuni atau sebesar 34%. Tipe Rumah Mewah menempati posisi kedua dengan jumlah 35 penghuni atau sebesar 29%. Tingginya jumlah penghuni pada tipe hunian menengah menunjukkan bahwa sebagian besar penghuni memiliki kebutuhan ruang yang cukup besar untuk mendukung aktivitas keluarga. Sementara itu, Tipe Komersil dan Tipe Town House masing-masing berjumlah 22 penghuni atau sebesar 18%. Berdasarkan hasil tersebut dapat disimpulkan bahwa mayoritas penghuni memilih tipe hunian menengah karena

dianggap mampu memenuhi kebutuhan keluarga sekaligus sesuai dengan kemampuan ekonomi yang dimiliki.

Preprocessing Data

Sebelum dilakukan proses klasifikasi menggunakan metode K-Nearest Neighbor, seluruh data terlebih dahulu melalui tahapan preprocessing. Tahapan ini sangat penting karena kualitas data akan sangat mempengaruhi hasil klasifikasi yang dihasilkan. Proses preprocessing dilakukan untuk memastikan bahwa data yang digunakan dalam penelitian berada dalam kondisi yang baik, lengkap, dan siap untuk diproses menggunakan algoritma KNN. Tahapan preprocessing yang dilakukan meliputi Data Cleaning, Transformasi Data, dan Normalisasi Data. Ketiga tahapan tersebut dilakukan secara berurutan agar proses klasifikasi dapat berjalan dengan baik dan menghasilkan tingkat akurasi yang optimal.

Data Cleaning

Data cleaning merupakan proses pembersihan data dari berbagai kesalahan yang mungkin terjadi selama proses pengumpulan data. Kesalahan tersebut dapat berupa data kosong, data ganda, kesalahan penulisan, maupun data yang tidak sesuai dengan kebutuhan penelitian. Pada penelitian ini dilakukan pemeriksaan terhadap seluruh 120 data penghuni yang telah dikumpulkan. Hasil pemeriksaan menunjukkan bahwa sebagian kecil data memiliki ketidakkonsistenan dalam penulisan pekerjaan dan pendapatan sehingga dilakukan penyesuaian agar seluruh data memiliki format yang seragam.

Tabel 7. Hasil Data Cleaning

Keterangan	Jumlah Data
Data Awal	120
Data Tidak Konsisten	8
Data Diperbaiki	8
Data Akhir Digunakan	120

Berdasarkan Tabel di atas dapat diketahui bahwa terdapat 8 data yang mengalami ketidakkonsistenan. Seluruh data tersebut berhasil diperbaiki sehingga jumlah data yang digunakan dalam penelitian tetap sebanyak 120 data. Hasil data cleaning menunjukkan bahwa seluruh data yang digunakan telah memenuhi kriteria penelitian dan siap untuk diproses pada tahap berikutnya.

Transformasi Data

Setelah proses data cleaning selesai dilakukan, tahap berikutnya adalah transformasi data. Transformasi data dilakukan karena tidak seluruh variabel yang digunakan dalam penelitian berbentuk numerik. Beberapa variabel seperti pekerjaan dan tipe hunian masih berbentuk data kategorikal sehingga perlu diubah ke dalam bentuk angka agar dapat diproses menggunakan algoritma K-Nearest Neighbor (KNN). Proses transformasi data merupakan salah satu tahapan penting dalam data mining karena algoritma KNN menggunakan perhitungan matematis berupa jarak antar data. Oleh karena itu, seluruh atribut harus berada dalam bentuk numerik agar proses perhitungan dapat dilakukan secara optimal.

Transformasi Variabel Pekerjaan

Variabel pekerjaan yang semula berbentuk kategori diubah menjadi nilai numerik seperti yang ditunjukkan pada tabel berikut.

Tabel 8. Transformasi Variabel Pekerjaan

No	Pekerjaan	Nilai
1	Pegawai Swasta	1
2	Wiraswasta	2
3	Profesional	3
4	PNS	4

Berdasarkan Tabel di atas dapat diketahui bahwa setiap kategori pekerjaan diberikan kode numerik yang berbeda. Pemberian kode tersebut bertujuan untuk memudahkan proses pengolahan data menggunakan algoritma KNN tanpa mengubah makna dari kategori pekerjaan itu sendiri.

Transformasi Variabel Tipe Hunian

Selain pekerjaan, variabel tipe hunian juga perlu dilakukan transformasi agar dapat diproses secara matematis.

Tabel 9. Transformasi Variabel Tipe Hunian

No	Tipe Hunian	Nilai
1	Tipe Komersil	1
2	Tipe Villa	2
3	Tipe Rumah Mewah	3
4	Tipe Town House	4

Berdasarkan Tabel di atas dapat diketahui bahwa tipe hunian diberikan nilai berdasarkan kategori yang dimiliki. Semakin besar tipe hunian maka nilai numerik yang diberikan juga semakin tinggi.

Transformasi Variabel Karakteristik Penghuni

Variabel karakteristik penghuni digunakan sebagai kelas target dalam proses klasifikasi. Oleh karena itu, variabel ini juga perlu ditransformasikan ke dalam bentuk numerik.

Tabel 10. Transformasi Karakteristik Penghuni

No	Karakteristik Penghuni	Nilai
1	Keluarga Muda	1
2	Keluarga Berkembang	2
3	Keluarga Mapan	3

Transformasi terhadap kelas target dilakukan untuk mempermudah proses pengujian dan evaluasi hasil klasifikasi yang dilakukan oleh algoritma KNN.

Normalisasi Data

Setelah seluruh data berhasil ditransformasikan ke dalam bentuk numerik, langkah selanjutnya adalah melakukan normalisasi data. Normalisasi merupakan proses penyamaan skala data agar setiap variabel memiliki rentang nilai yang seimbang. Tahapan ini sangat penting karena metode KNN menggunakan perhitungan jarak antar data. Apabila terdapat perbedaan skala yang terlalu besar antar variabel, maka variabel dengan nilai terbesar akan mendominasi proses perhitungan dan menyebabkan hasil klasifikasi menjadi kurang optimal. Sebagai contoh, variabel usia

memiliki rentang nilai antara 20 sampai 60 tahun, sedangkan variabel jumlah anggota keluarga hanya memiliki rentang nilai antara 1 sampai 8 orang. Jika tidak dilakukan normalisasi maka variabel usia akan memiliki pengaruh yang lebih besar dibandingkan variabel lainnya. Dalam penelitian ini digunakan metode Min-Max Normalization dengan rumus:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan:

X' = Nilai hasil normalisasi

X = Nilai data asli

X_{min} = Nilai minimum dari suatu atribut

X_{max} = Nilai maksimum dari suatu atribut

Melalui proses normalisasi, seluruh variabel akan berada pada rentang nilai 0 sampai 1 sehingga memiliki kontribusi yang seimbang dalam proses perhitungan jarak.

Tabel 11. Contoh Hasil Normalisasi Data

Usia	Pekerjaan	Pendapatan	Anggota Keluarga	Tipe Hunian
0,45	0,25	0,52	0,50	0,33
0,52	0,50	0,61	0,50	0,67
0,61	0,75	0,70	0,75	1,00
0,38	0,25	0,44	0,25	0,33
0,70	1,00	0,82	0,75	1,00

Berdasarkan Tabel di atas dapat diketahui bahwa seluruh data telah berada pada rentang nilai antara 0 hingga 1. Hasil tersebut menunjukkan bahwa proses normalisasi berhasil dilakukan dan data siap digunakan dalam proses klasifikasi.

Pembagian Data Training dan Data Testing

Setelah proses preprocessing selesai dilakukan, data kemudian dibagi menjadi dua kelompok yaitu data training dan data testing. Data training digunakan untuk membentuk pola klasifikasi yang akan digunakan oleh algoritma KNN, sedangkan data testing digunakan untuk menguji kemampuan model dalam memprediksi data baru yang belum pernah diproses sebelumnya. Dalam penelitian ini digunakan perbandingan pembagian data sebesar 80:20.

Tabel 12. Pembagian Data Penelitian

Jenis Data	Jumlah Data	Persentase
Data Training	96	80%
Data Testing	24	20%
Total	120	100%

Berdasarkan Tabel di atas dapat diketahui bahwa sebanyak 96 data digunakan sebagai data training dan 24 data digunakan sebagai data testing. Penggunaan komposisi 80% data training dan 20% data testing dipilih karena mampu memberikan keseimbangan antara proses pembelajaran model dan proses pengujian model. Semakin banyak data training yang digunakan maka pola klasifikasi yang terbentuk akan semakin baik, sedangkan data testing tetap diperlukan untuk mengukur performa model secara objektif.

Implementasi Metode K-Nearest Neighbor (KNN)

Tahap berikutnya adalah implementasi algoritma K-Nearest Neighbor untuk melakukan prediksi karakteristik penghuni Citraland Gama City. Metode KNN bekerja dengan cara mencari sejumlah tetangga terdekat dari data yang akan diprediksi. Kedekatan antar data dihitung menggunakan rumus Euclidean Distance. Data yang memiliki jarak paling kecil dianggap memiliki karakteristik yang paling mirip. Dalam penelitian ini digunakan nilai $K = 5$ karena berdasarkan beberapa kali pengujian awal nilai tersebut menghasilkan performa klasifikasi yang lebih baik dibandingkan nilai K lainnya.

Tabel 13. Contoh Data Testing

Variabel	Nilai
Usia	38 Tahun
Pekerjaan	Wiraswasta
Pendapatan	Rp12.000.000
Jumlah Anggota Keluarga	4 Orang
Tipe Hunian	80
Karakteristik	Belum Diketahui

Data pada Tabel di atas merupakan salah satu data testing yang akan diklasifikasikan menggunakan algoritma KNN.

Perhitungan Manual K-Nearest Neighbor

Untuk mengetahui proses kerja algoritma KNN secara lebih jelas, dilakukan perhitungan manual terhadap salah satu data testing. Rumus Euclidean Distance yang digunakan adalah:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan:

- $d(x, y)$ = Jarak antara data testing dan data training
 x_i = Nilai atribut ke- i pada data testing
 y_i = Nilai atribut ke- i pada data training
 n = Jumlah atribut yang digunakan
 Σ = Penjumlahan seluruh atribut
 $\sqrt{\quad}$ = Akar kuadrat

Tabel 14. Contoh Data Training Terdekat

Data	Jarak	Karakteristik
D1	0,152	Keluarga Berkembang
D2	0,198	Keluarga Berkembang
D3	0,214	Keluarga Berkembang
D4	0,287	Keluarga Mapan
D5	0,315	Keluarga Berkembang

Berdasarkan Tabel di atas dapat diketahui bahwa terdapat lima tetangga terdekat yang digunakan dalam proses klasifikasi. Dari lima data tersebut, empat data termasuk kategori Keluarga Berkembang dan satu data termasuk kategori Keluarga Mapan. Karena mayoritas tetangga terdekat berasal dari kategori Keluarga Berkembang, maka hasil prediksi untuk data testing tersebut adalah Keluarga Berkembang. Proses voting mayoritas inilah yang menjadi dasar pengambilan keputusan dalam algoritma K-Nearest Neighbor.

Hasil Klasifikasi Data Testing

Setelah seluruh data testing diproses menggunakan metode KNN, diperoleh hasil klasifikasi sebagai berikut.

Tabel 15. Hasil Prediksi Karakteristik Penghuni

Karakteristik	Jumlah Prediksi
Keluarga Muda	7
Keluarga Berkembang	10
Keluarga Mapan	7
Total	24

Berdasarkan Tabel di atas dapat diketahui bahwa kategori yang paling banyak diprediksi oleh sistem adalah Keluarga Berkembang dengan jumlah 10 data. Hasil tersebut sejalan dengan distribusi data awal yang menunjukkan bahwa sebagian besar penghuni Citraland Gama City berada pada kategori keluarga berkembang. Hal ini menunjukkan bahwa pola yang ditemukan oleh algoritma KNN sesuai dengan kondisi aktual data penelitian. Selain itu, jumlah prediksi untuk kategori Keluarga Muda dan Keluarga Mapan masing-masing sebanyak 7 data. Distribusi tersebut menunjukkan bahwa model mampu mengidentifikasi seluruh kategori karakteristik penghuni yang digunakan dalam penelitian.

Evaluasi Hasil Klasifikasi Menggunakan Confusion Matrix

Setelah proses klasifikasi dilakukan menggunakan metode K-Nearest Neighbor (KNN), tahap berikutnya adalah melakukan evaluasi terhadap hasil prediksi yang diperoleh. Evaluasi ini bertujuan untuk mengetahui tingkat kemampuan algoritma dalam mengklasifikasikan karakteristik penghuni Citraland Gama City berdasarkan data yang digunakan. Dalam penelitian ini evaluasi dilakukan menggunakan Confusion Matrix. Confusion Matrix merupakan salah satu metode evaluasi yang digunakan untuk membandingkan antara data aktual dengan hasil prediksi yang dihasilkan oleh sistem. Melalui Confusion Matrix dapat diketahui jumlah data yang berhasil diklasifikasikan dengan benar maupun data yang mengalami kesalahan klasifikasi. Penggunaan Confusion Matrix dalam penelitian ini sangat penting karena dapat memberikan gambaran mengenai performa algoritma KNN secara lebih rinci. Selain itu, metode ini juga dapat digunakan sebagai dasar dalam menghitung nilai akurasi klasifikasi yang dihasilkan.

Tabel 16. Confusion Matrix Hasil Klasifikasi

Data Aktual	Keluarga Muda	Keluarga Berkembang	Keluarga Mapan	Total
Keluarga Muda	6	1	1	8
Keluarga Berkembang	1	8	1	10
Keluarga Mapan	0	1	5	6
Total	7	10	7	24

Berdasarkan Tabel di atas dapat diketahui bahwa terdapat beberapa data yang berhasil diprediksi dengan benar dan beberapa data yang mengalami kesalahan klasifikasi. Pada kategori Keluarga Muda terdapat 8 data aktual. Dari jumlah tersebut sebanyak 6 data berhasil diprediksi dengan benar sebagai Keluarga Muda, sedangkan 1 data diprediksi sebagai Keluarga Berkembang dan 1 data diprediksi sebagai Keluarga Mapan. Pada kategori Keluarga Berkembang terdapat 10 data aktual. Dari

jumlah tersebut sebanyak 8 data berhasil diklasifikasikan dengan benar sebagai Keluarga Berkembang. Sementara itu terdapat masing-masing 1 data yang diprediksi sebagai Keluarga Muda dan Keluarga Mapan. Pada kategori Keluarga Mapan terdapat 6 data aktual. Sebanyak 5 data berhasil diprediksi dengan benar sebagai Keluarga Mapan, sedangkan 1 data diprediksi sebagai Keluarga Berkembang. Berdasarkan hasil tersebut dapat diketahui bahwa sebagian besar data berhasil diklasifikasikan dengan benar oleh algoritma KNN. Kesalahan klasifikasi yang terjadi relatif kecil dan umumnya terjadi pada data yang memiliki karakteristik yang hampir sama antar kategori.

Pengujian Tingkat Akurasi Metode K-Nearest Neighbor

Setelah diperoleh hasil Confusion Matrix, langkah berikutnya adalah menghitung tingkat akurasi klasifikasi yang dihasilkan oleh metode K-Nearest Neighbor. Akurasi digunakan untuk mengetahui seberapa besar kemampuan model dalam memprediksi data secara tepat dibandingkan dengan kondisi sebenarnya. Semakin tinggi nilai akurasi yang diperoleh maka semakin baik performa model dalam melakukan klasifikasi. Perhitungan akurasi dilakukan menggunakan rumus sebagai berikut:

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Seluruh\ Data\ Testing} \times 100\%$$

Berdasarkan Confusion Matrix yang telah diperoleh, jumlah prediksi yang benar adalah:

Keluarga Muda = 6 data

Keluarga Berkembang = 8 data

Keluarga Mapan = 5 data

Total prediksi benar: 6+8+5=19

Jumlah data testing yang digunakan: 24 Data

Sehingga perhitungan akurasi menjadi: Akurasi = 19 / 24 x 100% = 79,17%

Tabel 17. Hasil Pengujian Akurasi

Keterangan	Jumlah
Data Testing	24
Prediksi Benar	19
Prediksi Salah	5
Akurasi	79%

1. Berdasarkan Tabel diatas dapat diketahui bahwa metode K-Nearest Neighbor menghasilkan tingkat akurasi sebesar 79%. Nilai tersebut menunjukkan bahwa dari 24 data testing yang digunakan, sebanyak 19 data berhasil diprediksi dengan benar sesuai dengan karakteristik aktual yang dimiliki. Sedangkan sebanyak 5 data mengalami kesalahan klasifikasi. Hasil ini menunjukkan bahwa metode KNN memiliki kemampuan yang cukup baik dalam melakukan prediksi karakteristik penghuni berdasarkan atribut usia, pekerjaan, tingkat pendapatan, jumlah anggota keluarga, dan tipe hunian. Tingkat akurasi sebesar 79% juga menunjukkan bahwa pola hubungan antar variabel yang digunakan dalam penelitian mampu dikenali dengan baik oleh algoritma KNN sehingga menghasilkan klasifikasi yang cukup akurat.

Pembahasan

Berdasarkan hasil pengolahan data, penghuni Citraland Gama City memiliki karakteristik yang beragam. Mayoritas penghuni berada pada rentang usia produktif 31–50 tahun, dengan pekerjaan dominan sebagai pegawai swasta dan wiraswasta. Sebagian besar penghuni memiliki tingkat pendapatan menengah hingga menengah atas, yaitu antara Rp5.000.000 sampai Rp15.000.000 per bulan. Jumlah anggota keluarga yang paling banyak ditemukan adalah 3–4 orang, sedangkan tipe hunian yang dominan adalah Tipe Villa dan Tipe Rumah Mewah. Kondisi ini menunjukkan bahwa mayoritas penghuni merupakan keluarga dengan aktivitas ekonomi yang aktif dan kebutuhan hunian yang relatif tinggi.

Metode K-Nearest Neighbor (KNN) diterapkan melalui tahapan data cleaning, transformasi data, normalisasi, pembagian data training dan testing, perhitungan jarak menggunakan Euclidean Distance, serta penentuan nilai K. Hasil implementasi menunjukkan bahwa metode KNN mampu mengelompokkan penghuni berdasarkan kemiripan karakteristik yang dimiliki. Metode ini relatif sederhana, mudah diterapkan, dan mampu memberikan hasil klasifikasi yang cukup baik pada data penghuni perumahan.

Hasil pengujian menunjukkan bahwa metode KNN menghasilkan tingkat akurasi sebesar 79%. Nilai tersebut menunjukkan bahwa sebagian besar data penghuni berhasil diklasifikasikan dengan benar. Meskipun masih terdapat beberapa kesalahan klasifikasi akibat kemiripan karakteristik antar kelompok, hasil ini menunjukkan bahwa variabel usia, pekerjaan, pendapatan, jumlah anggota keluarga, dan tipe hunian memiliki kontribusi yang baik dalam proses prediksi karakteristik penghuni.

Hasil prediksi karakteristik penghuni memberikan informasi yang bermanfaat bagi pengelola Citraland Gama City dalam memahami profil penghuni secara lebih akurat. Informasi tersebut dapat digunakan sebagai dasar penyusunan kebijakan, peningkatan pelayanan, serta perencanaan fasilitas yang sesuai dengan kebutuhan penghuni. Selain itu, penerapan metode KNN membantu mempercepat proses analisis data dan mengurangi risiko kesalahan yang sering terjadi pada pengolahan data secara manual, sehingga pengambilan keputusan dapat dilakukan dengan lebih efektif dan tepat sasaran.

4. Simpulan

Berdasarkan hasil penelitian, implementasi metode K-Nearest Neighbor (KNN) untuk prediksi karakteristik penghuni Citraland Gama City Kota Medan terbukti mampu membantu proses klasifikasi data penghuni secara lebih cepat dan sistematis. Metode KNN dapat mengelompokkan karakteristik penghuni berdasarkan variabel usia, pekerjaan, tingkat pendapatan, jumlah anggota keluarga, dan tipe hunian. Hasil pengujian menunjukkan bahwa metode KNN menghasilkan tingkat akurasi sebesar 79%, sehingga metode ini cukup baik digunakan untuk memprediksi karakteristik penghuni. Selain itu, hasil klasifikasi yang diperoleh dapat membantu pihak pengelola dalam memahami karakteristik penghuni dan mendukung proses pengambilan keputusan yang lebih tepat berdasarkan data yang tersedia.

5. Daftar Pustaka

Abdullah, M. (2022). *Metodologi penelitian kuantitatif*. Aswaja Pressindo.
Hardani, dkk. (2023). *Metode penelitian kualitatif dan kuantitatif*. Pustaka Ilmu.
Kusrini, & Luthfi, E. T. (2022). *Algoritma data mining*. Andi.

- Limbong, T., Simarmata, J., & Iskandar, A. (2022). *Data mining: Teori dan implementasi*. Yayasan Kita Menulis.
- Nofriansyah, D., & Nurcahyo, G. W. (2022). *Konsep data mining dan penerapannya*. Deepublish.
- Prasetyo, E. (2022). *Data mining: Konsep dan aplikasi menggunakan MATLAB*. Andi.
- Santoso, B. (2023). *Machine learning dan data mining untuk analisis data*. Elex Media Komputindo.
- Saputra, D., & Nofriansyah, D. (2023). *Machine learning untuk pemula*. Deepublish.
- Sinambela, L. P. (2023). *Metodologi penelitian kuantitatif*. Mitra Wacana Media.
- Siregar, S. (2022). *Metode penelitian kuantitatif: Dilengkapi dengan perbandingan perhitungan manual dan SPSS*. Kencana.
- Siregar, V. M. M. (2022). *Kecerdasan buatan dan data mining*. Yayasan Kita Menulis.
- Sugiyono. (2022). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.
- Suyanto. (2022). *Data mining untuk klasifikasi dan klasterisasi data*. Informatika.
- Wanto, A., Windarto, A. P., Hartama, D., & Parlina, I. (2022). *Data mining: Algoritma dan implementasi*. Yayasan Kita Menulis.
- Windarto, A. P., Wanto, A., & Hartama, D. (2023). *Penerapan data mining pada berbagai bidang*. Yayasan Kita Menulis.