Implementasi Algoritma Reduct dan Decision Tree untuk Mengenali Pola Klasifikasi Pelanggan pada Sistem E-Commerce

1st Hans Alessandro Manurung
Sistem Informasi
Telkom University
Bandung, Indonesia
hansalessandrom@student.telkomuniversity.ac.id

2nd Rohmat Saedudin
Sistem Informasi
Telkom University
Bandung, Indonesia
rohmatsaedudin@telkomuniversity.ac.id

3rd Muhammad Fathinuddin
Sistem Informasi
Telkom University
Bandung, Indonesia
muhammadfathinuddin@telkomuniversity.ac.id

Pertumbuhan e-commerce membuat volume dan kompleksitas data transaksi pelanggan semakin tinggi. Hal ini bisa menurunkan kinerja dan interpretasi model klasifikasi jika terlalu banyak atribut digunakan. Penelitian ini bertujuan menerapkan algoritma Reduct berbasis Rough Set Theory sebagai metode seleksi atribut sebelum klasifikasi pelanggan dengan Decision Tree. Penelitian dilakukan secara kuantitatif menggunakan data transaksi pelanggan e-commerce dari Kaggle, yang terdiri dari 48 transaksi dan empat kategori produk. Proses penelitian ini dimulai dari pra-pemrosesan data, diskritisasi atribut numerik, Reduct untuk seleksi atribut guna memperoleh minimal reduct, konstruksi model Decision Tree, dan evaluasi model menggunakan metrik akurasi, presisi, recall, serta F1 - score. Hasilnya, algoritma Reduct berhasil mereduksi tujuh atribut menjadi dua atribut utama, yaitu Total Transaksi (Rp) dan Kategori Produk, tanpa mengurangi kemampuan klasifikasi. Model Decision Tree dengan atribut hasil reduksi menghasilkan struktur pohon yang lebih sederhana, mudah dipahami, dan kinerjanya lebih baik dibandingkan model tanpa reduksi atribut. Temuan menunjukkan bahwa integrasi Reduct dan Decision Tree berhasil menyederhanakan data sambil meningkatkan kualitas klasifikasi pelanggan e-commerce. Kata kunci: Reduct, Teori Himpunan Kasar, Decision Tree, Klasifikasi Pelanggan, E- Commerce

Kata kunci— ***Reduct, Rough Set Theory, Decision Tree,*** **Klasifikasi Pelanggan,** ***E-Commerce***

I. PENDAHULUAN

Volume dan kompleksitas data transaksi pelanggan meningkat secara signifikan seba ga i akibat dari perkembangan sistem e-commerce yang semakin pesat. N data besar dapat dikumpulkan dari semua tindakan pelanggan, seperti nilai transaksi, frekuensi pembelian, dan pemilihan kategori produk. Data ini dapat digunakan untuk memahami pola perilaku pelanggan dan membantu pengambilan keputusan bisnis berbasis data (Dewi et al., 2024). Di tengah persaingan pasar digital yang semakin ketat, salah satu komponen penting dalam meningkatkan daya saing bisnis e-commerce adalah pengoptimalan data transaksi pelanggan [1]. Namun, lebih banyak fitur dalam data transaksi pelanggan membuat analisis data lebih sulit, terutama pada tahap klasifikasi pelanggan. Kondisi di mana lebih banyak atribut digunakan berdampak pada efisiensi komputasi dan kinerja model klasifikasi dapat menyebabkan masalah curse of dimensionality. Selain meningkatkan kompleksitas pemrosesan data, atribut yang berlebihan atau tidak relevan dapat menyebabkan overfitting dan interpretasi yang sulit dari hasil klasifikasi (Diwandari & Hidayat, 2021). Oleh karena itu, penentuan fitur yang tepat ada la h langkah penting da la m proses klasifikasi pelanggan e- commerce [2]. Decision Tree ada la h sa la h satu dari banyak algoritma klasifikasi yang digunakan da la m penelitian yang berkaitan dengan kla sifikasi pelanggan e-commerce. Algoritma ini banyak digunakan karena kemampuan untuk menghasilkan model klasifikasi dalam bentuk aturan keputusan yang mudah dipahami dan ditafsirkan, yang sesuai untuk membantu analisis dan pengambilan keputusan bisnis (Diwandari & Hidayat, 2021). Decision Tree kompetitif da la m mengkla sifikasikan pela nggan e-commerce, menurut penelitian Dewi et al. (2024). Meskipun demikian, kinerja Decision Tree sangat bergantung pada kualitas dan relevansi atribut yang digunakan seba ga i input, sehingga penggunaan atribut yang tidak tepat dapat menurunkan efektivitas model klasifikasi [3]. Metode seleksi atribut yang mampu mempertahankan informasi penting da la m data tanpa meningkatkan kompleksitas model diperlukan untuk mengatasi masa la h atribut yang berlebihan. Teori Rough Set ada la h metode seleksi atribut yang mampu menentukan hubungan ketergantungan antara atribut kondisi dan atribut keputusan secara objektif. Metode ini tidak memerlukan asumsi probabilistik tambahan. Rough Set Theory dapat menghasilkan minimal reduct, yaitu himpunan atribut minimum yang tetap mempertahankan kemampuan data untuk membedakan kela s keputusan (Pawlak, 1991). Ini dapat dicapai melalui gagasan reduksi. Metode ini dinilai cocok untuk data transaksi e-commerce pelanggan yang memiliki atribut diskrit dan heterogen [4]. Diharapkan bahwa integrasi algoritma Reduct dan Decision Tree akan memungkinkan pembuatan model klasifikasi pelanggan e- commerce yang lebih efisien dan mudah dipahami. Model klasifikasi mungkin lebih sederhana dan kinerjanya lebih baik karena jumla h atribut yang relevan yang lebih sedikit. Metode ini sejalan denga n paradigma penelitian kuantitatif, yang menekankan pengujian empiris terukur dan sistematis hubungan antarvariabel (Creswell & Creswell, 2018)[5].

II. KAJIAN TEORI

II .1 Data Mining

Data mining adalah proses menemukan pola, hubungan, dan pengetahuan penting dari kumpulan data

yang sangat besar dengan menggunakan teknik komputasi dan statistik. Data mining adalah dasar penelitian ini untuk mengolah data transaksi e-commerce pelanggan karena sangat penting dalam sistem e-commerce untuk menganalisis data transaksi pelanggan dan menemukan pola perilaku pelanggan yang dapat membantu pengambilan keputusan bisnis (Diwandari & Hidayat, 2021). [6].

II.2 Decision Tree

Algoritma klasifikasi yang dikenal seba ga i "pohon keputusan" menciptakan model dalam bentuk pohon dengan setiap node, cabang, dan daun mewa kili kelas keputusan. Algoritma ini banyak digunakan karena dapat membuat aturan keputusan yang mudah dipahami dan ditafsirkan. Decision Tree digunakan da la m sistem e-commerce untuk mengkategorikan pela nggan berdasarkan atribut transaksi. Ini memungkinkan analisis perilaku pelanggan yangjelas dan sistematis (Diwandari & Hidayat, 2021; Dewi et al., 2024).

II.3 Rough Set Theory (Reduct)

Untuk menangani ketidakpastia n dan ketidaktepatan data, Rough Set Theory ada la h metode matematis yang tidak memerlukan asumsi tambahan seperti probabilitas a tau fungsi keanggotaan. Relasi ketidakbedaan, yaitu relasi yang mengelompokkan objek yang tidak dapat dibedakan berdasarkan karakteristik tertentu, adalah konsep utama dari set rumit. Metode in i memungkinkan untuk menganalisis ketergantungan antara ciri-ciri keputusan dan ciri-ciri kondisi secara objektif (Pawlak, 1991). Da la m teori Rough Set, reduksi ada la h ide utama untuk metode seleksi atribut. Tujuan reduksi adalah untuk mempertahankan kemampuan data untuk membedakan kelas keputusan dengan mengumpulkan atribut minimal. Dengan menggunakan reduct, atribut yang tidak relevan atau redundant dapat dieliminasi tanpa mengurangi kualitas informasi yang diperlukan untuk proses klasifikasi.. Sebelum memulai proses klasifikasi, penelitian in i menggunakan reduct untuk menyederhanakan fitur data pelanggan e-commerce (Pawlak, 1991).

III. METODE

Dengan mengintegrasikan algoritma Reduct dan Decision Tree, sistematika penyelesa ia n masalah dalam penelitian ini disusun secara terstuktur untuk mencapai tujuan klasifikasi pelanggan e-commerce.

A. Sistematika Penyelesaian Masalah

1. Identifikasi Masalah

   Penelitian dimulai dengan menemukan masalah pada sistem e-commerce: data transaksi pelanggan yang sangat kompleks karena banyaknya atribut yang digunakan. Kondisi ini berpotensi mengganggu pemrosesan dan kinerja model klasifikasi pelanggan, sehingga diperlukan metode untuk memilih atribut yang tepat.

2. Pengumpulan Data

   Data yang digunakan dalam penelitian ini berasal dari dataset transaksi e-commerce pelanggan yang dikumpulkan melalui platform Ka ggle; dataset ini serupa dengan dataset yang digunakan pada Tugas Akhir dan terdiri dari data transaksi pelanggan yang memiliki beberapa atribut kondisi serta satu atribut keputusan, yaitu kategoripelanggan.

3. Pra-pemrosesan Data

   Untuk memastikan kualitas data sebelum dianalisis, tahap pra-pemrosesan dilakukan. Tahap ini mencakup membersihkan data dan menyesuaikan formatnya a ga r sesuai dengan metode yang digunakan dalam penelitian.

4. Diskritasi Data

   Dengan mengelompokkan nilai ke dalam interval tertentu, atribut numerik pada data transaksi pelanggan dilakukan melalui proses diskritisasi. Tahap ini diperlukan karena teori Rough Set bekerja denga n baik pada data diskrit.

5. Pembentukan Tabel Keputusa n

   Data hasil pra-pemrosesan dan diskritisasi disusun da la m bentuk tabelkeputusan yang terdiri dari atribut kondisi dan atribut keputusan. Analisis hubungan ketergantungan antar atribut didasarkan pada tabel keputusan ini.

6. Seleksi Atribut Menggunakan Reduct

   Tujuan dari tahap ini ada la h untuk menghilangkan atribut yang tidak diperlukan tanpa mengurangi kemampuan data untuk membedakan kelas keputusan. Untuk mencapai tujuan ini, algoritma reduksi yang berbasis pada teori set keras digunakan untuk mela kukan seleksi atribut.

7. Pembangunan Model Klasifikasi

   Model klasifikasi pelanggan dibangun dengan algoritma Decision Tree yang berpusat pada atribut hasil reduksi. Selain itu, model klasifikasi juga dibangun dengan menggunakan atribut penuh tanpa reduksi sebagai pembanding.

8. Analisis dan Evaluasi Hasil

   Analisis hasil klasifikasi dilakukan untuk mengevaluasi kinerja model dan membandingkan hasil klasifikasi sebelum dan sesudah penerapan Reduct. Tujuan analisis ini adalah untuk mengetahui seberapa efektif penggunaan atribut terhadap kinerja dan kompleksita s model klasifikasi pelanggan e-commerce.

B. Metode Evaluasi

Metode Evaluasi digunakan untuk mengevaluasi kinerja model klasifikasi pelanggan e-commerce yang dibangun menggunakan algoritma Decision Tree, baik sebelum maupun setelah penerapan seleksi atribut menggunakan Reduct. Evaluasi dilakukan denga n menggunakan Confusion Matrix seba ga i dasar perhitungan metrik evaluasi yang digunakan dalam penelitian ini.

1) Confusion Matrix

Untuk menunjukkan perbedaan antara kela s hasil prediksi dan kelas aktual dalam model klasifikasi, confusion matrix digunakan. Statistik ini menunjukkan jumla h data yang diklasifikasikan dengan benar dan salah untuk setiap kela s pelanggan. Perhitungan nilai ketepatan, ketepatan, recall, dan skor F1 dilakukan berdasarkan data confusion matrix.

2) Accuracy

Accuracy adalah ukuran tingkat ketepatan model dalam mengklasifikasikan seluruh data pela nggan secara keseluruhan. Nilai ketepatan ditentukan dengan membandingkan jumlah data yang diklasifikasikan dengan benar dengan jumla h total data yang diuji. Semakin tinggi nilai ketepatan, semakin baik kemampuan model untuk mengklasifikasikan data e-commerce pelanggan.

Rumus Accuracy : *Accuracy = Jumlah Klasifikasi Benar / Jumlah Seluruh Data*

3) Precision

Nilai ketepatan, yang diperoleh dari perbandingan antara jumlah data yang diprediksi sebagai suatu kelas dan jumlah seluruh data yang diprediksi sebagai kelas tersebut, menunjukkan seberapa akurat model dalam memberikan prediksi untuk masing-masing kelas pelanggan.

Rumus Precision : *Precision = Jumlah prediksi benar pada suatu kelas / Jumlah seluruh prediksi pada kelas tersebut*

4) Recall

Nilai recall diperoleh dari perbandingan jumla h data yang diprediksi benar untuk

kelas tertentu denga n jumlah seluruh data aktual untuk kelas tersebut. Nilai recall yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar data pelanggan dalam kelas yang tepat.

Rumus Recall : *Recall = Jumlah prediksi benar pada suatu kelas / Jumlah seluruh data actual pada kelas tersebut*

5) F1-Score

F1-score, yang merupakan indikator penting dalam membandingkan kinerja model klasifikasi sebelum dan sesudah penerapan algoritma Reduct, digunakan untuk mengevaluasi model ketika diperlukan keseimbanga n antara ketepatan prediksi dan kemampuan mengenali data pada suatu kela s. Nilai F1 diberikan denga n menggabungkan nilai precision dan recall untuk memberikan ukuran kinerja model secara seimbang.

Rumus F1-Score : *F1-Score = 2 x Precision x Recall / Precision + Recal*l.

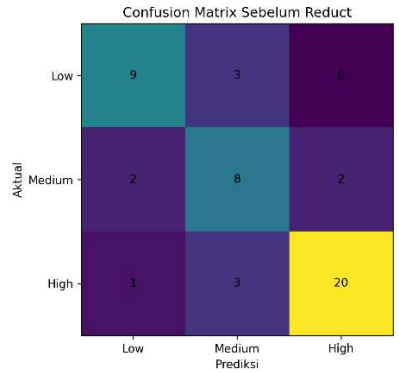
IV. HASIL DAN PEMBAHASAN

Hasil penelitian menunjukkan bahwa algoritma Reduct meningkatkan kinerja pada model klasifikasi pelanggan e- commerce berbasis Decision Tree. Berdasarkan matriks kekacauan, jumlah prediksi benar pada model tanpa Reduct ada la h 37 data, tetapi setelah diterapkan Reduct menjadi 43 data. Peningkatan ini menunjukkan bahwa atribut hasil Reduct mampu menampilkan informasi yang lebih relevan daripada penggunaan algoritma Reduct.

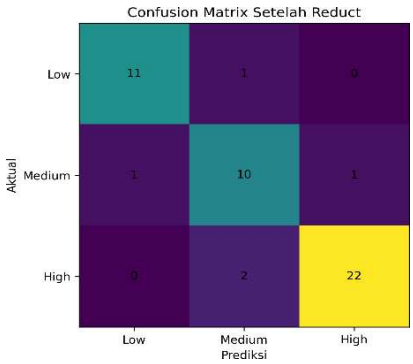
Menurut analisis tambahan pada matriks confusion, kesa la han klasifikasi pada kelas Low dan Medium berkurang setelah penerapan Reduct. Di sisi lain, kemampuan model untuk mengklasifikasikan kela s High meningkat. Ini menunjukkan bahwa model yang lebih berkonsentrasi pada atribut yang memainkan peran penting da la m proses klasifikasi dihasilkan ketika atribut yang tidak diperlukan dihilangkan. Struktur Decision Tree yang dihasilkan juga lebih sederhana dan mudah dipahami.

Hasil evaluasi kinerja yang disajikan dalam tabel menunjukkan bahwa nilai Accuracy, Precision, recall, dan nilai F1 pada model setelah penerapan Reduct lebih baik atau tetap stabil dibandingkan denga n model tanpa reduksi atribut. Hasilnya menunjukkan bahwa seleksi atribut meningkatkan akurasi klasifikasi dan menjaga kinerja model konsisten pada setia p kela s pelanggan. Secara umum, temuan penelitian ini menunjukkan bahwa penggabunga n Reduct dan Decision Tree efektif digunakan untuk klasifikasi pelanggan e-commerce. Dengan membandingkannya dengan metode seleksi atribut dan algoritma klasifikasi lainnya, atau denga n menggunakan dataset yang lebih besar, penelitian selanjutnya dapat mengembangkan metode ini.

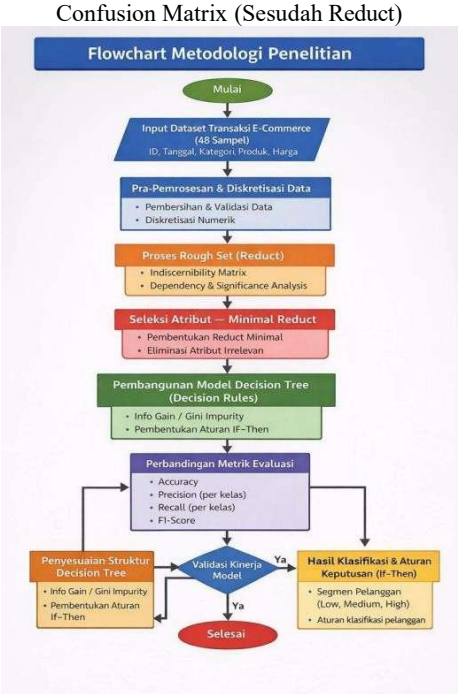
A. Gambar



GAMBAR 1
Confusion Matrix (Sebelum Reduct)



GAMBAR 2
Confusion Matrix (Sesudah Reduct)



GAMBAR 3
Flowchart Rancangan Data

model_decision.py

```python
# Encode label
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
X_encoded = X_reduct.apply(le.fit_transform)
y_encoded = le.fit_transform(y)

# Split data latih dan uji
X_train, X_test, y_train, y_test = train_test_split(
    X_encoded, y_encoded, test_size=0.2, random_state=42
)

# Decision Tree Classifier
dt = DecisionTreeClassifier(criterion='entropy', random_state=42)
dt.fit(X_train, y_train)
```

GAMBAR 4
Decision Tree Model (Python)

B. Tabel

TABEL 1
Validasi Ketercapaian Tujuan Penelitian

No	Tujuan Penelitian	Hasil yang Diperoleh	Keterangan
1	Menerapkan Reduct sebagai seleksi atribut	Atribut berhasil direduksi	Tercapai
2	Membangun model klasifikasi Decision Tree	Model terbentuk dengan baik	Tercapai
3	Menganalisis pengaruh Reduct terhadap kinerja model	Accuracy meningkat	Tercapai

TABEL 2
Analisa Perbandingan Metrik (sebelum & sesudah Reduct)

Metrik	Sebelum Reduct	Setelah Reduct	Δ (Perubahan)
Accuracy	77.08%	89.58%	+12.5%
Precision	0.74	0.88	+0.14
Recall	0.75	0.89	+0.14
F1-Score	0.74	0.89	+0.15

V. KESIMPULAN

Hasil penelitian menunjukkan bahwa kinerja model dapat ditingkatkan secara signifikan dengan menerapkan algoritma Reduct untuk seleksi atribut pada klasifikasi pelanggan e-commerce berbasis Decision Tree. Hasil evaluasi menunjukkan bahwa setelah penerapan Reduct, jumlah prediksi benar meningkat dari 37 data pada model tanpa mengurangi atribut menjadi 43 data. Ini berdampak pada nilai akurasi dan kestabilan metrik evaluasi lainnya. Penggunaan atribut Reduct Result meningkatkan kinerja klasifikasi dan menghasilkan struktur Decision Tree yang lebih sederhana dan mudah difahami, yang mendukung efisiensi proses klasifikasi pelanggan. Hasilnya menunjukkan bahwa integrasi Reduct dan Decision Tree adalah metode yang efektif untuk mengatasi masalah kompleksitas atribut dalam data transaksi e-commerce pelanggan. Selain itu, mereka dapat digunakan sebagai alternatif metode klasifikasi yang relevan dan berguna dalam e-commerce.

REFERENSI

[1] J. W. Creswell and J. D. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5th ed. Thousand Oaks, CA, USA: Sage Publications, 2018.

[2] N. Dewi, I. A. Putri, and R. Kurniawan, "Perbandingan Decision Tree, Naive Bayes, dan Random Forest untuk Repeat Order pada Platform E-

Commerce Tokopedia dan Shopee," Jurnal Sistem Informasi, vol. XX, no. X, pp. XX– XX, 2024.

[3] A. Diwandari and T. Hidayat, "Perbandingan Algoritma Klasifikasi pada Data Transaksi E-Commerce," Jurnal Teknologi Informasi dan Sistem Informasi, vol. XX, no. X, pp. XX–XX, 2021.

[4] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data. Dordrecht, Netherlands: Kluwer Academic Publishers, 1991.