

Penerapan Algoritma C4.5 Untuk Menentukan Prioritas Fitur Dalam Memprediksi Potensi Penjualan Di Toko Rumahbayitaz

Muhammad Naufal Fikri
Department of Information Systems
Telkom University
Bandung, Indonesia
muhnaufalfikri@student.telkomuniversity.ac.id

Irfan Darmawan
Department of Information Systems
Telkom University
Bandung, Indonesia
irfandarmawan@telkomuniversity.ac.id

Oktariani Nurul Pratiwi
Department of Information Systems
Telkom University
Bandung, Indonesia
onurulp@telkomuniversity.ac.id

Abstract— UMKM di Indonesia menghadapi tantangan dalam menentukan strategi produk baru, terutama risiko *overstock* dan *understock* akibat keputusan pengadaan yang bersifat spekulatif. Penelitian ini bertujuan membangun model prediksi kelarisan produk baru menggunakan algoritma C4.5 berbasis *decision tree*. Data diambil dari laporan *Shopee Seller Center* milik Toko RumahBayiTAZ untuk periode 2023–2024, dengan atribut seperti *Produk Dilihat*, *Total Pengunjung*, *Total Pembeli Baru*, dan *Total Potensi Pembeli*. Untuk mengatasi ketidakseimbangan kelas, diterapkan teknik SMOTE (*Synthetic Minority Oversampling Technique*). Model dibangun dengan pustaka *ChefBoost* dalam bahasa pemrograman Python dan dievaluasi menggunakan *Confusion Matrix* serta teknik *K-Fold Cross Validation*. Hasil penelitian menunjukkan akurasi rata-rata 88,9%, dengan fitur *Produk Dilihat* sebagai simpul akar pohon keputusan. Model ini tidak hanya mampu mengklasifikasikan produk ke dalam kategori *Laris*, *Sedang*, dan *Tidak Laris*, tetapi juga membantu pelaku usaha mengidentifikasi fitur-fitur penting untuk perencanaan stok dan strategi promosi produk baru.

Keywords— Algoritma C4.5, Klasifikasi, UMKM, SMOTE, Pohon Keputusan

I. PENDAHULUAN

Dalam era *e-commerce* yang sangat kompetitif, pengelolaan stok menjadi tantangan utama bagi UMKM, termasuk Toko RumahBayiTAZ yang menjual perlengkapan bayi secara daring. Permasalahan *overstock* (kelebihan stok) menyebabkan biaya penyimpanan meningkat dan risiko kerugian membesar, sedangkan *understock* (kekurangan stok) membuat toko gagal memenuhi permintaan pelanggan.

Kondisi ini semakin sulit dihadapi ketika manajemen harus memutuskan stok awal untuk produk baru yang belum memiliki riwayat penjualan. Biasanya keputusan diambil secara spekulatif atau intuitif, tanpa dasar data historis yang memadai. Padahal, bahkan sebelum terjadi penjualan, produk baru di marketplace sudah menghasilkan indikator awal seperti jumlah tampilan produk, total pengunjung, atau jumlah pengguna yang memasukkan produk ke keranjang.

Permasalahan utama penelitian ini adalah bagaimana memanfaatkan data awal tersebut untuk memprediksi kelarisan produk baru. Algoritma C4.5 dipilih karena mampu menangani atribut numerik maupun kategorikal, menghasilkan model berbasis pohon keputusan yang mudah diinterpretasi, serta mengungkap atribut mana yang paling berkontribusi terhadap klasifikasi.

Penelitian ini bertujuan membangun model prediksi kelarisan produk baru pada Toko RumahBayiTAZ dengan algoritma C4.5. Dengan model ini, diharapkan pengelola toko dapat mengambil keputusan pengadaan stok awal dan strategi promosi secara lebih tepat dan berbasis data.

II. METODOLOGI

A. Sumber Data

Data yang digunakan untuk penelitian ini diperoleh secara sekunder dari *Shopee Seller Center* milik Toko RumahBayiTAZ, mencakup periode Januari 2023 hingga Desember 2024. Dataset tersebut terdiri dari berbagai atribut penting yang merepresentasikan performa produk di platform daring. Atribut-atribut tersebut meliputi Nama Produk sebagai identitas unik setiap item yang dijual, Jumlah penjualan yang menunjukkan total unit terjual, *Produk Dilihat* yang mencerminkan jumlah tayangan halaman produk oleh pengguna, *Total Pengunjung* sebagai indikator banyaknya pengguna unik yang mengakses halaman produk, *Total Pembeli Baru* yang menunjukkan akuisisi pelanggan baru, *Total Potensi Pembeli* sebagai sinyal minat pembelian melalui aktivitas seperti memasukkan produk ke keranjang, serta *Total Pesanan* yang merekam jumlah transaksi berhasil.

Data diunduh dalam format *Excel/CSV* yang terstruktur, sehingga memudahkan proses integrasi dan *pra-pemrosesan* selanjutnya untuk analisis prediktif. Pendekatan pengumpulan data ini memastikan bahwa seluruh informasi historis penjualan dapat dianalisis secara sistematis, mendukung pembangunan model klasifikasi kelarisan produk berbasis algoritma C4.5.

B. Pra-pemrosesan Data

Proses *pra-pemrosesan* dilakukan untuk memastikan kualitas data sebelum digunakan dalam pembangunan model klasifikasi. Tahap pertama adalah pembersihan data, yang meliputi penghapusan entri duplikat yang dapat menyebabkan bias dalam distribusi, serta penanganan nilai kosong (*missing values*) yang dapat mengganggu perhitungan atribut. Teknik imputasi sederhana digunakan untuk mengganti nilai kosong dengan nilai median atau modus, bergantung pada tipe atribut, sehingga konsistensi dataset tetap terjaga.

Selanjutnya dilakukan transformasi nilai numerik pada atribut “Jumlah” menjadi variabel target kategorikal dengan tiga kelas: *Laris*, *Sedang*, dan *Tidak Laris*. Penetapan kelas dilakukan menggunakan pendekatan distribusi kuartil pada data historis penjualan. Nilai di atas batas *upper outlier*

dikategorikan sebagai Laris, rentang inter-kuartil sebagai Sedang, dan nilai di bawah median sebagai Tidak Laris. Pendekatan ini bertujuan mencerminkan pola penjualan realistik sesuai distribusi aktual produk.

Selain itu, dilakukan *feature engineering* untuk menciptakan atribut turunan yang memperkaya informasi. Rasio Konversi dihitung sebagai perbandingan antara jumlah pesanan dengan total potensi pembeli ditambah satu, memberikan indikasi seberapa efektif suatu produk dikonversi dari minat ke transaksi. Efektivitas Lihat dihitung sebagai rasio antara jumlah tampilan produk (*Produk Dilihat*) dengan total potensi pembeli ditambah satu, yang mencerminkan efektivitas eksposur dalam mendorong minat pembelian. Kedua atribut ini diharapkan meningkatkan kemampuan model dalam menangkap sinyal awal kelarisan produk baru.

C. Penanganan Ketidakseimbangan Data

Dataset awal menunjukkan distribusi kelas yang tidak seimbang, di mana jumlah contoh pada kelas Laris jauh lebih sedikit dibandingkan kelas Sedang atau Tidak Laris. Ketidakseimbangan ini dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas, sehingga prediksi pada kelas minoritas menjadi kurang akurat.

Untuk mengatasi hal tersebut, penelitian ini menggunakan teknik *SMOTE* (*Synthetic Minority Oversampling Technique*). *SMOTE* menghasilkan data sintesis untuk kelas minoritas dengan membuat contoh baru berdasarkan interpolasi antara contoh minoritas yang berdekatan. Proses ini meningkatkan jumlah contoh pada kelas Laris hingga setara dengan kelas lainnya, menciptakan distribusi yang lebih seimbang tanpa sekadar menggandakan data asli.

Penerapan *SMOTE* dilakukan pada data latih di setiap iterasi *K-Fold Cross Validation*, sehingga memastikan bahwa proses balancing tidak menyebabkan kebocoran data (*data leakage*) ke data uji. Dengan distribusi yang seimbang pada data latih, model dapat belajar pola dari semua kelas secara adil, meningkatkan kemampuan generalisasi terutama untuk mendeteksi produk yang benar-benar berpotensi laris.

D. Algoritma C4.5

Algoritma *C4.5* merupakan metode *decision tree* yang populer karena kemampuannya menangani atribut numerik dan kategorikal sekaligus. Dalam penelitian ini, *C4.5* digunakan untuk mengklasifikasikan produk ke dalam kategori Laris, Sedang, atau Tidak Laris berdasarkan atribut performa awal seperti *Produk Dilihat*, Total Pengunjung, Total Pembeli Baru, dan Total Potensi Pembeli.

Prinsip kerja *C4.5* adalah memilih atribut yang memiliki nilai *Information Gain Ratio* tertinggi untuk membagi data pada setiap simpul pohon. Proses ini dilakukan secara rekursif hingga semua data diklasifikasikan dengan baik atau kriteria penghentian tertentu tercapai. Fitur dengan kontribusi informasi yang rendah akan dieliminasi secara otomatis dari struktur pohon.

Implementasi algoritma dilakukan dengan pustaka *ChefBoost* dalam bahasa *Python*. *ChefBoost* mendukung konfigurasi parameter seperti *rule generation* yang diatur menjadi *strict* untuk memastikan aturan yang dihasilkan lebih deterministik dan mudah diinterpretasi. Model juga dikonfigurasi dengan visualisasi pohon keputusan yang mempermudah analisis terhadap urutan fitur penting dan aturan klasifikasi yang terbentuk.

E. Kerangka Kerja

Penelitian ini menggunakan kerangka kerja *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) untuk memastikan setiap tahap analisis data dilakukan secara sistematis, terstruktur, dan sesuai praktik terbaik industri. *CRISP-DM* terdiri dari enam fase utama yang saling terkait, masing-masing memiliki peran penting dalam keseluruhan proses pembangunan model.

Pada fase *Business Understanding*, peneliti mendefinisikan masalah bisnis yang ingin diselesaikan, yaitu risiko *overstock* dan *understock* akibat sulitnya memprediksi kelarisan produk baru di Toko RumahBayiTAZ. Tujuan bisnis diformulasikan agar penelitian dapat mendukung pengambilan keputusan pengadaan stok yang lebih akurat dan berbasis data.

Fase *Data Understanding* berfokus pada pengumpulan, eksplorasi, dan analisis awal data penjualan historis dari *Shopee Seller Center*. Tahapan ini mencakup identifikasi atribut-atribut penting seperti *Produk Dilihat*, Total Pengunjung, dan Total Potensi Pembeli yang dapat menjadi indikator awal kelarisan produk.

Pada fase *Data Preparation*, dilakukan pembersihan data dari duplikasi dan *missing values*, transformasi atribut penjualan menjadi label kelas Laris, Sedang, dan Tidak Laris, serta *feature engineering* untuk menghasilkan variabel tambahan seperti Rasio Konversi dan Efektivitas Lihat. Data juga diseimbangkan menggunakan teknik *SMOTE* untuk memastikan distribusi kelas yang adil.

Fase *Modeling* melibatkan pembangunan model klasifikasi menggunakan algoritma *C4.5*. Algoritma ini dipilih karena mampu menangani atribut numerik dan kategorikal serta menghasilkan *pohon keputusan* yang mudah diinterpretasi oleh pengelola toko. Model dikembangkan dengan pustaka *ChefBoost* pada *Python*, dengan konfigurasi *rule generation* ketat untuk menghasilkan aturan yang jelas.

Pada fase *Evaluation*, performa model dinilai menggunakan *Confusion Matrix* dan metrik seperti akurasi, *precision*, *recall*, dan *F1-score*. Evaluasi dilakukan dengan pendekatan *K-Fold Cross Validation* ($k=5$) untuk memastikan hasil yang stabil dan mengurangi risiko *overfitting*.

Tahap akhir, *Deployment*, diwujudkan dalam bentuk aplikasi berbasis *Streamlit* yang dirancang untuk memudahkan pengguna non-teknis di Toko RumahBayiTAZ memahami hasil model. Aplikasi ini memungkinkan pengguna mengunggah berkas aturan pohon keputusan (*rules*) hasil pelatihan dalam format *Python*, lalu secara otomatis mengekstrak dan menampilkan urutan prioritas fitur yang digunakan dalam proses klasifikasi.

Dengan antarmuka yang interaktif dan mudah digunakan, sistem membantu pengguna yang kesulitan membaca struktur kode fungsi *findDecision()* untuk langsung melihat daftar atribut yang paling berpengaruh dalam model. Informasi tentang prioritas fitur ini mendukung pengambilan keputusan stok awal dan strategi promosi produk baru secara lebih terukur dan berbasis data, bahkan untuk pengguna yang tidak memiliki latar belakang teknis.

F. Evaluasi Model

Evaluasi model dilakukan secara menyeluruh untuk memastikan akurasi, konsistensi, dan kemampuan generalisasi pada data baru. Metode utama yang digunakan

adalah *Confusion Matrix*, yang memetakan jumlah prediksi benar dan salah pada masing-masing kelas. *Confusion Matrix* memberikan informasi terperinci tentang kesalahan klasifikasi antar kelas, membantu mengidentifikasi bias atau kelemahan model.

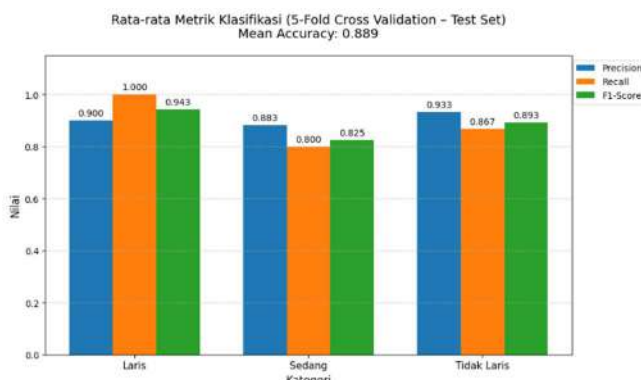
Selain itu, dihitung metrik-metrik kunci seperti akurasi, *precision*, *recall*, dan *F1-score*. Akurasi mengukur proporsi prediksi yang benar dari keseluruhan data, sedangkan *precision* mengindikasikan seberapa tepat model dalam memprediksi kelas tertentu. *Recall* mengukur sejauh mana model mampu menemukan semua contoh sebenarnya dari suatu kelas, dan *F1-score* memberikan keseimbangan antara *precision* dan *recall*.

Penilaian dilakukan dengan pendekatan *K-Fold Cross Validation* dengan $k=5$, membagi data menjadi lima subset. Setiap iterasi menggunakan empat subset sebagai data latih dan satu subset sebagai data uji, memastikan evaluasi model dilakukan pada data yang berbeda-beda. Pendekatan ini membantu mengurangi risiko *overfitting* dan memberikan estimasi performa yang lebih stabil dan dapat diandalkan untuk penerapan dunia nyata.

III. HASIL DAN PEMBAHASAN

A. Hasil Evaluasi Model

Before Evaluasi model dilakukan menggunakan pendekatan *K-Fold Cross Validation* dengan $k=5$ untuk memastikan hasil yang stabil dan meminimalkan risiko *overfitting*. Setelah proses *oversampling* dengan teknik *SMOTE*, data latih memiliki distribusi kelas yang seimbang, sehingga model mampu belajar pola dari ketiga kelas target secara adil.



GAMBAR 1
Rata-rata Metrik Klasifikasi

Gambar diatas menampilkan hasil evaluasi model menggunakan *5-Fold Cross Validation* dengan metrik *Precision*, *Recall*, dan *F1-Score* pada masing-masing kategori kelas, yaitu Laris, Sedang, dan Tidak Laris.

Pada kategori Laris, model mencatat nilai *Precision* sebesar 0.900, *Recall* sempurna di angka 1.000, dan *F1-Score* 0.943. Hal ini menunjukkan bahwa model mampu mendeteksi semua produk yang benar-benar laris (*Recall* 100%), dengan tingkat kesalahan sangat rendah pada prediksi positif (*Precision* mendekati 1). *F1-Score* yang tinggi mengindikasikan keseimbangan baik antara *Precision* dan *Recall* pada kelas ini.

Untuk kategori Sedang, nilai *Precision* tercatat 0.883, *Recall* 0.800, dan *F1-Score* 0.825. Metrik yang sedikit lebih

rendah dibandingkan kelas Laris mengindikasikan tantangan model dalam membedakan produk yang berada di batas antara kelas Sedang dan Tidak Laris. Hal ini wajar karena kategori Sedang secara konseptual merupakan kelas tengah dengan batasan kuartil yang lebih sempit, sehingga rentan overlap.

Pada kategori Tidak Laris, model memperoleh *Precision* sebesar 0.933, *Recall* 0.867, dan *F1-Score* 0.893. Nilai *Precision* yang tinggi menandakan prediksi kelas Tidak Laris cukup tepat, meskipun *Recall* sedikit lebih rendah dibanding *Precision*, menunjukkan bahwa sebagian kecil contoh Tidak Laris masih diklasifikasikan ke kelas lain.

Hasil evaluasi menunjukkan bahwa model *C4.5* mencapai rata-rata akurasi 89,9% pada data uji di kelima fold. *Confusion Matrix* yang dihasilkan memperlihatkan distribusi prediksi yang cukup baik pada kelas Laris, Sedang, dan Tidak Laris, dengan sebagian besar prediksi tepat pada kelas yang sesuai.

Selain akurasi, metrik *precision*, *recall*, dan *F1-score* juga dihitung untuk menilai kinerja model pada setiap kelas. Nilai *precision* dan *recall* yang tinggi pada kelas Laris menandakan kemampuan model dalam mendeteksi produk yang benar-benar laris, sementara performa pada kelas Sedang dan Tidak Laris juga tetap stabil meskipun terdapat beberapa kesalahan minor pada batas antara kelas Sedang dan Tidak Laris.

Tabel berikut menyajikan contoh hasil akurasi rata-rata per fold.

TABEL 1
HASIL AKURASI RATA-RATA PER FOLD

| Fold | Akurasi (%) |
|------------------|-------------|
| 1 | 78 |
| 2 | 89 |
| 3 | 89 |
| 4 | 100 |
| 5 | 89 |
| Rata-rata | 88,9 |

B. Struktur Pohon Keputusan

Model *C4.5* yang dibangun pada penelitian ini menghasilkan struktur *pohon keputusan* yang jelas dan terstruktur, sehingga memudahkan interpretasi bagi pengelola Toko RumahBayiTAZ. Proses pembentukan pohon dilakukan berdasarkan perhitungan *Information Gain Ratio*, memilih atribut dengan nilai informasi tertinggi pada setiap simpul untuk memaksimalkan pemisahan kelas target.

Hasil pelatihan menunjukkan bahwa atribut *Produk Dilihat* muncul secara konsisten sebagai simpul akar (*root node*) dari pohon keputusan. Hal ini menegaskan bahwa visibilitas produk di etalase *Shopee* merupakan indikator awal paling penting dalam memprediksi kelarisan produk. Produk yang lebih sering dilihat cenderung menarik lebih banyak perhatian calon pembeli, meningkatkan peluang masuk kategori Laris.

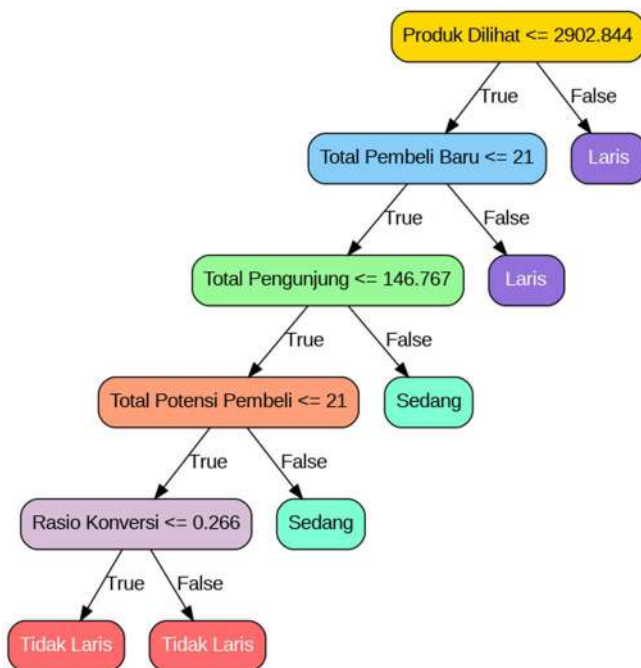
Pada level berikutnya, model memanfaatkan atribut Total Pembeli Baru untuk membagi produk yang memiliki jumlah tampilan lebih rendah. Logika ini mencerminkan pentingnya indikator akuisisi pelanggan baru sebagai penentu tambahan kelarisan. Produk dengan banyak pembeli baru meskipun

tampilannya tidak terlalu tinggi masih memiliki potensi besar untuk laris karena berhasil mengonversi ketertarikan menjadi transaksi aktual.

Total Pengunjung menjadi pembagi pada level selanjutnya, menilai seberapa luas eksposur produk ke pengguna unik. Atribut ini membantu model membedakan produk yang ramai dikunjungi namun tidak selalu dikonversi menjadi pembeli, sehingga mencegah overestimasi kelarisan hanya dari tampilan halaman.

Atribut Total Potensi Pembeli digunakan pada level pohon yang lebih dalam untuk mengklasifikasikan produk dengan lebih halus, khususnya pada batas antara kategori Sedang dan Tidak Laris. Total Potensi Pembeli menggambarkan interaksi lebih aktif seperti memasukkan produk ke keranjang, yang menjadi sinyal minat pembelian meski belum terjadi transaksi.

Berikut adalah contoh aturan keputusan (*rules*) yang dihasilkan oleh model.



GAMBAR 2
Pohon Keputusan Hasil Algoritma C4.5

Contoh aturan keputusan (*rules*) yang dihasilkan oleh model menunjukkan bahwa jika Produk Dilihat melebihi nilai ambang 2902, maka produk diklasifikasikan sebagai Laris. Apabila *Produk Dilihat* berada di bawah atau sama dengan 2902 namun Total Pembeli Baru melebihi 21, produk tetap masuk kategori Laris. Jika *Produk Dilihat* kurang dari atau sama dengan 2902, Total Pembeli Baru tidak lebih dari 21, tetapi Total Pengunjung melebihi 146, maka produk diklasifikasikan sebagai Sedang. Sementara itu, kondisi Total Potensi Pembeli yang berada di bawah atau sama dengan 21 mengindikasikan klasifikasi Tidak Laris. Fitur tambahan seperti Rasio Konversi dan Efektivitas Lihat muncul pada beberapa cabang di tingkat lebih dalam pada pohon keputusan, tetapi tidak konsisten menjadi pembagi utama pada semua jalur. Hal ini menunjukkan bahwa nilai *information gain*-nya relatif lebih rendah, sehingga kontribusinya dalam membedakan kelas kelarisan dianggap kurang signifikan oleh model.

Secara keseluruhan, struktur *pohon keputusan* ini membantu menjelaskan secara transparan bagaimana kombinasi atribut performa awal produk digunakan untuk menentukan potensi kelarisan. Hasil model memberikan panduan praktis kepada pengelola toko untuk memfokuskan perhatian pada fitur-fitur kunci sejak produk pertama kali diunggah di *Shopee Seller Center*, mendukung pengambilan keputusan stok awal dan strategi promosi yang lebih berbasis data.

C. Implementasi dan Interpretasi Bisnis

Before Sebagai bagian dari fase *Deployment* pada kerangka kerja *CRISP-DM*, penelitian ini menghasilkan sistem pendukung keputusan berbasis aplikasi *Streamlit* yang dirancang untuk mempermudah penggunaan model oleh pihak non-teknis, seperti pengelola Toko RumahBayiTAZ. Aplikasi ini tidak hanya bertujuan menjalankan prediksi, tetapi juga membantu pengguna memahami logika di balik klasifikasi yang dihasilkan oleh model *C4.5*.

Aplikasi *Streamlit* yang dikembangkan memungkinkan pengguna mengunggah berkas aturan pohon keputusan (*rules*) hasil pelatihan model dalam format *Python*. Biasanya, fungsi *findDecision()* yang dihasilkan *ChefBoost* memiliki struktur logika bersarang yang sulit dibaca langsung oleh pengguna awam. Untuk mengatasi hambatan ini, sistem dirancang agar dapat mem-parsing isi fungsi tersebut secara otomatis dan mengekstrak urutan prioritas fitur yang digunakan pada setiap simpul pohon.

Dengan fitur ini, pengguna tidak perlu lagi membaca kode program secara manual. Sistem secara otomatis menampilkan daftar fitur yang diurutkan berdasarkan level kemunculannya dalam pohon keputusan—dimulai dari simpul akar (*root node*) hingga cabang yang lebih dalam. Informasi ini sangat penting karena fitur pada level atas memiliki *information gain* tertinggi dan menjadi faktor penentu utama dalam klasifikasi kelarisan produk.

Contoh keluaran antarmuka *Streamlit* menampilkan urutan fitur prioritas seperti berikut.



GAMBAR 3
Tampilan Streamlit

Urutan fitur prioritas yang dihasilkan oleh model menempatkan Produk Dilihat sebagai indikator utama yang merepresentasikan visibilitas produk di etalase daring dan mempengaruhi ketertarikan awal pembeli. Selanjutnya, Total Pembeli Baru mencerminkan kemampuan produk dalam menarik konsumen baru yang belum pernah bertransaksi sebelumnya, menjadi sinyal penting bagi pertumbuhan pasar. Atribut Total Pengunjung menunjukkan seberapa banyak calon pembeli yang tertarik mengunjungi halaman produk, memberikan gambaran tentang eksposur dan jangkauan promosi. Total Potensi Pembeli menggambarkan tingkat

minat lebih lanjut melalui aksi aktif seperti memasukkan produk ke keranjang belanja, menandakan adanya niat pembelian yang lebih serius. Terakhir, Rasio Konversi memberikan gambaran seberapa efektif interaksi pengguna tersebut berubah menjadi pesanan aktual, sehingga menjadi indikator efisiensi dalam mengonversi minat menjadi transaksi.

Melalui visualisasi urutan fitur ini, pengguna Toko RumahBayiTAZ dapat dengan mudah memahami indikator kunci yang perlu diperhatikan ketika menilai potensi kelarisan produk baru. Pengelola dapat fokus pada atribut-atribut yang paling berpengaruh bahkan sebelum produk memiliki riwayat penjualan.

Selain menampilkan prioritas fitur, aplikasi *Streamlit* juga mendukung keterbukaan (*transparency*) dan interpretabilitas model (*explainable AI*), sehingga hasil klasifikasi bukan sekadar prediksi “kotak hitam” (*black box*), melainkan dapat dijelaskan dengan logika yang mudah dimengerti. Hal ini membantu membangun kepercayaan pengguna pada sistem, serta mempermudah kolaborasi antara tim manajemen, promosi, dan pengadaan stok.

Dengan dukungan sistem ini, Toko RumahBayiTAZ diharapkan dapat mengambil keputusan pengadaan stok awal dan strategi promosi dengan lebih tepat sasaran. Hal ini berkontribusi pada pengurangan risiko *overstock* maupun *understock*, peningkatan efisiensi rantai pasok, dan pengelolaan inventaris yang lebih cerdas dan berbasis data.

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi berbasis algoritma *C4.5* untuk memprediksi kelarisan produk baru di Toko RumahBayiTAZ menggunakan data historis yang diperoleh dari *Shopee Seller Center*. Dengan pendekatan kerangka kerja *CRISP-DM (Cross Industry Standard Process for Data Mining)*, proses analisis dilakukan secara sistematis mulai dari pemahaman masalah bisnis hingga tahap *deployment* ke sistem yang dapat digunakan oleh pihak non-teknis.

Model yang dihasilkan menunjukkan performa yang baik dengan akurasi rata-rata 88,9% berdasarkan evaluasi menggunakan *K-Fold Cross Validation* ($k=5$). Hasil struktur pohon keputusan menempatkan Produk Dilihat sebagai atribut utama pada simpul akar, diikuti oleh Total Pembeli Baru, Total Pengunjung, dan Total Potensi Pembeli pada level berikutnya. Hal ini menunjukkan bahwa indikator awal seperti visibilitas produk dan interaksi calon pembeli memiliki kontribusi signifikan dalam menentukan potensi kelarisan bahkan sebelum terjadi transaksi. Visibilitas produk direpresentasikan oleh Produk Dilihat, yang mencerminkan seberapa menarik tampilan produk di etalase daring dan menjadi faktor utama untuk menarik perhatian calon konsumen. Total Pembeli Baru mengindikasikan keberhasilan produk dalam mengakuisisi pelanggan yang sebelumnya belum pernah bertransaksi, menunjukkan potensi pasar yang berkembang. Total Pengunjung memberikan gambaran tentang seberapa luas eksposur produk kepada pengguna unik di platform, menjadi indikator penting untuk menilai tingkat ketertarikan awal. Sementara itu, Total Potensi Pembeli mencerminkan minat lebih lanjut melalui aksi aktif seperti memasukkan produk ke keranjang belanja, menandakan adanya niat pembelian yang lebih serius meskipun transaksi belum terjadi.

Kontribusi utama penelitian ini adalah penyediaan sistem pendukung keputusan (*Decision Support System*) berbasis aplikasi *Streamlit* yang memungkinkan pengguna non-teknis untuk mengunggah berkas aturan pohon keputusan (*rules*) hasil pelatihan model. Aplikasi ini mengekstrak dan menampilkan urutan prioritas fitur secara otomatis, sehingga mempermudah pemahaman terhadap logika klasifikasi yang dihasilkan. Dengan sistem ini, pengelola Toko RumahBayiTAZ dapat membuat keputusan pengadaan stok awal dan strategi promosi produk baru dengan lebih tepat sasaran, mengurangi risiko *overstock* dan *understock*, serta meningkatkan efisiensi pengelolaan inventaris.

Untuk pengembangan ke depan, penelitian ini dapat diperluas dengan mengintegrasikan data dari platform e-commerce lain atau kanal penjualan offline untuk memberikan cakupan yang lebih luas dan representatif. Penambahan fitur kontekstual seperti event promosi besar (*flash sale*, *Harbolnas*), musim liburan, harga pesaing, dan ulasan pelanggan juga diharapkan dapat meningkatkan akurasi model. Selain itu, perbandingan dengan algoritma lain seperti *Random Forest*, *Gradient Boosting*, atau *XGBoost* dapat dilakukan untuk mengeksplorasi potensi peningkatan performa klasifikasi yang lebih lanjut.

REFERENCES

- [1] N. Azwanti, “Optimalisasi pengelolaan stok barang menggunakan algoritma *C4.5*,” *Jurnal Sistem Informasi dan Teknologi Informasi*, vol. 4, no. 1, pp. 12–20, 2018.
- [2] M. Barwa, “Manajemen stok barang dengan teknik data mining,” *Jurnal Teknik dan Informatika*, vol. 7, no. 3, pp. 45–58, 2019.
- [3] A. Izyuddin and S. Wibisono, “Prediksi penjualan produk AC dengan algoritma *C4.5*,” *Jurnal Informatika dan Sistem Informasi*, vol. 12, no. 2, pp. 87–95, 2020.
- [4] R. Rofiani, L. Oktaviani, D. Vernanda, and T. Hendriawan, “Penerapan metode klasifikasi decision tree dalam prediksi kanker paru-paru menggunakan algoritma *C4.5*,” *Jurnal Tekno Kompak*, vol. 18, no. 1, pp. 34–45, 2021.
- [5] A. I. Shafarindu, E. Patimah, Y. M. Siahaan, A. W. Wardhana, B. V. Haekal, and D. S. Prasvita, “Pengelompokan data penjualan supermarket menggunakan algoritma *C4.5*,” *Jurnal Teknologi Informasi*, vol. 9, no. 2, pp. 29–42, 2021.
- [6] D. E. Sinaga, A. P. Windarto, and R. A. Nasution, “Evaluasi algoritma *C4.5* menggunakan teknik cross validation dan confusion matrix pada data penjualan,” *Jurnal Sistem Informasi dan Manajemen Data*, vol. 10, no. 3, pp. 65–78, 2022.
- [7] D. Nababan and A. V. Tanlim, “Optimalisasi stok barang dengan analisis data historis penjualan,” *Jurnal Sistem dan Teknologi Informasi*, vol. 5, no. 1, pp. 18–27, 2019.
- [8] S. Sukmawati, “Analisis pola penjualan menggunakan teknik data mining,” *Jurnal Komputer dan Informatika*, vol. 8, no. 1, pp. 23–36, 2020.
- [9] P. P. Putra and A. S. Chan, “Pemanfaatan algoritma *C4.5* untuk analisis data penjualan e-commerce,” *Jurnal Rekayasa Sistem dan Teknologi Informasi*, vol. 14, no. 2, pp. 40–56, 2021.
- [10] A. D. Saputra and A. Qoiriah, “Perbandingan teknik klasifikasi untuk prediksi penjualan produk di pasar daring,” *Jurnal Sistem Cerdas dan Aplikasi Komputer*, vol. 7, no. 2, pp. 12–23, 2020.
- [11] F. Nasari, “Strategi pengelolaan stok berbasis data mining menggunakan algoritma decision tree,” *Jurnal Ekonomi dan Bisnis Digital*, vol. 11, no. 4, pp. 75–88, 2021.
- [12] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.

- [13] A. Kristianus Lalo, P. Batarius, and Y. C. Hoar Siki, "Implementasi decision tree untuk prediksi produk terlaris di e-commerce," *Jurnal Sistem Informasi dan Komputerisasi*, vol. 3, no. 3, pp. 22–35, 2020.
- [14] G. L. Pritalia, "Analisis data penjualan supermarket menggunakan metode klasifikasi," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 5, no. 3, pp. 55–67, 2021.
- [15] R. Rofiani, "Penerapan decision tree untuk prediksi kebutuhan stok di pasar ritel," *Jurnal Informatika dan Sistem*, vol. 8, no. 1, pp. 30–40, 2021.
- [16] N. F. Muhidin, D. Gustian, and Sihabudin, "Analisis dan penerapan algoritma C4.5 untuk memprediksi kualitas penelitian dan publikasi ilmiah," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, pp. 2454–2460, 2022.
- [17] J. Erbani, P.-É. Portier, E. Egyed-Zsigmond, and D. Nurbakova, "Confusion matrices: A unified theory," *IEEE Access*, vol. 12, pp. 181372–181379, 2024.
- [18] Murlena and D. Apriana, "Penerapan data mining untuk memprediksi ketersediaan stok produk HNI HPAI menggunakan algoritma C4.5," *Arcitech: Journal of Computer Science and Artificial Intelligence*, vol. 2, no. 1, pp. 19–32, 2022.
- [19] M. Liandana and I. M. D. Susila, "Pengaruh jumlah fitur pada algoritma machine learning dalam memprediksi aktivitas jatuh," *Jurnal Sistem dan Informatika*, vol. 17, no. 2, pp. 109–120, 2023.
- [20] D. Ruswanti, D. Susilo, and Riani, "Implementasi CRISP-DM pada data mining untuk melakukan prediksi pendapatan dengan algoritma C4.5," *GO Infotech: Jurnal Ilmiah STMIK AUB*, vol. 30, no. 1, pp. 111–121, 2024.
- [21] N. C. Yulianto and A. Y. Nisrina, "Klasifikasi penderita diabetes mellitus menggunakan metode SVM, regresi logistik biner, decision tree, dan random forest," Institut Teknologi Sepuluh Nopember, 2023.