

ANALISIS SENTIMEN ULASAN PELANGGAN BERBASIS *DEEP LEARNING* UNTUK Mendukung CRM Analytics dalam Mencegah CHURN PELANGGAN PADA *E-COMMERCE*

Dzidan Aditya Gumilang, Msy Anggita Shafira,
Jhon Kevin Andrianto Sinaga, Fathoni, Ali Ibrahim

Sistem Informasi, Universitas Sriwijaya
Jl. Raya Palembang-Prabumulih KM. 32, Indralaya, Ogan Ilir,
Sumatera Selatan 30862, Indonesia
dzdnaditya@gmail.com

ABSTRAK

Tingkat customer churn yang tinggi merupakan tantangan utama bagi perusahaan e-commerce. Selama ini, sebagian besar model prediksi churn hanya bergantung pada data transaksi dan jarang melibatkan opini tekstual pelanggan. Penelitian ini bertujuan memprediksi probabilitas churn dengan menggabungkan analisis sentimen ulasan dengan variabel perilaku Customer Relationship Management (CRM) berupa RFM (Recency, Frequency, Monetary) dan data logistik. Algoritma Long Short-Term Memory (LSTM) digunakan untuk mengklasifikasikan sentimen dari 8.196 data ulasan. Hasil ekstraksi sentimen ini kemudian diintegrasikan dengan data transaksi historis untuk melatih algoritma Random Forest dalam memprediksi churn. Hasil pengujian menunjukkan model LSTM mencapai akurasi 88,44% dalam klasifikasi teks sentimen. Pada tahap klasifikasi churn, model Random Forest menghasilkan nilai Recall 84%, yang membuktikan keandalannya dalam mendeteksi pelanggan berisiko. Selanjutnya, analisis Feature Importance menemukan bahwa total nilai belanja (Monetary), tingginya rasio ongkos kirim, dan keterlambatan paket merupakan faktor paling dominan yang memicu churn, sedangkan frekuensi belanja tidak memberikan pengaruh yang signifikan.

Kata kunci : *Customer Churn, Analisis Sentimen, Deep Learning, Random Forest, E-commerce, CRM*

1. PENDAHULUAN

Persaingan ketat di industri e-commerce saat ini menuntut perusahaan untuk menghadapi masalah tingginya angka customer churn (kondisi di mana pelanggan berhenti berbelanja atau pindah ke platform kompetitor). Mengingat biaya untuk mencari pelanggan baru jauh lebih mahal daripada mempertahankan pelanggan lama, tingginya tingkat churn berdampak langsung pada merosotnya margin keuntungan perusahaan. Untuk mengatasi hal ini, berbagai metode machine learning (ML) telah banyak diterapkan untuk memprediksi pelanggan mana yang berpotensi pergi. Pendekatan ML seperti Random Forest atau Gradient Boosting terbukti lebih unggul dibanding metode statistik tradisional karena kemampuannya mengenali pola perilaku dari jutaan baris data historis pelanggan [1], [2], [3]. Performa prediksi algoritma ini juga dapat ditingkatkan lebih jauh dengan melakukan optimasi seleksi fitur untuk membuang data yang tidak relevan [4].

Seiring dengan semakin rumitnya data transaksi di platform digital, arsitektur deep learning (DL) dan model gabungan (hybrid ensemble) juga mulai banyak diaplikasikan. Sebagai contoh, penelitian Saha et al. (2024) menggunakan Convolutional Neural Network (CNN) untuk fokus pada titik-titik krusial dalam riwayat transaksi pengguna [5]. Penggabungan beberapa model algoritma sekaligus juga terbukti mampu menutupi kelemahan satu algoritma tertentu, sehingga hasil prediksinya menjadi lebih stabil dan akurat pada berbagai kondisi data [6], [7].

Namun, jika ditelaah dari berbagai studi terdahulu, masih terdapat satu celah penelitian yang menarik. Sebagian besar model prediksi churn saat ini terlalu bergantung pada data kuantitatif, seperti frekuensi transaksi atau nominal belanja. Padahal, angka-angka transaksi ini hanya menunjukkan rekam jejak masa lalu, tetapi tidak bisa menjelaskan "alasan" di balik keputusan pelanggan untuk berhenti. Di sisi lain, masih sedikit penelitian yang memanfaatkan ulasan teks pelanggan (unstructured data). Ulasan pelanggan sangat penting karena mencerminkan emosi dan tingkat kepuasan mereka secara langsung. Sangat mungkin terjadi, seorang pelanggan yang transaksinya terlihat masih aktif secara statistik sebenarnya sudah bersiap untuk churn yang ditandai dengan ulasan-ulasan negatif.

Berangkat dari masalah tersebut, penelitian ini mencoba membangun model prediksi churn yang menggabungkan dua sisi sekaligus: data kualitatif (opini teks) dan data kuantitatif (riwayat transaksi). Secara teknis, alur penelitian ini dirancang dalam dua tahap utama. Pertama, membangun model deep learning berbasis Long Short-Term Memory (LSTM) untuk membaca sentimen dari 8.196 ulasan pelanggan dan mengubahnya menjadi skor numerik. Kedua, skor sentimen tersebut akan digabungkan dengan variabel riwayat belanja (Recency, Frequency, Monetary) serta catatan logistik pengiriman (seperti rasio ongkir dan keterlambatan paket) untuk melatih model Random Forest.

Fokus utama evaluasi dalam pemodelan ini adalah memaksimalkan nilai Recall untuk memastikan sistem tidak kebobolan dalam mendeteksi pelanggan berisiko tinggi (false negative). Di akhir analisis, penelitian ini juga akan membongkar faktor operasional mana yang paling memicu pelanggan e-commerce untuk pergi (Feature Importance), sehingga temuan akhirnya dapat dijadikan landasan praktis bagi perusahaan dalam memperbaiki strategi retensi pelanggannya.

2. TINJAUAN PUSTAKA

Berbagai literatur terdahulu menempatkan *customer churn* sebagai fokus utama dalam analitik bisnis karena dampaknya terhadap kelangsungan perusahaan. Mengingat kompleksitas perilaku pelanggan, berbagai metode komputasi telah dikembangkan untuk mengoptimalkan pemanfaatan data historis guna mendeteksi potensi *churn* sejak dini [6], [8].

Dalam upaya mitigasi tersebut, pemanfaatan *machine learning* (ML) seperti *Random Forest*, *XGBoost*, dan *Gradient Boosting* terbukti lebih baik dibandingkan metode statistik tradisional. Khususnya, model *ensemble* dinilai lebih andal karena kemampuannya menggabungkan beberapa algoritma untuk menghasilkan prediksi yang lebih stabil. Lebih lanjut, integrasi teknik interpretasi seperti SHAP (*Shapley Additive exPlanations*) memungkinkan peneliti untuk menganalisis model *black-box*, sehingga perusahaan dapat memahami faktor spesifik yang memicu pelanggan untuk beralih [9], [10].

Akurasi model prediksi ini sangat bergantung pada kualitas fitur yang digunakan. Proses seleksi fitur (*feature selection*) menjadi tahapan penting untuk mengeliminasi data yang tidak relevan atau *noise*. Sebagai contoh, penerapan metode optimasi seperti *Improved Ant Colony Optimization* telah dikonfirmasi mampu menyaring variabel yang paling relevan, yang pada akhirnya meningkatkan efisiensi komputasi dan akurasi prediksi secara keseluruhan [11].

Seiring makin dinamisnya data *e-commerce*, arsitektur *deep learning* (DL) mulai banyak diadaptasi untuk menangkap pola hubungan yang non-linear. Model seperti *Convolutional Neural Network* (CNN) maupun jaringan hibrida terbukti memberikan peningkatan akurasi dibandingkan ML konvensional, meskipun model ini sering kali membutuhkan sumber daya komputasi yang lebih besar [5], [12]. Namun, sebagian besar algoritma DL dalam literatur prediksi *churn* masih terbatas pada pemrosesan data perilaku kuantitatif (seperti frekuensi dan nominal transaksi), sehingga belum dapat menjelaskan alasan psikologis di balik keputusan pelanggan tersebut.

Untuk menggali konteks kualitatif tersebut, pendekatan analisis sentimen terhadap teks ulasan pelanggan mulai banyak dieksplorasi. Beberapa studi awal menggunakan teknik *text embedding* tradisional seperti TF-IDF, CBOW, dan GloVe untuk mengklasifikasikan opini. Namun, untuk teks ulasan

yang panjang dan memiliki konteks sekuensial yang kompleks, arsitektur berbasis *Recurrent Neural Network* (RNN), khususnya varian *Long Short-Term Memory* (LSTM), terbukti lebih baik dalam mengingat konteks kalimat jangka panjang dan mengekstraksi polaritas sentimen secara presisi [13], [14].

Merujuk pada paparan literatur di atas, terdapat peluang untuk membangun kerangka prediksi *churn* yang lebih komprehensif. Oleh karena itu, sintesis yang mengintegrasikan data perilaku pelanggan secara kuantitatif dengan hasil analisis sentimen ulasan berbasis LSTM diharapkan mampu melengkapi kelemahan model-model sebelumnya. Melalui pendekatan ini, strategi *Customer Relationship Management* (CRM) dapat dirumuskan secara lebih proaktif dan tepat sasaran berdasarkan pemahaman analitis atas perilaku dan emosi pelanggan.

3. METODE PENELITIAN

3.1. Kerangka Penelitian

Penelitian ini menggunakan pendekatan campuran dengan mengintegrasikan analisis teks (kualitatif) dan data kuantitatif untuk memprediksi *customer churn*. Data dikumpulkan melalui studi dokumentasi dari *dataset* publik sekunder *e-commerce* yang mencakup riwayat transaksi dan ulasan pelanggan. Pelaksanaan penelitian dibagi ke dalam lima tahapan utama. Tahap pertama adalah pengumpulan dan prapemrosesan data (*data cleansing* dan *text preprocessing*). Tahap kedua adalah klasifikasi ulasan ke dalam skor sentimen menggunakan algoritma *deep learning Long Short-Term Memory* (LSTM). Tahap ketiga berfokus pada pembentukan variabel perilaku pelanggan berbasis RFM (*Recency, Frequency, Monetary*) dan data operasional pengiriman. Pada tahap keempat, data sentimen dan data transaksi diintegrasikan pada tingkat pelanggan individu untuk melatih algoritma *machine learning* dalam memprediksi probabilitas *churn*. Tahap kelima adalah evaluasi performa model dan analisis variabel prediktor menggunakan *Feature Importance* untuk merumuskan rekomendasi strategi *Customer Relationship Management* (CRM) yang didorong oleh *AI-driven predictive analytics* [15].

3.2. Sumber dan Pengumpulan Data

Data penelitian bersumber dari *dataset e-commerce* publik kaggle yang memuat riwayat transaksi terstruktur dan opini teks pelanggan. Data transaksi diekstraksi dengan menggabungkan tabel *customers*, *orders*, *order_items*, dan *order_payments* berdasarkan ID pelanggan dan ID pesanan. Untuk analisis sentimen, sampel ulasan dibatasi sebanyak 8.196 data menggunakan teknik *purposive sampling*. Pembatasan ini bertujuan untuk menyaring teks yang memiliki konteks relevan, menjaga keseimbangan kelas sentimen, dan menjaga efisiensi komputasi komputer selama proses pelatihan model.

3.3. Pra-pemrosesan Data

Kualitas data merupakan faktor penentu utama terhadap tingkat akurasi model analitik yang dibangun. Sebelum memasuki tahap pelatihan algoritma (training), dataset mentah harus melalui serangkaian teknik pra-pemrosesan untuk mengatasi anomali, menghilangkan noise, serta mengubah format data agar kompatibel dengan arsitektur komputasi. Mengingat penelitian ini menggunakan dua dimensi data yang berbeda, teknik pra-pemrosesan dijalankan melalui dua jalur utama, yakni pemrosesan teks opini dan pemrosesan atribut transaksi.

Pada jalur pra-pemrosesan data tekstual, ulasan pelanggan yang umumnya tidak terstruktur dan memuat kesalahan penulisan disiapkan agar dapat dipahami oleh model deep learning (LSTM). Tahapan Natural Language Processing (NLP) yang diaplikasikan sejalan dengan standar optimalisasi klasifikasi sentimen [16], [17], yang meliputi:

- Case Folding & Cleansing*: Menyeragamkan seluruh karakter menjadi huruf kecil (*lowercase*) serta mengeleminasi tanda baca, angka, tautan, dan simbol non-alfabet yang tidak memiliki bobot semantik.
- Tokenization*: Memecah kalimat ulasan yang utuh menjadi potongan-potongan kata tunggal (*token*) agar algoritma dapat menganalisis struktur bahasa secara parsial.
- Stopword Removal*: Menghapus kata-kata penghubung atau kata umum (seperti "yang", "dan", "di") yang sering muncul namun tidak membawa makna spesifik terkait polaritas sentimen.
- Word Embedding & Padding*: Mengonversi token kata menjadi representasi vektor numerik (*word embedding*) dan menyeragamkan panjang urutan data (*padding*) sehingga memenuhi standar dimensi matriks masukan (*input shape*) pada LSTM.

Sementara itu, pada jalur pra-pemrosesan data kuantitatif, riwayat transaksi terstruktur diproses untuk mengekstraksi indikator kepuasan dan perilaku *churn*. Tahapan pemrosesan atribut numerik ini meliputi:

- Penanganan *Missing Values*: Mengidentifikasi dan menangani nilai kosong pada atribut transaksi menggunakan teknik imputasi statistik atau penghapusan baris (*dropping*) jika data gagal direkonstruksi.
- Feature Engineering*: Mengonstruksi variabel prediktor baru dari perhitungan matematis atribut mentah. Proses ini menghasilkan fitur loyalitas operasional berupa *Recency* (hari sejak transaksi terakhir), *Frequency* (total pesanan), *Monetary* (total pembelanjaan), keterlambatan pengiriman, dan rasio ongkos kirim.
- Normalisasi Data: Menyeragamkan skala pengukuran pada fitur numerik menggunakan *Standard Scaler* untuk memastikan algoritma klasifikasi tidak bias terhadap variabel yang memiliki rentang nilai lebih besar.

- Feature Selection*: Mengeliminasi atribut demografis atau spesifikasi produk yang tidak relevan, dan murni mempertahankan fitur perilaku (*behavioral*). Praktik reduksi fitur ini secara langsung merujuk pada temuan Al-Shourbaji *et al.*, yang membuktikan bahwa seleksi karakteristik pelanggan yang esensial akan mereduksi *noise*, mencegah terjadinya *overfitting*, dan secara signifikan meningkatkan akurasi algoritma dalam memprediksi status *churn* pelanggan [18].

3.4. Arsitektur Deep Learning untuk Analisis Sentimen

Tahap ini menggunakan arsitektur komputasi *Long Short-Term Memory* (LSTM) untuk mengekstraksi nilai sentimen dari ulasan pelanggan. LSTM dipilih karena mampu mengatasi masalah *vanishing gradient* dan baik dalam menangani data sekuensial berupa teks panjang [19]. Dalam penelitian ini, model LSTM dikonfigurasi untuk menerima *input* berupa *word embedding*, mempelajari relasi antar kata, dan menghasilkan *output* probabilitas biner yang mengklasifikasikan ulasan menjadi sentimen positif (1) atau negatif (0).

3.5. Pemodelan Prediksi Churn dan Integrasi Data

Setelah skor sentimen diperoleh dari model LSTM, data tersebut digabungkan dengan fitur perilaku transaksi berdasarkan identitas pelanggan (*customer_unique_id*). Variabel target dalam tahap ini adalah status *churn*, di mana pelanggan yang tidak bertransaksi selama enam bulan terakhir diberi label 1 (*churn*), dan pelanggan yang masih aktif diberi label 0.

Prediksi *churn* dilakukan menggunakan algoritma *ensemble*, yaitu *Random Forest* dan *XGBoost*. Metode *ensemble learning* ini dinilai sangat efektif untuk menangani hubungan non-linear dan imbalans data, serta terbukti secara konsisten memberikan performa akurasi prediksi yang lebih tangguh dibandingkan algoritma klasifikasi tradisional [1], [2]. Integrasi *Explainable AI* (XAI) menggunakan pendekatan seperti SHAP juga diterapkan untuk mengetahui variabel transaksi atau sentimen mana yang paling berkontribusi dalam memicu keputusan pelanggan untuk meninggalkan layanan [20], [21].

3.6. Teknik Evaluasi Model

Performa model dievaluasi pada dua tahap: evaluasi klasifikasi sentimen (LSTM) dan evaluasi prediksi *churn* (*ensemble learning*). Pengukuran dilakukan menggunakan *Confusion Matrix* untuk membandingkan hasil prediksi dengan data aktual (*ground truth*). Metrik yang digunakan meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

Dalam konteks CRM, evaluasi model ini mempertimbangkan risiko bisnis operasional. Penelitian ini berfokus pada optimalisasi nilai *Recall* untuk meminimalkan *False Negative* (sistem gagal

mendeteksi pelanggan yang akan pergi), karena dapat berpotensi menghilangkan pendapatan perusahaan. Namun demikian, nilai *Precision* tetap dipertahankan pada batas yang wajar agar perusahaan tidak salah sasaran dalam mendistribusikan anggaran retensi, seperti pemberian voucher diskon, kepada pelanggan yang sebenarnya masih berstatus loyal.

4. HASIL DAN PEMBAHASAN

4.1. Evaluasi Model Deep Learning untuk Klasifikasi Sentimen

Tahap evaluasi bertujuan untuk mengukur kinerja model *deep learning* dalam mengklasifikasikan sentimen ulasan pelanggan *e-commerce*. Pengujian ini menggunakan empat metrik utama: *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Evaluasi difokuskan pada 8.196 data uji dengan rincian distribusi kelas seperti yang disajikan pada Tabel 1.

Tabel 1. Distribusi data sentimen pelanggan

Label	Kelas Sentimen	Jumlah Data
0	Sentimen Negatif	2.881
1	Sentimen Positif	5.315
Total		8.196

Tabel 1 menunjukkan adanya ketidakseimbangan kelas (*imbalanced data*), di mana jumlah ulasan bersentimen positif hampir dua kali lipat dibandingkan ulasan negatif. Meskipun proporsi ini mencerminkan tingkat kepuasan pelanggan yang baik secara umum, 2.881 ulasan negatif tetap menjadi indikator penting. Kelompok ulasan negatif ini merupakan representasi pelanggan yang memiliki kecenderungan tinggi untuk melakukan *churn*.

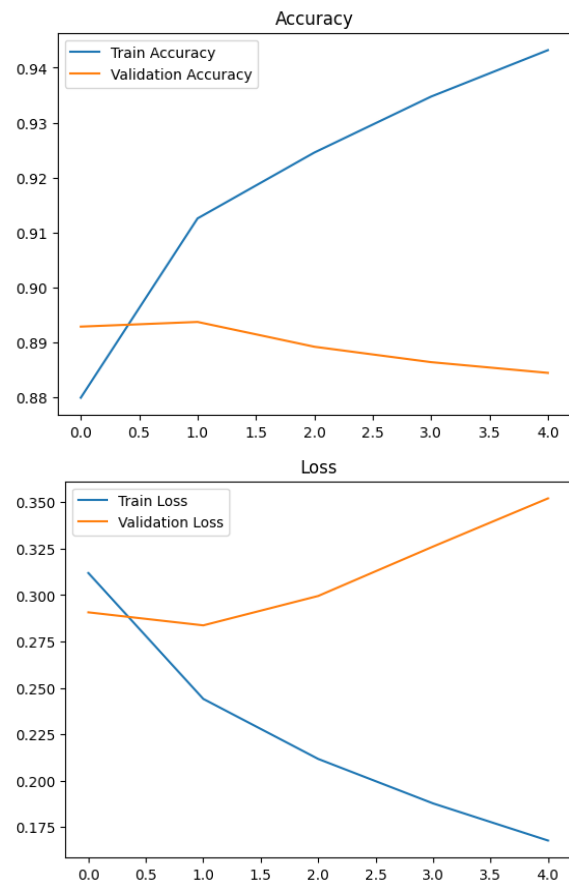
Hasil pengujian model *deep learning* menunjukkan tingkat akurasi sebesar 88,44%. Rincian performa klasifikasi untuk masing-masing kelas sentimen disajikan pada Tabel 2.

Tabel 2. Hasil klasifikasi model *deep learning*

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Support
Negatif (0)	0.83	0.84	0.84	2881
Positif (1)	0.91	0.91	0.91	5315
Accuracy			0.88	8196
Macro Avg	0.87	0.87	0.87	8196
Weighted Avg	0.88	0.88	0.88	8196

Berdasarkan Tabel 2, model memiliki performa yang lebih baik dalam mengenali sentimen positif dibandingkan sentimen negatif. Nilai *Precision*, *Recall*, dan *F1-Score* untuk kelas positif mencapai 0.91, sedangkan metrik untuk kelas negatif berada pada rentang 0.83 hingga 0.84. Perbedaan performa ini wajar terjadi karena model dilatih dengan jumlah data latih (*support*) yang lebih banyak pada kelas positif. Meskipun demikian, nilai *F1-Score* yang stabil di atas 0.80 untuk kedua kelas menunjukkan bahwa model

sudah cukup baik dalam mengekstraksi dan mengklasifikasikan konteks teks ulasan.



Gambar 1. Grafik *accuracy* dan *loss model*

Performa model juga dievaluasi melalui kurva pelatihan pada Gambar 1. Kurva akurasi menunjukkan tren peningkatan yang stabil antara data latih dan data validasi tanpa memunculkan perbedaan (*gap*) yang lebar. Hal ini mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik dan terhindar dari *overfitting*. Sejalan dengan itu, grafik *loss* yang terus melandai menunjukkan model secara efektif mampu meminimalkan tingkat kesalahan prediksi (*error*) selama iterasi pelatihan.

Dalam konteks penerapan analitik bisnis (*Customer Relationship Management / CRM*), model klasifikasi sentimen ini dapat dimanfaatkan sebagai sistem peringatan dini. Saat sistem mendeteksi ulasan bersentimen negatif, perusahaan dapat mengambil tindakan preventif dengan merespons keluhan secara *real-time* sebelum pelanggan beralih ke platform kompetitor.

Hasil pengujian ini membuktikan bahwa opini kualitatif pelanggan dapat diekstraksi menjadi data kuantitatif yang terukur. Pada tahapan penelitian selanjutnya, probabilitas sentimen ini akan diintegrasikan bersama variabel operasional dan riwayat transaksi untuk membangun model prediksi *churn* pelanggan yang lebih komprehensif.

4.2. Ekstraksi Fitur Perilaku Pelanggan (*Feature Engineering*)

Skor sentimen yang diperoleh pada tahap sebelumnya baru mewakili opini pelanggan secara kualitatif. Untuk membangun model prediksi yang utuh, riwayat transaksi pelanggan perlu diekstraksi menjadi fitur perilaku kuantitatif. Karena data transaksi *e-commerce* umumnya tersebar dalam beberapa tabel relasional (seperti tabel pesanan, pembayaran, dan rincian barang), tahap awal yang dilakukan adalah proses agregasi. Tahapan ini bertujuan menyatukan seluruh riwayat transaksi tersebut agar setiap pelanggan memiliki satu profil data tunggal (*customer-centric*).

Ekstraksi fitur perilaku belanja pada penelitian ini dilakukan menggunakan pendekatan RFM (*Recency, Frequency, Monetary*). Variabel *Recency* dihitung berdasarkan selisih hari antara transaksi terakhir pelanggan dan batas akhir waktu pada *dataset*. Variabel *Frequency* direpresentasikan oleh total

pesanan unik yang berhasil diselesaikan, sedangkan *Monetary* merupakan akumulasi nominal pengeluaran pelanggan selama menggunakan platform tersebut.

Meskipun demikian, indikator RFM standar belum cukup untuk merepresentasikan perilaku pelanggan secara menyeluruh. Dalam transaksi *e-commerce*, masalah logistik sering kali menjadi faktor ketidakpuasan utama (*pain point*). Oleh karena itu, penelitian ini menambahkan dua variabel operasional pendukung. Variabel pertama adalah Keterlambatan Pengiriman (*Delivery Delay*), yang dihitung dari selisih antara estimasi waktu tiba dan waktu aktual paket diterima oleh pelanggan. Variabel kedua adalah Rasio Ongkos Kirim, yakni persentase perbandingan antara biaya pengiriman (*freight value*) dan harga barang. Kedua fitur ini diekstraksi karena tingginya biaya kirim dan keterlambatan kedatangan paket berpotensi memicu pengalaman negatif yang berujung pada *churn*.

Tabel 3. Contoh hasil ekstraksi fitur pelanggan

ID Pelanggan	Recency	Frequency	Monetary	Delay	Rasio Ongkir	Status Churn
0000366f...	160	1	142	-5	0.09	0
0000b849...	163	1	27	-5	0.44	0
0000f46a...	585	1	86	-2	0.25	1
0000f6cc...	369	1	44	-12	0.68	1
0004aac8...	336	1	197	-8	0.09	1

Tahap akhir dalam penyiapan data ini adalah pelabelan target *churn*. Berbeda dengan sistem berbasis langganan yang memiliki status pembatalan yang jelas (*unsubscribe*), pelanggan *e-commerce* memiliki karakteristik non-kontraktual, yaitu dapat berhenti berbelanja tanpa perlu menghapus akun. Untuk mengatasi hal tersebut, penelitian ini menetapkan batasan waktu ketidakaktifan (vakum). Pelanggan yang tidak melakukan pembelian sama sekali selama 180 hari (sekitar 6 bulan) terakhir diklasifikasikan sebagai *churn* (label 1), sedangkan pelanggan yang masih bertransaksi dalam kurun waktu tersebut diberi label aktif (label 0).

Dataset perilaku operasional yang memuat indikator RFM, variabel logistik, dan label target ini selanjutnya digabungkan dengan probabilitas sentimen untuk digunakan pada tahap pelatihan algoritma prediksi *churn*.

4.3. Hasil Prediksi Churn Pelanggan

Pada tahap pemodelan, fitur perilaku yang telah diekstraksi digunakan untuk melatih algoritma *Random Forest Classifier*. Pemilihan algoritma *ensemble* ini didasarkan pada kemampuannya dalam memproses pola data yang kompleks sekaligus menekan risiko *overfitting* pada data transaksi *e-commerce*. Evaluasi model prediksi *churn* ini menggunakan 19.220 baris data uji (20% dari total *dataset*) yang dipisahkan dari data latih agar model dapat diuji secara objektif.

Hasil pengujian menunjukkan bahwa model mencapai akurasi keseluruhan sebesar 71,16%. Namun, dalam analisis *machine learning*, mengandalkan nilai akurasi tunggal berpotensi menghasilkan interpretasi yang bias, terutama pada distribusi data yang tidak seimbang. Oleh karena itu, evaluasi performa model dilanjutkan dengan menganalisis metrik *Precision*, *Recall*, dan *F1-Score* untuk setiap kelas target, sebagaimana dirangkum pada Tabel 4.

Tabel 4. Hasil evaluasi prediksi *churn* dengan model *random forest*

Kelas	Precision	Recall	F1-Score	Support (Jumlah Data)
0 (Aktif)	0.51	0.41	0.45	5.590
1 (Churn)	0.77	0.84	0.80	13.630
Accuracy			0.71	19.220
Macro Avg	0.64	0.62	0.63	19.220
Weighted Avg	0.70	0.71	0.70	19.220

Berdasarkan kolom *Support* pada Tabel 4, terlihat adanya ketidakseimbangan proporsi data (*class imbalance*). Jumlah pelanggan yang terindikasi *churn* (13.630 data) jauh lebih besar dibandingkan jumlah pelanggan aktif (5.590 data). Meskipun tidak seimbang secara statistik, fenomena ini wajar terjadi dalam realitas bisnis *e-commerce*. Konsumen platform digital umumnya memiliki loyalitas yang rendah, di

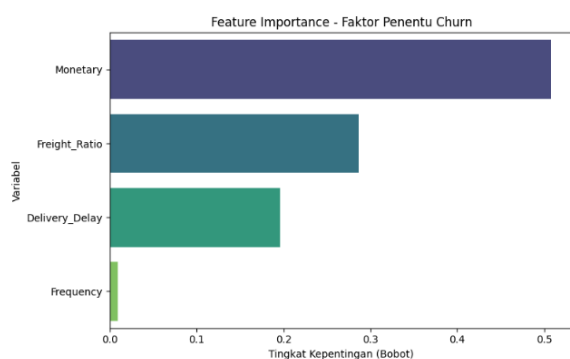
mana sebagian besar dari mereka merupakan pembeli satu kali (*one-time buyer*) yang tidak akan kembali bertransaksi tanpa adanya pendorong retensi yang kuat.

Evaluasi paling krusial pada pemodelan ini terletak pada nilai *Recall* untuk kelas *churn* (1) yang mencapai 0.84. Dalam kerangka strategi *Customer Relationship Management* (CRM), tingginya sensitivitas model ini merupakan indikator yang sangat positif. Angka tersebut mengindikasikan bahwa dari keseluruhan populasi pelanggan yang sebenarnya berniat meninggalkan aplikasi, model berhasil mendeteksi 84% di antaranya secara akurat.

Dari perspektif manajerial, model klasifikasi ini sesuai dengan kebutuhan efisiensi bisnis perusahaan. Dalam upaya menekan tingkat *churn*, kesalahan mendeteksi pelanggan aktif sebagai pelanggan *churn* (*False Positive*) dianggap memiliki risiko yang jauh lebih rendah dibandingkan gagal mendeteksi pelanggan yang benar-benar berisiko pergi (*False Negative*). Penurunan performa pada kelas aktif (ditunjukkan oleh nilai *F1-Score* kelas 0 sebesar 0.45) merupakan bentuk kompromi (*trade-off*) yang rasional guna memaksimalkan deteksi *churn* dan mencegah hilangnya pendapatan perusahaan.

4.4. Analisis Faktor Penentu Churn (Feature Importance)

Salah satu kelebihan algoritma *Random Forest* adalah kemampuannya memberikan interpretabilitas (*explainability*). Artinya, model tidak hanya memprediksi status *churn*, tetapi juga mampu mengidentifikasi faktor prediktor yang dominan dalam memicu keputusan tersebut. Besaran pengaruh dari setiap variabel divisualisasikan melalui grafik *Feature Importance* pada Gambar 2.



Gambar 2. Grafik *feature importance*

Berdasarkan grafik tersebut, kontribusi masing-masing variabel terhadap probabilitas *churn* bervariasi, dengan rincian bobot sebagai berikut:

- Total Pembelanjaan / *Monetary* (Bobot: 0,5075): Variabel ini memiliki kontribusi terbesar, yakni memengaruhi lebih dari 50% keputusan *churn*. Pelanggan dengan akumulasi transaksi bernilai tinggi umumnya memiliki ekspektasi layanan yang lebih besar. Penurunan drastis atau berhentinya aktivitas belanja dari kelompok ini

dideteksi oleh model sebagai sinyal *churn* yang paling kuat.

- Rasio Ongkos Kirim / *Freight Ratio* (Bobot: 0,2869): Berada di urutan kedua, temuan ini mengonfirmasi bahwa biaya logistik merupakan titik kritis bagi konsumen. Jika proporsi ongkos kirim dinilai terlalu tinggi dibandingkan harga barang, pelanggan memiliki kecenderungan lebih besar untuk enggan melakukan transaksi berulang.
- Keterlambatan Pengiriman / *Delivery Delay* (Bobot: 0,1961): Selisih antara estimasi waktu pengiriman dan waktu kedatangan aktual menyumbang pengaruh hampir 20%. Keterlambatan logistik berdampak langsung pada pengalaman berbelanja dan berpotensi mengikis tingkat kepercayaan pelanggan terhadap platform.
- Frekuensi Pembelian / *Frequency* (Bobot: 0,0094): Berbeda dengan asumsi ritel konvensional, frekuensi transaksi ternyata memiliki pengaruh yang sangat kecil (di bawah 1%) terhadap keputusan *churn* pada dataset ini. Hal ini mengindikasikan bahwa di ekosistem *e-commerce*, pelanggan lebih mementingkan proporsi nilai (harga barang dibanding ongkos kirim) dan kelancaran logistik dibandingkan dengan kuantitas kunjungan atau frekuensi belanja.

Berdasarkan pemetaan bobot tersebut, dapat ditarik implikasi manajerial bahwa strategi retensi pelanggan *e-commerce* tidak cukup jika hanya mengandalkan promosi atau diskon produk. Perusahaan perlu memprioritaskan perbaikan efisiensi operasional logistik, khususnya dalam merumuskan strategi penekanan ongkos kirim dan memastikan ketepatan waktu pengiriman paket kepada pelanggan.

4.5. Implikasi Manajerial untuk Customer Relationship Management (CRM)

Berdasarkan temuan dari pemodelan prediksi *churn* dan sentimen, terdapat beberapa langkah praktis yang bisa diterapkan oleh pihak manajemen *e-commerce* untuk memperbaiki strategi retensi pelanggan, yaitu:

- Fokus Retensi pada Pelanggan *High-Monetary*: Karena variabel total belanja (*Monetary*) memegang pengaruh paling dominan (lebih dari 50%) terhadap *churn*, perusahaan perlu mengurangi strategi sebaran promo secara massal. Anggaran promosi seperti voucher atau *cashback* sebaiknya diprioritaskan untuk pelanggan dengan riwayat transaksi besar yang mulai terlihat pasif. Menyelamatkan satu pelanggan setia yang sering berbelanja bernilai lebih tinggi ketimbang membuang anggaran untuk pelanggan yang hanya berbelanja satu kali (*one-time buyer*).

- b. Penyesuaian Subsidi Ongkos Kirim: Posisi rasio ongkos kirim sebagai faktor penentu kedua menunjukkan bahwa konsumen sangat sensitif terhadap biaya logistik. Tim CRM perlu membuat skema subsidi ongkos kirim yang lebih dinamis, misalnya dengan memberikan potongan tambahan secara otomatis bagi pembeli di zona pengiriman mahal. Hal ini penting agar biaya kirim tetap rasional dan pelanggan tidak membatalkan pesannya saat proses *checkout*.
- c. Kompensasi Keterlambatan Otomatis: Keterlambatan pengiriman merupakan masalah operasional yang paling sering merusak pengalaman belanja. Selain mengevaluasi kinerja ekspedisi, perusahaan dapat memanfaatkan sistem prediksi ini untuk langkah antisipasi. Jika sistem mendeteksi bahwa sebuah paket akan terlambat, aplikasi dapat diatur untuk otomatis mengirimkan permohonan maaf dan kompensasi ringan (seperti tambahan poin) sebelum pelanggan sempat mengajukan komplain atau memberi ulasan buruk.
- d. Pengurangan *Spamming* Notifikasi: Frekuensi kunjungan terbukti sangat kecil pengaruhnya (di bawah 1%) terhadap loyalitas, sehingga perusahaan tidak perlu memaksakan pelanggan untuk membuka aplikasi setiap hari. Terlalu banyak mengirim notifikasi pemasaran justru dapat mengganggu kenyamanan pelanggan dan memicu mereka untuk menghapus aplikasi.

Secara keseluruhan, integrasi analisis sentimen dan algoritma prediksi ini membantu tim CRM beralih dari sekadar bereaksi terhadap komplain menjadi lebih proaktif. Perusahaan kini dapat mendeteksi potensi hilangnya pelanggan dan mengambil tindakan pencegahan sejak dini, jauh sebelum pelanggan tersebut benar-benar beralih ke kompetitor.

5. KESIMPULAN DAN SARAN

Secara keseluruhan, penelitian ini menunjukkan bahwa opini ulasan pelanggan dapat digabungkan dengan data transaksi historis untuk memprediksi churn. *Model Deep Learning* (LSTM) yang digunakan untuk klasifikasi sentimen memberikan hasil yang baik dengan tingkat akurasi 88,44%. Ketika skor sentimen ini digabungkan dengan data perilaku belanja (RFM dan logistik) ke dalam model *Random Forest*, algoritma tersebut mampu mendeteksi pelanggan yang akan churn dengan tingkat *Recall* mencapai 84%. Angka sensitivitas yang tinggi ini menunjukkan bahwa model cukup peka dan dapat diandalkan oleh perusahaan sebagai sistem peringatan dini (*early warning system*) sebelum pelanggan benar-benar berhenti menggunakan layanan.

Dari segi faktor penyebab churn, hasil analisis *Feature Importance* memberikan temuan yang cukup menarik. Keputusan pelanggan *e-commerce* untuk pergi ternyata tidak dipengaruhi oleh seberapa sering mereka berbelanja. Faktor yang paling memicu

pelanggan untuk churn justru didominasi oleh total uang yang dihabiskan (*Monetary*), beban ongkos kirim yang dirasa terlalu mahal, serta pengalaman buruk karena paket yang terlambat datang.

Berangkat dari temuan tersebut, disarankan agar pihak manajemen *e-commerce* mulai mengurangi sebaran promo secara massal. Akan lebih efektif jika fokus perbaikan diarahkan pada operasional logistik, seperti menyesuaikan subsidi ongkos kirim dan mengevaluasi kinerja ekspedisi pengiriman. Promo atau insentif juga sebaiknya difokuskan pada pelanggan dengan riwayat transaksi besar (*high-monetary*) yang sudah mulai jarang berbelanja. Sementara itu, untuk penelitian selanjutnya, sangat disarankan menggunakan teknik resampling seperti SMOTE untuk menyeimbangkan distribusi data (*class imbalance*). Selain itu, mencoba membandingkan performa model ini dengan algoritma boosting lain seperti XGBoost atau LightGBM juga disarankan untuk melihat apakah akurasi prediksinya bisa lebih tinggi.

DAFTAR PUSTAKA

- [1] O. O. Fowowe and R. Agboluaje, "Leveraging Predictive Analytics for Customer Churn: A Cross-Industry Approach in the US Market," *International Journal of Applied Sciences and Radiation Research*, vol. 2, no. 1, 2025, doi: 10.22399/ijasar.20.
- [2] A. Manzoor, M. Atif Qureshi, E. Kidney, and L. Longo, "A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3402092.
- [3] Ang Li, T. Yang, X. Zhan, Y. Shi, and H. Li, "Utilizing Data Science and AI for Customer Churn Prediction in Marketing," *Journal of Theory and Practice of Engineering Science*, vol. 4, no. 05, 2024, doi: 10.53469/jtpes.2024.04(05).10.
- [4] I. Al-Shourbaji, A. Jabbari, S. Rizwan, M. Mehanawi, P. Mansur, and M. Abdalraheem, "An Improved Ant Colony Optimization to Uncover Customer Characteristics for Churn Prediction," *Computational Journal of Mathematical and Statistical Sciences*, vol. 4, no. 1, 2025, doi: 10.21608/cjmss.2024.298501.1059.
- [5] X. Liu, G. Xia, X. Zhang, W. Ma, and C. Yu, "Customer churn prediction model based on hybrid neural networks," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-79603-9.
- [6] S. Saha, C. Saha, M. M. Haque, M. G. R. Alam, and A. Talukder, "ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3349950.
- [7] T. Zdziebko, P. Sulikowski, W. Sałabun, M. Przybyła-Kasperek, and I. Bąk, "Optimizing Customer Retention in the Telecom Industry: A

- Fuzzy-Based Churn Modeling with Usage Data,” *Electronics (Switzerland)*, vol. 13, no. 3, 2024, doi: 10.3390/electronics13030469.
- [8] V. Kumar and W. Reinartz, *Customer Management Relationship Concept, Strategy, and Tools*, vol. 10, no. 3, 2018.
- [9] H. Huang *et al.*, “Improving the explainability of CNN-LSTM-based flood prediction with integrating SHAP technique,” *Ecol. Inform.*, vol. 84, 2024, doi: 10.1016/j.ecoinf.2024.102904.
- [10] S. Li and Z. Shen, “Explainable Customer Churn Prediction Model Based on Deep Learning,” in *ACM International Conference Proceeding Series*, 2024, doi: 10.1145/3654823.3654875.
- [11] M. Imani, M. Joudaki, A. Beikmohammadi, and H. R. Arabnia, “Customer Churn Prediction: A Systematic Review of Recent Advances, Trends, and Challenges in Machine Learning and Deep Learning,” 2025, doi: 10.3390/make7030105.
- [12] M. Z. Alotaibi and M. A. Haq, “Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods,” *Engineering, Technology and Applied Science Research*, vol. 14, no. 3, 2024, doi: 10.48084/etasr.7480.
- [13] S. Beltozar-Clemente, O. Iparraguirre-Villanueva, F. Pucuhuayla-Revatta, J. Zapata-Paulini, and M. Cabanillas-Carbonell, “Predicting customer abandonment in recurrent neural networks using short-term memory,” *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 1, 2024, doi: 10.1016/j.joitmc.2024.100237.
- [14] D. H. Rudd, H. Huo, and G. Xu, “Causal Analysis of Customer Churn Using Deep Learning,” in *2021 International Conference on Digital Society and Intelligent Systems, DSInS 2021*, 2021, doi: 10.1109/DSInS54396.2021.9670561.
- [15] Y. Gaidhani *et al.*, “AI-Driven Predictive Analytics for CRM to Enhance Retention Personalization and Decision-Making,” *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 4, 2025, doi: 10.14569/IJACSA.2025.0160456.
- [16] S. Heristian, M. Napiah, and W. Erawati, “Analisis Sentimen Ulasan Pelanggan Menggunakan Algoritma Naive Bayes pada Aplikasi Gojek,” *Computer Science (CO-SCIENCE)*, vol. 5, no. 1, pp. 35–41, Jan. 2025, doi: 10.31294/coscience.v5i1.7775.
- [17] M. Lamonda, E. Budiarto, H. Rahendra Herlianto, and W. Wiyanto, “PERBANDINGAN METODE SUPPORT VECTOR MACHINE (SVM) DAN MULTINOMIAL NAIVE BAYES (MNB) PADA ANALISIS SENTIMEN PENGGUNA APLIKASI SHOPEE DI GOOGLE PLAY STORE TERKAIT FITUR CASH ON DELIVERY (COD),” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 5, 2025, doi: 10.36040/jati.v9i5.14822.
- [18] I. Al-Shourbaji, A. Jabbari, S. Rizwan, M. Mehanawi, P. Mansur, and M. Abdalraheem, “An Improved Ant Colony Optimization to Uncover Customer Characteristics for Churn Prediction,” *Computational Journal of Mathematical and Statistical Sciences*, vol. 4, no. 1, 2025, doi: 10.21608/cjmss.2024.298501.1059.
- [19] H. Alizadegan, B. Rashidi Malki, A. Radmehr, H. Karimi, and M. A. Ilani, “Comparative study of long short-term memory (LSTM), bidirectional LSTM, and traditional machine learning approaches for energy consumption prediction,” *Energy Exploration and Exploitation*, vol. 43, no. 1, 2025, doi: 10.1177/01445987241269496.
- [20] A. Batta, A. K. Kar, and S. Satpathy, “Cross-Platform Analysis of Seller Performance and Churn for Ecommerce Using Artificial Intelligence,” *Journal of Global Information Management*, vol. 31, no. 1, 2023, doi: 10.4018/JGIM.322439.
- [21] Y. Li and K. Yan, “Prediction of bank credit customers churn based on machine learning and interpretability analysis,” *Data Science in Finance and Economics*, vol. 5, no. 1, 2025, doi: 10.3934/dsfe.2025002