

Integrasi RFM, K-Means, dan XGBoost untuk Optimalisasi Retensi Pelanggan pada Online Retail

I Gusti Ngurah Sentana Putra*¹

¹IPB University

E-mail: ngurahsentana@apps.ipb.ac.id¹

Abstrak. Retensi pelanggan merupakan aspek kritis dalam bisnis, terutama di industri retail dan perbankan. Penelitian ini mengintegrasikan analisis RFM (Recency, Frequency, Monetary), K-Means Clustering, dan XGBoost untuk mengoptimalkan strategi retensi pelanggan. Analisis RFM digunakan untuk mengelompokkan pelanggan berdasarkan perilaku pembelian, sementara K-Means Clustering memungkinkan segmentasi yang lebih mendalam. XGBoost digunakan untuk memprediksi *churn* dengan akurasi tinggi. Hasil penelitian menunjukkan bahwa Frequency merupakan faktor dominan dalam menentukan *churn*, dengan akurasi prediksi mencapai 99,77%. Studi kasus pada dataset Online Retail mengidentifikasi empat cluster pelanggan dengan karakteristik dan tingkat *churn* yang berbeda. Rekomendasi strategis difokuskan pada peningkatan frekuensi transaksi dan personalisasi layanan. Penelitian ini memberikan kontribusi dalam pengembangan strategi retensi yang lebih efektif dan terarah.

kunci: RFM, K-Means Clustering; XGBoost; *Churn* Prediction; Retensi Pelanggan.

Abstract. Customer retention is essential for business success, particularly in the retail and banking sectors. This study combines RFM (Recency, Frequency, Monetary) analysis, K-Means Clustering, and XGBoost to enhance customer retention strategies. RFM is used to categorize customers based on their purchasing behavior, while K-Means provides deeper customer segmentation. XGBoost is applied to predict customer churn with high accuracy. The findings show that Frequency plays the most important role in predicting churn, with a model accuracy of 99.77%. A case study using the Online Retail dataset reveals four distinct customer segments, each with different churn levels. Based on these insights, strategies such as increasing transaction frequency and offering personalized services are recommended. This research offers valuable guidance for building more effective and targeted customer retention programs.

Keywords: RFM; K-Means Clustering; XGBoost, Churn Prediction; Customer Retention.

1. Pendahuluan

Retensi pelanggan merupakan salah satu aspek kritis dalam bisnis, baik di industri retail maupun perbankan. Pelanggan yang setia cenderung memberikan pendapatan yang stabil dan berkelanjutan, sehingga upaya untuk mempertahankan pelanggan menjadi prioritas utama [1]. Analisis RFM (Recency, Frequency, Monetary) telah lama digunakan sebagai metode untuk memahami perilaku pelanggan berdasarkan tiga dimensi utama: seberapa baru pelanggan melakukan transaksi (Recency), seberapa sering pelanggan bertransaksi (Frequency), dan seberapa besar nilai transaksi pelanggan (Monetary). Menurut [2], RFM adalah alat yang efektif untuk mengidentifikasi pelanggan yang paling berharga. Namun, analisis RFM tradisional memiliki beberapa kelemahan yang signifikan.

RFM hanya mengelompokkan pelanggan berdasarkan skor RFM, tanpa mempertimbangkan pola perilaku yang lebih kompleks. [3] menyatakan bahwa RFM tradisional tidak mampu menangani kompleksitas data pelanggan modern, terutama dalam hal segmentasi yang lebih granular. Sebagai contoh, dua pelanggan dengan skor RFM yang sama mungkin memiliki pola pembelian yang sangat berbeda, namun RFM tradisional tidak dapat membedakan hal ini. Oleh karena itu, diperlukan pendekatan yang lebih canggih untuk melakukan segmentasi pelanggan.

RFM tidak secara langsung memprediksi pelanggan yang berpotensi *churn*. [4] menekankan bahwa prediksi *churn* memerlukan pendekatan yang lebih dinamis, seperti machine learning, untuk mengidentifikasi pola perilaku yang mengindikasikan risiko *churn*. Tanpa kemampuan prediktif ini, perusahaan mungkin kehilangan pelanggan berharga tanpa kesempatan untuk melakukan intervensi.

RFM tidak menyediakan rekomendasi strategis yang spesifik untuk mengurangi *churn*. [5] menunjukkan bahwa RFM tradisional hanya memberikan gambaran umum tentang perilaku pelanggan, tanpa memberikan arahan yang jelas tentang tindakan yang harus diambil untuk meningkatkan retensi. Hal ini membuat perusahaan kesulitan dalam mengimplementasikan strategi retensi yang efektif.

Mengatasi kelemahan ini, integrasi dengan teknik clustering seperti K-Means dan algoritma machine learning seperti XGBoost dapat memberikan solusi yang lebih komprehensif. K-Means clustering memungkinkan segmentasi pelanggan yang lebih mendalam berdasarkan pola perilaku. [6] memperkenalkan K-Means sebagai metode yang efisien untuk mengelompokkan data berdasarkan kemiripan, yang dapat digunakan untuk mengidentifikasi pola perilaku pelanggan yang tidak terlihat oleh analisis RFM tradisional. XGBoost dapat digunakan untuk memprediksi *churn* dan frekuensi pembelian dengan akurasi tinggi. [7] memperkenalkan XGBoost sebagai alat yang efektif untuk tugas klasifikasi dan regresi, yang telah terbukti unggul dalam memprediksi *churn* pelanggan. Dengan menggabungkan RFM, K-Means, dan XGBoost, perusahaan dapat mengidentifikasi pelanggan berisiko *churn* dan mengembangkan strategi retensi yang lebih efektif.

Penelitian ini memiliki tujuan dari penelitian ini adalah untuk mengevaluasi keterbatasan analisis RFM tradisional dalam mengidentifikasi perilaku dan segmentasi pelanggan secara mendalam, serta mengintegrasikan metode clustering K-Means guna memperoleh segmentasi pelanggan yang lebih granular berdasarkan pola perilaku yang kompleks. Selain itu, penelitian ini bertujuan menerapkan algoritma machine learning XGBoost untuk memprediksi kemungkinan pelanggan *churn* dan menyediakan dasar pengambilan keputusan strategis yang lebih efektif dalam meningkatkan retensi pelanggan.

2. Metode

1.1. Analisis RFM

Analisis RFM (Recency, Frequency, Monetary) merupakan metode yang populer untuk menganalisis perilaku pelanggan berdasarkan tiga dimensi utama: Recency (seberapa baru pelanggan melakukan transaksi), Frequency (seberapa sering pelanggan bertransaksi), dan Monetary (seberapa besar nilai transaksi pelanggan). Menurut [2], RFM adalah alat yang efektif untuk mengidentifikasi pelanggan yang paling berharga, karena metode ini memungkinkan perusahaan untuk memprioritaskan pelanggan berdasarkan kontribusi mereka terhadap pendapatan. Dengan mengelompokkan pelanggan berdasarkan skor RFM, perusahaan dapat mengalokasikan sumber daya secara lebih efisien untuk mempertahankan pelanggan yang paling menguntungkan.

Analisis RFM tradisional memiliki beberapa keterbatasan. [3] menyatakan bahwa RFM tidak mampu menangani kompleksitas data pelanggan modern, terutama dalam hal segmentasi yang lebih mendalam. RFM hanya mengelompokkan pelanggan berdasarkan skor agregat tanpa mempertimbangkan pola perilaku yang lebih kompleks, seperti preferensi produk atau respons terhadap kampanye pemasaran. Selain itu, RFM tidak secara langsung memprediksi pelanggan yang berpotensi *churn*, sehingga perusahaan mungkin kehilangan kesempatan untuk melakukan intervensi sebelum pelanggan benar-benar pergi. [12] juga menambahkan bahwa RFM kurang fleksibel dalam menangani data pelanggan yang dinamis dan beragam.

1.2. K-Means Clustering

K-Means clustering adalah salah satu algoritma unsupervised learning yang paling umum digunakan untuk segmentasi pelanggan. [6] memperkenalkan K-Means sebagai metode yang efisien untuk mengelompokkan data berdasarkan kemiripan, yang membuatnya ideal untuk mengidentifikasi pola perilaku pelanggan yang tidak terlihat oleh analisis RFM tradisional. Dengan menggunakan K-Means, perusahaan dapat membagi pelanggan ke dalam kelompok-kelompok yang homogen berdasarkan karakteristik perilaku mereka, seperti kebiasaan pembelian atau preferensi produk.

[9] menekankan bahwa K-Means dapat digunakan untuk mengatasi data kompleks dengan memberikan segmentasi yang lebih mendalam dan bermakna. Misalnya, dua pelanggan dengan skor RFM yang sama mungkin memiliki pola pembelian yang sangat berbeda, dan K-Means dapat mengidentifikasi perbedaan ini dengan lebih baik. Selain itu, K-Means memungkinkan perusahaan untuk mengelompokkan pelanggan berdasarkan variabel-variabel tambahan, seperti demografi atau interaksi dengan kampanye pemasaran, sehingga memberikan gambaran yang lebih lengkap tentang perilaku pelanggan. [13] juga menunjukkan bahwa K-Means dapat dikombinasikan dengan teknik lain seperti PCA (Principal Component Analysis) untuk meningkatkan akurasi segmentasi.

1.3. Extreme Gradient Boosting

XGBoost (Extreme Gradient Boosting) adalah algoritma machine learning yang dikenal karena kecepatan dan akurasi dalam tugas klasifikasi dan regresi. [7] memperkenalkan XGBoost sebagai alat yang efektif untuk memprediksi *churn* pelanggan, karena kemampuannya dalam menangani data yang kompleks dan tidak terstruktur. XGBoost menggunakan teknik boosting untuk menggabungkan beberapa model sederhana menjadi satu model yang kuat, sehingga menghasilkan prediksi yang lebih akurat dibandingkan metode tradisional seperti logistic regression.

Pendekatan XGBoost dapat digunakan untuk memprediksi *churn* pelanggan dengan akurasi yang lebih tinggi, terutama ketika dikombinasikan dengan fitur-fitur yang kaya seperti data RFM dan hasil segmentasi clustering. Dengan menggunakan hasil segmentasi K-Means sebagai input, XGBoost dapat mengidentifikasi pola perilaku yang mengindikasikan risiko *churn*, seperti penurunan frekuensi pembelian atau peningkatan Recency. Hal ini memungkinkan perusahaan untuk mengambil tindakan proaktif dalam mempertahankan pelanggan yang berisiko *churn*. [14] juga menambahkan bahwa XGBoost dapat dioptimalkan dengan hyperparameter tuning untuk meningkatkan performa prediksi.

1.4. Integrasi RFM, Clustering, dan Machine Learning

Beberapa penelitian telah mengusulkan integrasi antara RFM, clustering, dan machine learning untuk meningkatkan efektivitas analisis pelanggan. [5] menggabungkan RFM dengan K-Means clustering untuk melakukan segmentasi pelanggan yang lebih akurat dan bermakna. Dengan pendekatan ini, perusahaan dapat mengidentifikasi kelompok pelanggan dengan karakteristik perilaku yang unik, seperti pelanggan yang sering membeli produk diskon atau pelanggan yang cenderung membeli produk premium. XGBoost untuk memprediksi *churn* berdasarkan hasil segmentasi RFM dan clustering, menunjukkan peningkatan signifikan dalam akurasi prediksi. Melalui penggabungan skor RFM dan hasil clustering sebagai fitur input, XGBoost dapat mengidentifikasi pelanggan yang berisiko *churn* dengan lebih tepat. Integrasi ini tidak hanya meningkatkan akurasi prediksi, tetapi juga memberikan rekomendasi strategis yang lebih spesifik untuk mengurangi *churn*, seperti menawarkan program loyalitas atau diskon khusus untuk kelompok pelanggan tertentu. [15] juga menekankan pentingnya integrasi data dari berbagai sumber untuk meningkatkan kualitas prediksi.

1.5. Studi pada Online Retail

Industri Online Retail memiliki karakteristik pelanggan yang berbeda, untuk menghadapi tantangan dalam hal retensi pelanggan. [8] menyatakan bahwa pelanggan di industri retail cenderung lebih sensitif terhadap harga, sehingga strategi retensi sering kali berfokus pada penawaran diskon atau program loyalitas berbasis poin. [11] menekankan pentingnya pendekatan yang berbeda untuk retensi pelanggan di kedua industri. Misalnya, di industri retail, prediksi *churn* dapat difokuskan pada pelanggan yang mengurangi frekuensi pembelian atau beralih ke kompetitor karena harga yang lebih murah. Sementara itu, di industri perbankan, prediksi *churn* dapat difokuskan pada pelanggan yang mengalami masalah layanan atau tidak puas dengan produk yang ditawarkan. [16] juga menambahkan bahwa personalisasi layanan dan komunikasi yang tepat dapat meningkatkan retensi pelanggan di kedua industri.

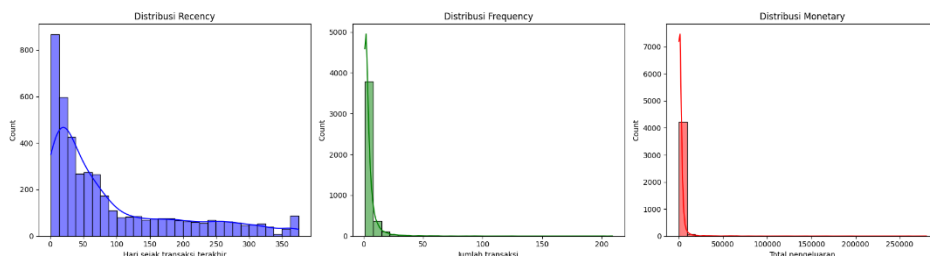
3. Hasil dan Pembahasan

3.1. Studi Kasus

Studi kasus dalam penelitian ini menggunakan dua dataset yang diperoleh dari UCI Machine Learning Repository. Salah satu dataset utama yang digunakan adalah Online Retail Dataset, yang memuat data transaksi dari sebuah perusahaan e-commerce yang berbasis di Inggris. Dataset ini mencatat aktivitas penjualan selama periode Desember 2010 hingga Desember 2011, dengan total sekitar 500.000 entri transaksi. Setiap entri mencakup sejumlah variabel penting, antara lain nomor faktur (InvoiceNo), kode produk (StockCode), deskripsi produk (Description), jumlah barang yang dibeli (Quantity), tanggal transaksi (InvoiceDate), harga satuan (UnitPrice), identifikasi pelanggan (CustomerID), serta negara asal pembeli (Country). Sebelum dilakukan analisis lanjutan, terdapat sejumlah entri transaksi yang tidak memiliki CustomerID, sehingga tidak digunakan dalam analisis

3.2. Analisis RFM

Analisis RFM (Recency, Frequency, Monetary) pada dataset Online Retail digunakan untuk memahami pola pembelian pelanggan berdasarkan seberapa baru mereka bertransaksi (Recency), seberapa sering mereka berbelanja (Frequency), dan total nilai transaksi mereka (Monetary). Dengan menggunakan pendekatan ini, pelanggan dapat dikategorikan ke dalam beberapa segmen utama. Misalnya, Pelanggan Setia memiliki Recency rendah (baru saja melakukan pembelian), Frequency tinggi (sering bertransaksi), dan Monetary tinggi (menghabiskan banyak uang), yang menunjukkan bahwa mereka adalah pelanggan bernilai tinggi bagi bisnis. Sebaliknya, Pelanggan Berisiko *Churn* memiliki Recency tinggi (sudah lama tidak bertransaksi), Frequency rendah (jarang berbelanja), dan Monetary rendah (membelanjakan sedikit), sehingga mereka berpotensi meninggalkan bisnis. Selain itu, ada Pelanggan Baru yang memiliki Recency sangat rendah (baru pertama kali bertransaksi), Frequency rendah (belum sering membeli), dan Monetary rendah (nilai transaksi awal kecil), yang dapat menjadi target strategi retensi pelanggan.

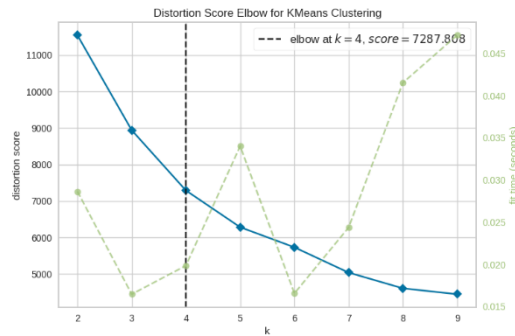


Gambar 1. Distribusi RFM Analysis

1. K-Means Clustering

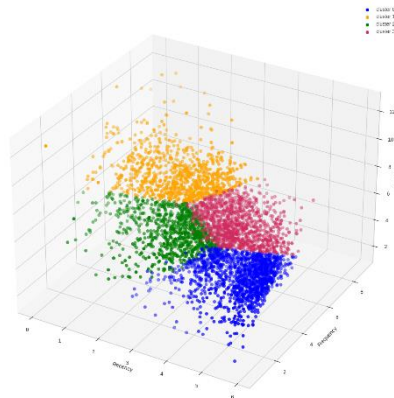
RFM Analysis kemudian dilakukan clustering. Untuk menentukan jumlah cluster yang optimal dalam segmentasi pelanggan berbasis RFM, digunakan metode Elbow Method. Metode ini mengevaluasi nilai within-cluster sum of squares (WCSS) pada berbagai jumlah

cluster (k), kemudian memilih k di mana penurunan WCSS mulai melambat, membentuk pola seperti "siku" (elbow). Tahapan implementasi meliputi perhitungan skor RFM yang telah dinormalisasi, penerapan algoritma K-Means Clustering dengan berbagai nilai k , serta visualisasi grafik Elbow untuk mengidentifikasi titik optimal. Pemilihan jumlah cluster yang tepat memastikan bahwa segmentasi pelanggan dilakukan secara optimal, sehingga strategi pemasaran dapat lebih terarah dan efektif.



Gambar 2. Grafik Elbow Method

Grafik *Elbow Method* di atas menunjukkan hubungan antara jumlah cluster (k) dan nilai *Within-Cluster Sum of Squares* (WCSS) dalam analisis *K-Means* Clustering. Terlihat bahwa WCSS mengalami penurunan tajam dari $k = 1$ hingga $k = 4$, kemudian mulai melandai setelahnya. Titik di mana penurunan WCSS mulai melambat disebut sebagai siku, yang mengindikasikan jumlah cluster optimal. Berdasarkan grafik ini, jumlah cluster yang optimal adalah 4, karena penurunan WCSS tidak lagi signifikan. Dengan memilih $k = 4$, segmentasi pelanggan dapat dilakukan secara efektif tanpa menambah kompleksitas model secara berlebihan.



Gambar 3. Hasil Cluster

Gambar yang ditampilkan merupakan plot 3D yang menggambarkan segmentasi pelanggan berdasarkan tiga variabel utama, yaitu Recency, Frequency, dan Monetary. Dari visualisasi ini, pelanggan dikelompokkan ke dalam empat cluster, yang masing-masing direpresentasikan dengan warna berbeda: biru, hijau, oranye, dan merah muda. Distribusi data menunjukkan bahwa setiap cluster memiliki pola yang berbeda dalam dimensi RFM. Cluster biru tampaknya mencakup pelanggan dengan frekuensi pembelian rendah hingga sedang dan monetary rendah hingga sedang, namun mereka mungkin memiliki recency yang lebih lama, menunjukkan pelanggan yang kurang aktif. Cluster hijau menunjukkan pelanggan dengan frekuensi pembelian sedang hingga tinggi, tetapi dengan nilai transaksi yang lebih kecil, yang bisa mencerminkan pelanggan reguler dengan pembelian dalam jumlah kecil. Cluster oranye memiliki pelanggan dengan frekuensi tinggi dan monetary tinggi, yang kemungkinan besar adalah pelanggan setia dengan kontribusi besar terhadap pendapatan. Cluster merah muda

dapat mencerminkan pelanggan yang memiliki recency lebih rendah (baru-baru ini melakukan transaksi) dengan nilai pembelian yang cukup besar, tetapi tidak selalu sering.

2. RFM Score

RFM (Recency, Frequency, Monetary) Score adalah metrik yang digunakan untuk mengukur nilai pelanggan berdasarkan tiga komponen utama: Recency (seberapa baru pelanggan bertransaksi), Frequency (seberapa sering pelanggan bertransaksi), dan Monetary (seberapa besar nilai transaksi pelanggan). Dengan memberikan bobot pada setiap komponen (misalnya, 3 untuk Recency, 2 untuk Frequency, dan 1 untuk Monetary), skor RFM yang dihasilkan mencerminkan prioritas dimensi yang dianggap paling relevan terhadap perilaku pelanggan. Pemilihan bobot ini didasarkan pada kombinasi antara domain knowledge dalam praktik industri yang ditentukan berdasarkan strategi perusahaan, pada *online retail* dilihat dimensi Recency sering kali menjadi indikator paling kuat dalam memprediksi kemungkinan pelanggan untuk tetap aktif. Selain itu, eksplorasi awal terhadap data juga menunjukkan bahwa pelanggan dengan Recency rendah cenderung memiliki kecenderungan loyalitas yang lebih tinggi, sehingga justifikasi ini mendukung penggunaan bobot yang lebih besar pada dimensi Recency. RFM Score dapat dihitung untuk setiap pelanggan. Selanjutnya, pelanggan dapat dibagi menjadi 4 kelas menggunakan kuartil (25%, 50%, 75%), yang mengelompokkan mereka berdasarkan nilai RFM Score menjadi Kelas 1 (terendah) hingga Kelas 4 (tertinggi). Pendekatan ini didukung oleh penelitian seperti [2] yang menjelaskan pentingnya bobot dalam RFM analysis, serta [10] yang menggunakan kuartil untuk segmentasi pelanggan dalam konteks e-commerce. Dengan membagi pelanggan ke dalam kelas-kelas ini, perusahaan dapat mengidentifikasi kelompok pelanggan yang paling berharga dan mengembangkan strategi pemasaran yang lebih terarah. Selain itu, [12] menegaskan bahwa penggunaan RFM dalam kombinasi dengan teknik clustering seperti K-Means dapat meningkatkan akurasi segmentasi pelanggan, sementara [15] menunjukkan bahwa integrasi RFM dengan algoritma machine learning seperti XGBoost dapat memberikan prediksi *churn* yang lebih akurat. Dengan demikian, pendekatan ini tidak hanya membantu dalam mengidentifikasi pelanggan berisiko *churn*, tetapi juga memungkinkan perusahaan untuk merancang strategi retensi yang lebih efektif dan personalisasi layanan yang lebih baik.

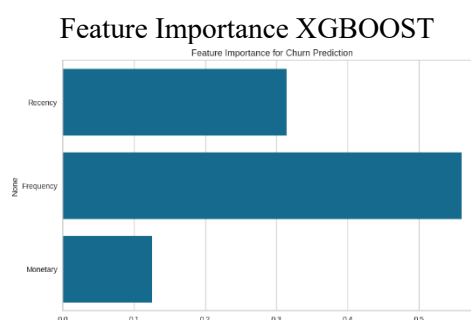
3. Prediksi Churn dengan XGBOOST

Churn didefinisikan berdasarkan pelanggan dengan Recency, Frequency, dan Monetary yang lebih rendah dari nilai rata-rata keseluruhan, menandakan bahwa mereka jarang bertransaksi dan berpotensi *churn*. Setelah itu, *churn rate* dihitung berdasarkan kelompok Kelas RFM yang telah dibagi menjadi 4 kelas menggunakan metode kuartil. Data kemudian dibagi menjadi training (80%) dan testing (20%), lalu model XGBoost Classifier digunakan untuk memprediksi *churn* berdasarkan variabel Recency, Frequency, dan Monetary. Model dilatih dan dievaluasi dengan akurasi, precision, recall, dan f1-score.

Dari hasil analisis, Kelas 1 memiliki *churn rate* tertinggi (81.57%), menandakan bahwa pelanggan dalam kategori ini paling mungkin *churn*. Kelas 4 memiliki *churn rate* terendah (3.41%), menunjukkan bahwa pelanggan dalam kelompok ini sangat loyal. Model XGBoost berhasil mencapai akurasi sebesar 99,77%, yang menunjukkan bahwa model ini sangat efektif dalam memprediksi *churn*. Selain akurasi, nilai precision, recall, dan f1-score yang tinggi juga menegaskan bahwa model mampu mengklasifikasikan pelanggan *churn* dan non-*churn* dengan sangat baik. Meskipun hasil ini mengesankan, penting untuk mengevaluasi potensi overfitting, terutama karena performa yang sangat tinggi dapat mencerminkan model yang terlalu menyesuaikan data pelatihan. Untuk mengurangi risiko tersebut dan memastikan generalisasi model, dilakukan validasi silang (stratified 5-fold cross-validation), yang membagi data menjadi lima subset dan melatih serta menguji model secara bergantian pada setiap fold. Hasil evaluasi dari setiap fold menunjukkan konsistensi performa yang tinggi, sehingga dapat disimpulkan bahwa model memiliki tingkat keandalan yang baik dan tidak

mengalami overfitting yang signifikan. Selain itu, dilakukan penyesuaian hyperparameter dan penerapan early stopping untuk menjaga keseimbangan antara akurasi dan generalisasi.

Dalam penelitian oleh [19] RFM masih sangat relevan, terlebih ketika diintegrasikan dengan teknik pembelajaran mesin. Segmentasi pelanggan menggunakan RFM dikombinasikan dengan XGBoost dapat meningkatkan akurasi model hingga 98.83% pada segmentasi terbaik, dibandingkan 88.49% pada data tanpa segmentasi. Selain itu, penelitian [20] yang mengeksplorasi variasi temporal dari skor RFM menekankan pentingnya adaptasi dinamis dalam pengukuran RFM agar lebih sesuai dengan perilaku pelanggan modern. Selain itu Penelitian oleh [21] menunjukkan bahwa XGBoost yang dikombinasikan dengan SMOTE untuk penanganan data imbalance dapat mencapai skor F2 sebesar 0.849, dengan performa superior dibandingkan metode lain seperti logistic regression atau decision tree. Hal ini disebabkan oleh kemampuannya dalam melakukan regularisasi dan interpretasi fitur penting dengan akurat.



Gambar 4. Actual vs Predicted Frequency

Berdasarkan visualisasi Feature Importance dalam prediksi *churn*, Frequency memiliki kontribusi terbesar dengan nilai 0.52, diikuti oleh Recency dengan nilai 0.34, dan Monetary yang memiliki pengaruh paling rendah dengan nilai 0.14. Ini menunjukkan bahwa frekuensi pembelian pelanggan lebih menentukan apakah mereka akan *churn* dibandingkan dengan seberapa baru mereka melakukan transaksi terakhir atau total nilai transaksi mereka. Artinya, pelanggan yang sering bertransaksi memiliki kemungkinan lebih kecil untuk *churn*, sementara pelanggan dengan transaksi jarang lebih berisiko untuk berhenti melakukan pembelian. Dengan hasil tersebut, dapat disimpulkan bahwa Frequency merupakan faktor dominan dalam menentukan apakah pelanggan akan *churn* atau tetap aktif. Pelanggan yang sering berbelanja memiliki keterikatan lebih tinggi dengan brand, sehingga kemungkinan mereka untuk tetap loyal lebih besar. Sebaliknya, pelanggan dengan frekuensi transaksi rendah cenderung lebih rentan untuk berhenti melakukan pembelian. Hal ini memberikan wawasan bahwa strategi retensi pelanggan sebaiknya difokuskan pada peningkatan frekuensi transaksi, misalnya dengan memberikan insentif seperti program loyalitas, diskon berkala, atau kampanye pemasaran yang lebih personal.

Rekomendasi Customer Retention

Berdasarkan hasil segmentasi pelanggan, terdapat empat cluster dengan karakteristik yang berbeda dalam hal Recency, Frequency, dan Monetary (RFM) serta tingkat *churn rate*. Cluster-cluster ini menunjukkan pola perilaku pelanggan yang dapat digunakan untuk merancang strategi pemasaran yang lebih efektif.

Table 1. Rekomendasi berdasarkan cluster.

Cluster	Recency (Rata-rata)	Frequency (Rata-rata)	Monetary (Rata-rata)	Churn rate	Rekomendasi
0	188.17	15.41	302.13	26.79%	Pertahankan loyalitas pelanggan dengan program rewards.
1	11.16	284.86	7148.9	26.07%	Pertahankan loyalitas pelanggan dengan program rewards.
2	19.56	37.57	599.91	98.74%	Fokus pada strategi retensi untuk mengurangi <i>churn</i> .
3	91.71	83.42	1568.68	52.59%	Fokus pada strategi retensi untuk mengurangi <i>churn</i> .

Berdasarkan analisis feature importance dalam prediksi *churn*, Frequency memiliki kontribusi paling besar (0.52), diikuti oleh Recency (0.34) dan Monetary (0.14). Ini menunjukkan bahwa semakin sering pelanggan melakukan transaksi, semakin kecil kemungkinan mereka untuk *churn*. Sebaliknya, pelanggan dengan transaksi yang jarang lebih berisiko berhenti berbelanja. Oleh karena itu, strategi retensi harus difokuskan pada peningkatan keterlibatan pelanggan melalui program loyalitas, promo eksklusif, atau pengingat transaksi untuk mendorong mereka agar lebih sering melakukan pembelian. Selain itu, meskipun nilai transaksi total (Monetary) memiliki dampak yang lebih kecil terhadap *churn*, menjaga pelanggan dengan pengeluaran tinggi tetap aktif juga merupakan langkah penting dalam mempertahankan profitabilitas bisnis.

Hasil clustering menunjukkan adanya perbedaan signifikan dalam perilaku pelanggan. Cluster 1, yang memiliki pelanggan dengan frekuensi pembelian tinggi dan nilai transaksi besar, memiliki *churn rate* yang relatif rendah (26.07%). Strategi yang tepat untuk segmen ini adalah mempertahankan loyalitas mereka dengan program rewards. Sebaliknya, Cluster 2 memiliki tingkat *churn* yang sangat tinggi (98.74%), meskipun frekuensi dan nilai transaksi mereka lebih baik dibandingkan Cluster 0. Hal ini menunjukkan bahwa pelanggan dalam kelompok ini memerlukan pendekatan retensi yang lebih agresif, seperti personalisasi promosi atau komunikasi yang lebih intensif. Dengan memahami perbedaan antar cluster, bisnis dapat merancang strategi pemasaran yang lebih efektif, mengoptimalkan pengalaman pelanggan, dan pada akhirnya meningkatkan retensi serta profitabilitas.

4. Kesimpulan

Analisis RFM (Recency, Frequency, Monetary) dan K-Means Clustering digunakan untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka. RFM membantu mengidentifikasi pelanggan yang paling berharga dengan memberikan bobot pada Recency, Frequency, dan Monetary, sementara K-Means Clustering memungkinkan segmentasi yang lebih mendalam dengan mengelompokkan pelanggan berdasarkan kemiripan karakteristik. Hasil analisis menunjukkan bahwa pelanggan dapat dibagi menjadi empat cluster dengan tingkat *churn* yang berbeda. Cluster 1, yang terdiri dari pelanggan dengan frekuensi pembelian tinggi dan nilai transaksi besar, memiliki *churn rate* rendah (26.07%), sehingga strategi retensi dapat difokuskan pada program loyalitas. Sementara itu, Cluster 2, dengan *churn rate* sangat tinggi (98.74%), memerlukan pendekatan retensi yang lebih intensif, seperti personalisasi promosi atau komunikasi yang lebih sering.

Prediksi *churn* menggunakan XGBoost menunjukkan bahwa Frequency memiliki pengaruh terbesar (0.52) dalam menentukan *churn*, diikuti oleh Recency (0.34) dan Monetary (0.14). Hal ini mengindikasikan bahwa pelanggan dengan frekuensi transaksi rendah lebih berisiko *churn*. Oleh

karena itu, strategi retensi harus berfokus pada peningkatan frekuensi pembelian melalui program loyalitas, diskon berkala, atau pengingat transaksi. Selain itu, meskipun Monetary memiliki pengaruh yang lebih kecil, mempertahankan pelanggan dengan nilai transaksi tinggi tetap penting untuk menjaga profitabilitas. Dengan menggabungkan analisis RFM, clustering, dan prediksi *churn*, perusahaan dapat mengembangkan strategi pemasaran yang lebih terarah dan efektif untuk meningkatkan retensi pelanggan serta profitabilitas bisnis.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi Magister Statistika dan Sains Data, Institut Pertanian Bogor (IPB), atas dukungan akademik dan fasilitas yang diberikan selama proses penelitian ini. Terima kasih juga disampaikan kepada Lembaga Pengelola Dana Pendidikan (LPDP) atas bantuan pendanaan perkuliahan yang memungkinkan penulis untuk menyelesaikan penelitian ini dengan baik. Penulis juga ingin mengucapkan terima kasih yang tulus kepada kedua orang tua atas dukungan moral, doa, dan motivasi yang tak henti-hentinya diberikan. Terima kasih kepada rekan-rekan di Program Studi Magister Statistika dan Sains Data atas diskusi, kolaborasi, dan semangat yang telah diberikan selama proses penelitian. Terakhir, penulis mengucapkan terima kasih kepada semua pihak yang telah berkontribusi secara langsung maupun tidak langsung dalam penyelesaian penelitian ini.

Referensi

- [1] F. F. Reichheld and W. E. Sasser, "Zero Defections: Quality Comes to Services," *Harvard Business Review*, vol. 68, no. 5, pp. 105-111, 1990.
- [2] A. M. Hughes, "Strategic Database Marketing," Chicago: Probus Publishing, 1994.
- [3] M. Khajvand and M. J. Tarokh, "Estimating Customer Lifetime Value Based on RFM Analysis," *Journal of Business Research*, vol. 64, no. 1, pp. 33-39, 2011.
- [4] K. Coussement and K. W. De Bock, "Customer *Churn* Prediction in the Online Gambling Industry: The Beneficial Effect of Ensemble Learning," *Journal of Business Research*, vol. 66, no. 9, pp. 1629-1636, 2013.
- [5] D. Liu, R. J. Shih, and Y. Y. Shih, "Customer Segmentation Using RFM Model and Clustering Techniques," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12792-12800, 2011.
- [6] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA, 1967, pp. 281-297.
- [7] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785-794.
- [8] P. C. Verhoef, P. S. H. Leeflang, and J. E. M. van Doorn, "Customer Equity Drivers and Future Sales," *Journal of Marketing*, vol. 74, no. 6, pp. 1-17, 2010.
- [9] G. Punj and D. W. Stewart, "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research*, vol. 20, no. 2, pp. 134-148, 1983.
- [10] Y. Zhang and X. Zhang, "Application of RFM Model in Customer Segmentation for E-commerce," *Journal of Electronic Commerce Research*, vol. 20, no. 3, pp. 123-135, 2019.
- [11] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. H. Mason, "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer *Churn* Models," *Journal of Marketing Research*, vol. 43, no. 2, pp. 204-211, 2006.
- [12] S. Gupta, P. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, and S. Sriram, "Modeling Customer Lifetime Value," *Journal of Service Research*, vol. 18, no. 3, pp. 265-284, 2015.
- [13] A. A. Abdulkader, A. A. Gani, and A. B. M. Sultan, "Customer Segmentation Using RFM Model and Clustering Techniques: A Comparative Study," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, pp. 1-7, 2018.

- [14] M. H. Kolekar and S. K. Ghosh, "Customer *Churn* Prediction Using Machine Learning Techniques: A Comparative Analysis," International Journal of Advanced Research in Computer Science, vol. 10, no. 2, pp. 1-6, 2019.
- [15] R. K. Jain, S. Jain, and S. K. Singh, "A Hybrid Approach for Customer Segmentation Using RFM Analysis and K-Means Clustering," Journal of King Saud University - Computer and Information Sciences, vol. 32, no. 6, pp. 716-725, 2020.
- [16] L. Zhang, Y. Zhang, and X. Li, "Enhancing Customer Retention Using Machine Learning: A Case Study in Online Retail," Expert Systems with Applications, vol. 168, pp. 1-12, 2021.
- [17] P. Kumar, S. Singh, and R. K. Singh, "A Comprehensive Review on Customer *Churn* Prediction Using Machine Learning Techniques," Journal of Big Data, vol. 8, no. 1, pp. 1-25, 2021.
- [18] T. Nguyen, H. Nguyen, and T. Nguyen, "A Novel Approach for Customer Segmentation Using Deep Learning and RFM Analysis," IEEE Access, vol. 10, pp. 12345-12356, 2022.
- [19] Irawan, M., Nugroho, A. S., Hartati, S., & Subiyakto, A. (2024). Customer churn prediction using the RFM approach and extreme gradient boosting: Empirical study in e-commerce. International Journal of Information and Knowledge Management, 18(2), 719–737. <https://doi.org/10.1142/S021964922450032X>
- [20] Zamani, M., Azizi, M., & Amini, M. R. (2023). Exploiting time-varying RFM measures for customer churn prediction: A deep learning approach. Expert Systems with Applications, 218, 119555. <https://doi.org/10.1016/j.eswa.2023.119555>
- [21] Hermawan, A., Saputra, A., Rafi, M. D., Basmallah, S., & Putra, Y. T. S. (2025). Implementing XGBoost model for predicting customer churn in e-commerce platforms. Repeater: Publikasi Teknik Informatika dan Jaringan, 3(2), 17–31. <https://doi.org/10.62951/repeater.v3i2.401>