

Implementasi Algoritma XGBoost untuk Klasifikasi *Customer Segment* dalam Mengidentifikasi Nilai dan Loyalitas Pelanggan

Nabil Abdul Majid^{1*}, Muhammad Rifky Hafan², Trafrika Nining Erina³, Savina Dyah Arwiana⁴, Aprilisa Arum Sari⁵

¹Teknik Informatika/Fakultas

Ilmu Komputer

Universitas Duta Bangsa

Surakarta

^{1*}230103173@mhs.udb.ac.id (penulis korespondensi)

²Teknik Informatika/Fakultas

Ilmu Komputer

Universitas Duta Bangsa

Surakarta

^{2*}230103169@mhs.udb.ac.id

³Teknik Informatika/Fakultas

Ilmu Komputer

Universitas Duta Bangsa

Surakarta

^{3*}trafrikaerina@gmail.com

⁴Teknik Informatika/Fakultas

Ilmu Komputer

Universitas Duta Bangsa

Surakarta

^{4*}savinadyah27@gmail.com

⁵Teknik Informatika/Fakultas

Ilmu Komputer

Universitas Duta Bangsa

Surakarta

^{5*}aprilisa_arumsari@udb.ac.id

Abstrak— *Customer Segment* sangat diperlukan agar perusahaan dapat memahami sifat pelanggan dan merancang strategi pemasaran yang lebih efisien. Studi ini bertujuan menerapkan algoritma Extreme Gradient Boosting (XGBoost) untuk mengelompokkan pelanggan ke dalam empat segmen, yaitu High_Value, Loyal, Regular, dan Occasional, dengan dasar karakteristik demografis serta perilaku pelanggan. Dataset yang digunakan merupakan Customer Segment Ritel Online dari Kaggle, dengan langkah penelitian mencakup pra-pemrosesan data, transformasi data, pemodelan dengan XGBoost, dan penilaian model. Temuan penelitian mengindikasikan bahwa model XGBoost dapat meraih akurasi hingga 75%. Analisis mendalam mengindikasikan bahwa *purchase frequency* dan *months active* adalah variabel yang paling berpengaruh dalam klasifikasi. Oleh karena itu, XGBoost sangat efisien dalam mengklasifikasikan segmen pelanggan dan dapat membantu dalam menentukan nilai serta kesetiaan pelanggan sebagai landasan untuk keputusan pemasaran.

Kata kunci— customer segmentation, XGBoost, klasifikasi, loyalitas pelanggan

Abstract— Customer segmentation is crucial for assisting companies in comprehending customer traits and creating more focused marketing approaches. This research seeks to apply the Extreme Gradient Boosting (XGBoost) algorithm to categorize customers into four groups: High Value, Loyal, Regular, and Occasional, utilizing demographic and behavioral traits. The dataset utilized was the Online Retail Customer Segmentation dataset sourced from Kaggle, and the research phases encompassed data preprocessing, data transformation, XGBoost modeling, and model assessment. The findings indicate that the XGBoost model reached a classification accuracy of 75% for customer segments. Additional analysis reveals that *purchase frequency* and *months active* are the key factors in the classification process. The results indicate that XGBoost is useful for segmenting customers and can aid in recognizing customer value and loyalty to inform marketing decisions.

Keywords— customer segmentation, XGBoost, classification, customer loyalty

I. PENDAHULUAN

Kemajuan teknologi informasi dan digitalisasi telah mendorong perusahaan untuk memanfaatkan data konsumen sebagai landasan dalam pengambilan keputusan bisnis. Salah satu metode yang sering digunakan adalah segmentasi pelanggan, yaitu proses mengkategorikan konsumen berdasarkan sifat, kebutuhan, dan perilaku yang mirip. Output

dari proses tersebut adalah beberapa Segment Pelanggan yang dapat digunakan perusahaan untuk memahami perbedaan karakteristik konsumen sehingga strategi pemasaran, peningkatan kualitas layanan, dan pengelolaan hubungan pelanggan bisa dilakukan dengan lebih efisien dan tepat sasaran.[1].

Dengan menggunakan Segmen Pelanggan, perusahaan dapat mengenali pelanggan yang

memiliki nilai tinggi, tingkat kesetiaan yang baik, serta pelanggan yang mungkin membutuhkan perhatian khusus untuk meningkatkan retensi pelanggan.

Seiring bertambahnya volume serta kompleksitas data pelanggan, penerapan machine learning menjadi salah satu opsi yang sering digunakan dalam analisis dan pengklasifikasian data. Salah satu algoritma yang dikenal dengan kinerja yang sangat baik dalam tugas klasifikasi adalah Extreme Gradient Boosting (XGBoost).[2]. Algoritma ini adalah pengembangan dari metode *Gradient Boosting* yang dibuat untuk meningkatkan akurasi prediksi, mengurangi kemungkinan *overfitting*, dan mempercepat proses komputasi. Berbagai penelitian telah menunjukkan keberhasilan penerapan XGBoost pada beragam bidang. Abdurrahman *et al.* menerapkan XGBoost pada klasifikasi penyakit diabetes dan menunjukkan performa klasifikasi yang baik setelah dilakukan optimasi parameter [3]. Harahap *et al.* menggunakan XGBoost pada klasifikasi penyakit paru-paru dan memperoleh hasil yang memuaskan dibandingkan metode pembandingan [4]. Selain itu, Dachi dan Sitompul menerapkan XGBoost pada klasifikasi keputusan kredit dan menunjukkan kemampuan algoritma dalam menangani permasalahan klasifikasi pada sektor keuangan [5]. Penelitian lain yang dilakukan oleh Yulianti *et al.* menunjukkan bahwa penerapan *hyperparameter tuning* pada XGBoost mampu meningkatkan performa klasifikasi nasabah kartu kredit [6]. Sementara itu, Beni *et al.* memanfaatkan XGBoost untuk memprediksi *customer churn* dan menunjukkan bahwa algoritma tersebut efektif dalam menganalisis perilaku pelanggan [7]. Di sisi lain, penelitian mengenai klasifikasi *Customer Segment* juga telah banyak dilakukan untuk membantu perusahaan mengidentifikasi karakteristik pelanggan sebagai dasar dalam penyusunan strategi pemasaran yang lebih tepat sasaran [8].

Meskipun demikian, sebagian besar penelitian terdahulu masih berfokus pada penerapan XGBoost di bidang kesehatan, keuangan, maupun prediksi *customer churn*. Penelitian yang secara khusus mengimplementasikan XGBoost untuk klasifikasi

customer segment dalam mengidentifikasi nilai dan loyalitas pelanggan masih relatif terbatas. Di samping itu, mayoritas studi hanya menitikberatkan pada pengukuran kinerja model tanpa menganalisis lebih dalam faktor-faktor yang berpengaruh terhadap pembentukan *Customer Segment*. Sehingga, studi ini menerapkan algoritma XGBoost untuk mengelompokkan pelanggan ke dalam empat kategori *Customer Segment*, yaitu *High_Value*, *Loyal*, *Regular*, dan *Occasional* [9]. Di samping menciptakan model klasifikasi, studi ini juga melakukan analisis *feature importance* untuk menentukan faktor-faktor yang paling memengaruhi nilai dan kesetiaan pelanggan [10]. Hasil penelitian diharapkan dapat menjadi dasar dalam penyusunan strategi pemasaran dan pengelolaan pelanggan yang lebih efektif sehingga mampu meningkatkan nilai bisnis perusahaan.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode penelitian kuantitatif dengan pendekatan *data mining*. Penelitian ini bertujuan untuk mengklasifikasikan *Customer Segment* menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) berdasarkan karakteristik dan perilaku pelanggan yang terdapat pada dataset *Online Retail Customer Segmentation* yang dipublikasikan oleh Rohit Kumar melalui platform Kaggle[9]. Dataset tersebut memuat atribut demografis, transaksi, dan perilaku pelanggan yang digunakan untuk mengklasifikasikan pelanggan ke dalam empat kategori, yaitu *High_Value*, *Loyal*, *Regular*, dan *Occasional*. Penggunaan data konsumen dalam proses *customer segmentation* telah banyak dimanfaatkan untuk membantu perusahaan mengenali karakteristik pelanggan, meningkatkan efisiensi strategi pemasaran, serta mendukung keputusan bisnis yang lebih tepat sasaran[11].

XGBoost adalah penyempurnaan dari metode Gradient Boosting Decision Tree (GBDT) yang membangun model secara bertahap melalui serangkaian pohon keputusan untuk memperbaiki kesalahan prediksi dari iterasi sebelumnya. XGBoost memanfaatkan fungsi objektif yang menyatukan loss function dan regularisasi untuk

meningkatkan akurasi prediksi, mengurangi overfitting, dan mempercepat komputasi pada data besar. Persamaan fungsi objektif XGBoost ditunjukkan pada Persamaan (1).

$$\text{Obj}^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^t \Omega(f_k)$$

dengan $\Omega(f)$ sebagai fungsi regularisasi yang mengontrol kompleksitas model sehingga menghasilkan prediksi yang lebih optimal[12].

Langkah-langkah penelitian yang dilaksanakan adalah sebagai berikut:

1. Seleksi Data (Data Selection)

Data penelitian didapat dari dataset Kaggle. Dataset terdiri dari 50.001 data dan berisi kolom atribut *customer_id*, *age*, *annual_income*, *months_active*, *avg_monthly_spend*, *purchase_frequency*, *avg_order_value*, *discount_usage_rate*, *return_rate*, *browsing_time_minutes*, *support_interactions*, *payment_method*, *region*, dan *customer_segment* sebagai variabel target.

Dataset tersebut dilengkapi dengan label *Customer Segment* yang terbagi menjadi empat kategori, yaitu *High_Value*, *Loyal*, *Occasional*, dan *Regular*. Oleh karena itu, penelitian ini tidak menciptakan segmen pelanggan baru, melainkan membuat model yang bisa mengenali dan memprediksi *Customer Segment* berdasarkan data yang sudah ada.

2. Pra-pemrosesan Data (Pre-processing)

Tahap pra-pemrosesan dilakukan agar data dapat digunakan dengan baik dalam proses klasifikasi. Langkah pertama adalah memeriksa nilai kosong atau data yang tidak terisi pada setiap atribut. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki data yang rusak atau hilang sehingga seluruh data dapat digunakan pada tahap berikutnya. Selanjutnya dilakukan tahap pemilihan fitur (*feature selection*) untuk menentukan atribut yang digunakan sebagai variabel prediktor. Pada penelitian ini atribut *customer_id* tidak digunakan karena hanya berfungsi sebagai

identitas pelanggan dan tidak memberikan informasi yang berpengaruh terhadap proses klasifikasi. Atribut yang digunakan sebagai fitur meliputi *age*, *annual_income*, *months_active*, *avg_monthly_spend*, *purchase_frequency*, *avg_order_value*, *discount_usage_rate*, *return_rate*, *browsing_time_minutes*, *support_interactions*, *payment_method*, dan *region*, sedangkan atribut *customer_segment* digunakan sebagai variabel target. Tahap ini bertujuan menghilangkan atribut yang tidak relevan sehingga model dapat melakukan proses pembelajaran secara lebih efektif dan meningkatkan performa klasifikasi[13].

3. Transformasi Data (Data Transformation)

Data kategorikal pada atribut *payment_method* dan *region* diubah menjadi bentuk numerik menggunakan metode *Label Encoding* agar dapat diproses oleh algoritma XGBoost. Selain itu, atribut *customer_segment* sebagai variabel target juga dikodekan ke dalam bentuk numerik untuk merepresentasikan empat kelas, yaitu *High_Value*, *Loyal*, *Regular*, dan *Occasional*. Selanjutnya, data dipisahkan menjadi data pelatihan dan data pengujian menggunakan metode *train-test split*. Sebanyak 80% data digunakan sebagai data pelatihan dan 20% sebagai data pengujian. Parameter *stratify* digunakan untuk menjaga proporsi setiap kelas sehingga distribusi data pada data pelatihan dan data pengujian tetap seimbang.

4. Data Mining (Modeling)

Proses penambangan data dilakukan menggunakan algoritma XGBoost Classifier. Algoritma ini dipilih karena memiliki kemampuan klasifikasi yang tinggi, mampu mengolah data berukuran besar secara efisien, serta menggunakan mekanisme regularisasi untuk mengurangi *overfitting*, sehingga menghasilkan performa prediksi yang baik [14].

Dalam penelitian ini, parameter yang digunakan meliputi *n_estimators* sebanyak 200, *max_depth* sebesar 6, *learning_rate* sebesar 0,1, dan *random_state* dengan nilai 42.

Proses pelatihan dilakukan dengan menggunakan data latih, sehingga model dapat memahami pola hubungan antara karakteristik pelanggan dan kategori *Customer Segment*. Setelah proses pelatihan selesai, model dipakai untuk memprediksi kategori pelanggan pada uji data sehingga diperoleh hasil klasifikasi untuk setiap kategori *Customer Segment*.

5. Evaluasi dan Interpretasi (Evaluation)

Model evaluasi dilaksanakan guna mengidentifikasi tingkat kinerja algoritma XGBoost dalam mengklasifikasikan *Customer Segment*. Pengukuran performa model menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* yang diperoleh melalui *Classification Report* untuk mengevaluasi hasil klasifikasi pada setiap kelas [15]. Selain itu, digunakan *Confusion Matrix* untuk menganalisis jumlah prediksi yang benar maupun kesalahan klasifikasi pada masing-masing kelas. Untuk mendukung interpretasi hasil, dilakukan analisis *Feature Importance* guna mengidentifikasi atribut yang memberikan kontribusi paling besar terhadap proses klasifikasi [16]. Selanjutnya, dilakukan visualisasi persebaran *Customer Segment*, analisis keterkaitan antarvariabel menggunakan *heatmap*, serta analisis pengeluaran pelanggan berdasarkan kategori *Customer Segment* menggunakan *boxplot* untuk memperoleh pemahaman yang lebih baik terhadap karakteristik data.

III. HASIL DAN PEMBAHASAN

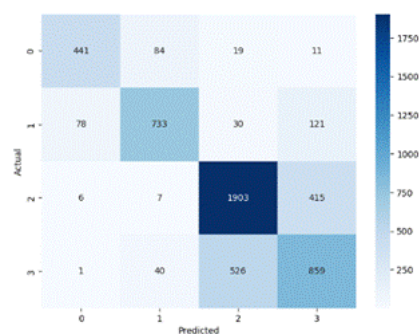
Bab ini menyajikan hasil implementasi algoritma XGBoost dalam mengklasifikasikan pelanggan ke dalam empat kategori *Customer Segment*, yaitu *High_Value*, *Loyal*, *Occasional*, dan *Regular*, berdasarkan karakteristik demografis dan perilaku pelanggan. Selain itu, pada bab ini juga dibahas hasil evaluasi model menggunakan *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, serta analisis *feature importance* untuk mengetahui faktor-faktor yang paling berpengaruh dalam pembentukan *Customer Segment*.

1. Hasil Klasifikasi Menggunakan XGBoost

Penelitian ini menggunakan algoritma XGBoost untuk mengklasifikasikan pelanggan menjadi empat kategori *Customer Segment*, yaitu *High_Value*, *Loyal*, *Occasional*, dan *Regular*. Hasil uji coba menunjukkan model memiliki tingkat akurasi 75%, artinya kemampuan model dalam mengklasifikasikan data cukup memuaskan.

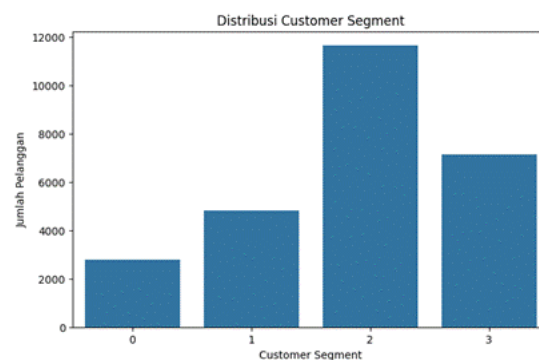
Tabel 1. Hasil Evaluasi Model XGBoost

Classification Report				
Segmen	Precision	Recall	F1-Score	Support
High_Value	0.84	0.79	0.82	555
Loyal	0.85	0.76	0.80	962
Occasional	0.77	0.82	0.79	2331
Regular	0.61	0.60	0.61	1426
accuracy			0.75	5274
macro avg	0.77	0.74	0.75	5274
weighted avg	0.75	0.75	0.75	5274



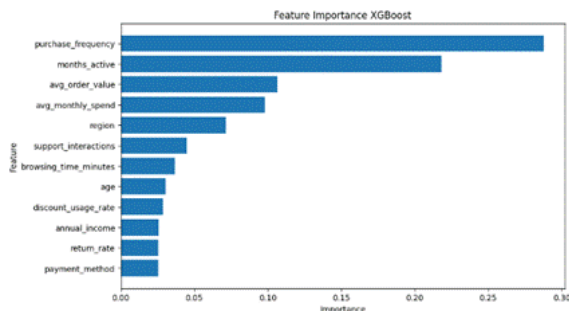
Gambar 1. Visualisasi Confusion Matrix XGBoost

Berdasarkan confusion matrix, yang membandingkan kelas aktual(actual) dengan kelas hasil prediksi(prediction), sebagian besar data berhasil diprediksi dengan tepat. Namun, masih ada kesalahan klasifikasi antara segmen *Occasional* dan *Regular*, yang menunjukkan bahwa kedua segmen tersebut memiliki karakteristik yang cukup mirip.



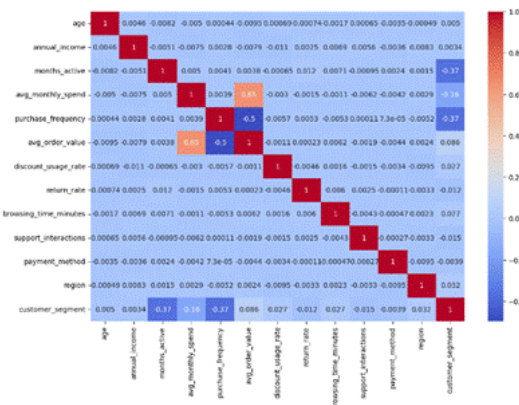
Gambar 2. Visualisasi Distribusi Customer Segment

Hasil distribusi pelanggan menunjukkan bahwa kelompok *Occasional* adalah yang terbesar dengan total 11.653 pelanggan, kemudian diikuti oleh kelompok *Regular* yang berjumlah 7.129 pelanggan, *Loyal* sebanyak 4.808 pelanggan, dan *High_Value* sebanyak 2.777 pelanggan.



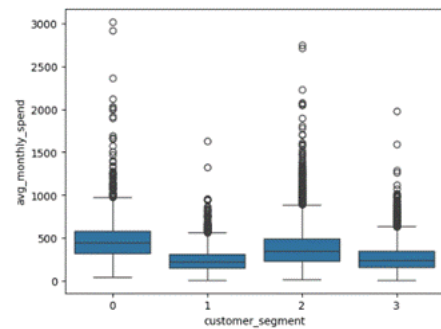
Gambar 3. Feature Importance XGBoost

Analisis Feature Importance atau fitur pentingnya menunjukkan bahwa *purchase_frequency* (0,2877) dan *months_active* (0,2180) adalah variabel paling berpengaruh dalam proses klasifikasi. Variabel lainnya yang penting adalah *avg_order_value* (0,1066) dan *avg_monthly_spend* (0,0980).



Gambar 4. Visualisasi Heatmap Korelasi Antar Variabel

Berdasarkan heatmap korelasi, hubungan yang paling kuat dan positif terjadi antara variabel *avg_monthly_spend* dan *avg_order_value* dengan nilai korelasi mencapai 0,65. Selain itu, frekuensi pembelian dan durasi aktivitas pelanggan memiliki hubungan yang paling kuat dalam membagi pelanggan ke dalam kategori *Customer Segment*.



Gambar 5. Boxplot Average Monthly Spend berdasarkan Customer Segment

Boxplot menunjukkan bahwa segmen *High_Value* memiliki pengeluaran bulanan yang lebih tinggi dibandingkan segmen lainnya dan beberapa pelanggan dalam segmen ini memiliki pengeluaran yang jauh melebihi rata-rata.

2. Pembahasan

Berdasarkan hasil evaluasi, algoritma XGBoost menunjukkan performa yang cukup baik dalam mengklasifikasikan pelanggan ke dalam empat kategori Customer Segment dengan akurasi sebesar 75%. Nilai precision, recall, dan F1-score yang relatif tinggi pada segmen *High_Value*, *Loyal*, dan *Occasional* menunjukkan bahwa model mampu mengenali karakteristik ketiga segmen tersebut dengan baik, sedangkan segmen *Regular* memiliki performa terendah sehingga menjadi segmen yang paling sulit diklasifikasikan. Hasil confusion matrix juga memperlihatkan bahwa kesalahan prediksi paling banyak terjadi antara segmen *Occasional* dan *Regular*, yang mengindikasikan adanya kemiripan karakteristik di antara kedua segmen tersebut. Selain itu, analisis feature importance menunjukkan bahwa *purchase_frequency* dan *months_active* merupakan variabel yang paling berpengaruh dalam proses klasifikasi, diikuti oleh *avg_order_value* dan *avg_monthly_spend*. Hal ini menunjukkan bahwa frekuensi pembelian, lamanya pelanggan aktif, serta tingkat pengeluaran pelanggan merupakan faktor utama dalam membedakan nilai dan

loyalitas pelanggan. Secara keseluruhan, temuan ini menegaskan bahwa XGBoost cukup efektif digunakan untuk klasifikasi *Customer Segment* serta mampu memberikan informasi penting yang dapat dimanfaatkan dalam penyusunan strategi pemasaran yang lebih tepat sasaran.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Extreme Gradient Boosting (XGBoost) untuk melakukan klasifikasi pelanggan ke dalam empat kategori *Customer Segment*, yaitu *High_Value*, *Loyal*, *Regular*, dan *Occasional* berdasarkan karakteristik dan perilaku pelanggan. Hasil pengujian menunjukkan bahwa model yang dirancang dapat mencapai akurasi 75% dengan kinerja klasifikasi yang baik pada mayoritas segmen pelanggan. Selain itu, analisis *feature importance* memberikan informasi mengenai faktor-faktor yang paling berpengaruh dalam proses pembentukan *Customer Segment*. Hasil klasifikasi dan analisis fitur tersebut dapat dimanfaatkan sebagai dasar dalam penyusunan strategi pemasaran yang lebih efektif serta pengelolaan hubungan pelanggan yang lebih optimal. Dengan demikian, tujuan penelitian untuk mengidentifikasi nilai dan loyalitas pelanggan melalui klasifikasi *Customer Segment* menggunakan algoritma XGBoost telah berhasil dicapai.

UCAPAN TERIMA KASIH

Penulis mengungkapkan rasa terima kasih kepada Tuhan Yang Maha Esa atas anugerah dan berkat-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penulis juga menyampaikan rasa terima kasih kepada dosen pembimbing, Program Studi, serta semua pihak yang telah memberikan dukungan, arahan, dan semangat selama proses penelitian. Harapan kami adalah penelitian ini bisa berguna untuk perkembangan ilmu pengetahuan dan studi di masa depan.

REFERENSI

- [1] J. Salminen, M. Mustak, M. Sufyan, and B. J. Jansen, "How can algorithms help in segmenting users and customers? A systematic review and research agenda for algorithmic customer segmentation," *J. Mark. Anal.*, vol. 11, no. 4, pp. 677–692, 2023, doi: 10.1057/s41270-023-00235-5.
- [2] A. Manzoor, M. A. 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, no. March, pp. 70434–70463, 2024, doi: 10.1109/ACCESS.2024.3402092.
- [3] G. Abdurrahman, H. Oktavianto, and M. Sintawati, "Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter Gridsearch dan Random Search Pada Klasifikasi Penyakit Diabetes," vol. 7, no. 3, pp. 193–198, 2022.
- [4] R. Harahap *et al.*, "PERBANDINGAN ALGORITMA RANDOM FOREST DAN XGBOOST UNTUK KLASIFIKASI PENYAKIT PARU-PARU BERDASARKAN DATA," no. 02, pp. 130–141.
- [5] J. Melvin and A. Soraya, "Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit," *J. Ris. Rumpun Mat. dan Ilmu Pengetah. Alam*, vol. 2, no. 2, 2023.
- [6] S. Elina, H. Yulianti, O. Soesanto, and Y. Sukmawaty, "Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit," vol. 4, no. 1, pp. 21–26, 2022.
- [7] B. T. Kristanti, A. Junaidi, and E. P. Mandyartha, "IMPLEMENTASI K-MEANS CLUSTER ING DALAM SEGMENTASI PELANGGAN BERDASARKAN USIA , PENDAPATAN , DAN MODEL RFM (STUDI KASUS : LANTIKYA STORE JOMBANG)," vol. 12, no. 3, 2024.
- [8] M. Alves and G. Tobias, *A review on customer segmentation methods for personalized customer targeting in e-commerce use cases*, vol. 21, no. 3. Springer Berlin Heidelberg, 2023. doi: 10.1007/s10257-023-00640-4.
- [9] R. Kumar, "Online Retail Customer Segmentation Dataset [Dataset]," Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/rohit8527kmr7518/online-retail-customer-classification-dataset>
- [10] I. G. Ngurah and S. Putra, "Integrasi RFM , K-Means , dan XGBoost untuk Optimalisasi Retensi Pelanggan pada Online Retail," vol. 5, no. 1, pp. 1–10, 2025.
- [11] M. S. Kasem, "Customer profiling , segmentation , and sales prediction using AI in direct marketing," *Neural Comput. Appl.*, vol. 36, no. 9, pp. 4995–5005, 2024, doi: 10.1007/s00521-023-09339-6.
- [12] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," pp. 785–794, 2016.
- [13] P. Dhal and C. Azad, *A comprehensive survey on feature selection in the various fields of machine learning*. Applied Intelligence, 2021.
- [14] P. Zhang, Y. Jia, and Y. Shang, "Research and application of XGBoost in imbalanced data," vol. 18, no. 263, 2022, doi: 10.1177/15501329221106935.
- [15] E. Bagli and G. Visani, "METRICS FOR MULTI-CLASS CLASSIFICATION: AN OVERVIEW," *arXiv*, pp. 1–17, 2020.
- [16] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.*, vol. 2, no. January, 2020, doi: 10.1038/s42256-019-0138-9