

Klasifikasi Produk Iphone dengan Menggunakan Algoritma XGBoost

Stevi Freshia Sihombing, Josua Prayuda Pakpahan, Andre Hasudungan Lubis*

Fakultas Teknik, Teknik Informatika, Universitas Medan Area, Medan, Indonesia

Email: ¹stevifreshia@gmail.com, ²josuapakpahan239@gmail.com, ^{3,*}andrelubis2201@gmail.com

Email Penulis Korespondensi: andrelubis2201@gmail.com

Abstrak—Produk iPhone telah menjadi simbol teknologi canggih dan gaya hidup modern yang sangat diminati oleh masyarakat global, termasuk di Indonesia. Dikenal dengan keunggulan pada sistem operasi iOS yang stabil dan eksklusif, produk ini menawarkan integrasi lintas perangkat yang mulus, pembaruan sistem yang konsisten, serta performa tinggi melalui dukungan prosesor generasi terbaru. iPhone juga memiliki daya tarik visual melalui desain minimalis dan elegan, serta fitur-fitur unggulan seperti kualitas kamera profesional, keamanan data tingkat tinggi, dan efisiensi daya. Popularitas iPhone yang tinggi menjadikannya sebagai salah satu produk paling kompetitif di pasar smartphone. Namun, keragaman model, fitur, dan harga dari tiap seri iPhone menyebabkan preferensi pengguna menjadi beragam dan kompleks. Dalam upaya memahami preferensi ini, diperlukan metode klasifikasi yang akurat untuk mengelompokkan produk sesuai daya tarik konsumen. Penelitian ini mengadopsi algoritma XGBoost yang dikenal efektif dalam menangani data kompleks dan besar. Dengan memanfaatkan data transaksi penjualan produk iPhone di pasar Indonesia, model ini dirancang untuk mengidentifikasi pola pembelian dan segmentasi pengguna. Hasil klasifikasi diharapkan dapat memberikan wawasan yang lebih mendalam bagi produsen maupun pemasar dalam merumuskan strategi pemasaran berbasis data yang lebih tepat sasaran.

Kata Kunci: Iphone; Preferensi Konsumen; Klasifikasi Produk; Data Transaksi; XGBoost

Abstract—The iPhone has become a symbol of advanced technology and a modern lifestyle that is highly sought after by the global community, including in Indonesia. Known for its stable and exclusive iOS operating system, this product offers seamless cross-device integration, consistent system updates, and high performance through the support of the latest generation of processors. The iPhone also has a visual appeal through a minimalist and elegant design, as well as superior features such as professional camera quality, high-level data security, and power efficiency. The high popularity of the iPhone makes it one of the most competitive products in the smartphone market. However, the diversity of models, features, and prices of each iPhone series causes user preferences to be diverse and complex. In an effort to understand these preferences, an accurate classification method is needed to group products according to consumer appeal. This study adopts the XGBoost algorithm which is known to be effective in handling complex and large data. By utilizing iPhone product sales transaction data in the Indonesian market, this model is designed to identify purchasing patterns and user segmentation. The classification results are expected to provide deeper insights for manufacturers and marketers in formulating more targeted data-based marketing strategies.

Keywords: Iphone; Consumer Preferences; Product Classification; Transaction Data; XGBoost

1. PENDAHULUAN

Produk iPhone merupakan salah satu perangkat telekomunikasi yang dikenal luas dikalangan masyarakat global, termasuk Indonesia [1]. Keunggulan utama dari produk ini yaitu terletak pada system operasi iOS yang stabil, pembaruan perangkat lunak yang konsisten, serta performa yang tinggi [2]. Design yang elegan serta kombinasi perangkat yang terintegrasi dengan canggih memperkuat posisi Iphone banyak diminati di kalangan pengguna. Selain dianggap sebagai simbol status dan inovasi, tingginya popularitas dan daya saing di pasar menjadikan Iphone topik menarik untuk dikaji melalui pendekatan klasifikasi berbasis kecerdasan buatan [3].

Dalam memahami mengenai preferensi pengguna terhadap tipe pada produk ini masi terbatas dikarenakan Iphone memiliki banyak varian dengan fitur dan harga yang berbeda. Hal ini menyebabkan produsen dan pemasar sulit dalam menentukan target pemasaran yang tepat untuk setiap segmen produk [4]. Dengan berbagai preferensi pelanggan tersebut dibutuhkan metode klasifikasi yang akurat untuk mengelompokkan produk berdasarkan daya tarik pelanggan. Dalam artikel ini algoritma XGBoost menjadi pilihan yang efisien dikarenakan mampu menangani data kompleks serta menghasilkan prediksi yang akurat dalam mengklasifikasikan produk sesuai kebutuhan target pasar.

Melalui teknik boosting yang iteratif dan penyederhanaan model dalam mencegah overfitting XGBoost dipilih karena mampu dalam menangani data dengan berbagai macam fitur kompleks [5]. Dengan mengelompokkan produk kedalam beberapa kategori model XGBoost secara efektif dapat mengidentifikasi pola dalam data pelanggan. Dengan adanya pendekatan menggunakan algoritma ini, diharapkan klasifikasi produk menjadi lebih akurat dan dapat mendukung dalam pengambilan keputusan.

Sejumlah penelitian sebelumnya telah meneliti isu klasifikasi produk smartphone berdasarkan preferensi konsumen dengan memanfaatkan berbagai metode pembelajaran mesin yang memiliki keunggulan dan kelemahan masing-masing. [6] Dalam penelitian ini, algoritma Random Forest digunakan sebagai metode klasifikasi yang utama. Random Forest diakui efektif dalam menangani data dengan dimensi tinggi dan variabel yang kompleks, serta memiliki kelebihan dalam mengurangi risiko overfitting lewat proses penggabungan hasil dari banyak pohon keputusan (decision trees). Studi tersebut menunjukkan bahwa Random Forest dapat memberikan segmentasi pasar yang tepat berdasarkan preferensi konsumen, sehingga membantu produsen dalam merumuskan strategi pemasaran yang lebih efektif. Di samping itu, pendekatan ini juga menawarkan kemampuan analisis yang cukup baik terhadap elemen-elemen penting yang berpengaruh pada pengelompokan produk.

Di sisi lain, [7] mengeksplorasi penerapan Support Vector Machine (SVM) untuk mengklasifikasikan produk elektronik, termasuk smartphone, dengan cara yang berbeda. SVM unggul dalam menangani dataset yang tidak linier dengan memanfaatkan fungsi kernel untuk menghasilkan margin pemisahan terbaik antara kelas. Studi ini menyoroti keefektifan SVM dalam mengelola dataset yang rumit dan fitur yang beragam, sehingga menghasilkan klasifikasi yang akurat. Akan tetapi, salah satu tantangan yang dihadapi adalah kebutuhan untuk melakukan proses penyetelan parameter kernel yang cukup mendalam guna mencapai performa yang optimal. Walaupun begitu, pendekatan ini masih menjadi opsi yang penting, khususnya dalam situasi data yang menunjukkan pola non-linier dan fitur interaksi yang rumit.

Selain dua metode tersebut, [8] melakukan penelitian dengan menggunakan Artificial Neural Networks (ANN) yang dapat mengidentifikasi pola non-linear serta interaksi fitur yang kompleks dalam data preferensi pengguna smartphone. ANN dikenal karena kemampuannya dalam mempelajari data besar dan beragam, yang membuatnya sangat sesuai untuk aplikasi klasifikasi produk dengan berbagai variabel. Studi ini menekankan kelebihan ANN dalam mengenali pola yang tidak mudah ditangkap oleh algoritma konvensional. Namun, mereka juga menekankan bahwa tanpa pengaturan arsitektur yang tepat dan teknik regulasi, ANN dapat mengalami overfitting, yang dapat mengurangi kemampuan generalisasi model saat berhadapan dengan data baru.

Metode lain yang juga banyak digunakan adalah K-Nearest Neighbors (KNN), sebagaimana dijelaskan oleh [9]. KNN adalah algoritma dasar yang melakukan pengelompokan berdasarkan kesamaan fitur antar data, dengan cara menemukan k tetangga terdekat dari data yang akan dikelompokkan. Keunggulan utama KNN terletak pada kemudahannya untuk diimplementasikan dan pemahaman hasil yang diperoleh. Studi ini sukses menerapkan KNN untuk segmentasi pasar produk smartphone berdasarkan atribut serta preferensi konsumen. Namun, tantangan yang dihadapi terutama terdapat pada efisiensi komputasi, karena KNN memerlukan perhitungan jarak yang intensif, terutama ketika ukuran dataset besar, sehingga memerlukan optimasi khusus untuk aplikasi berskala besar.

Selanjutnya, [10] menerapkan algoritma Gradient Boosting Machines (GBM) dalam penelitian klasifikasi produk elektronik, seperti smartphone, dengan tujuan untuk meningkatkan akurasi serta mengurangi kesalahan dalam prediksi. GBM adalah teknik ensemble yang mengintegrasikan beberapa model sederhana secara bertahap untuk menciptakan model yang lebih kuat dan tahan banting. Keunggulan GBM ada pada kemampuannya untuk mengelola fitur yang kompleks dan beragam, serta mekanisme boosting yang dapat secara perlahan memperbaiki kesalahan prediksi. Penelitian ini mengindikasikan bahwa GBM dapat menghasilkan klasifikasi yang lebih tepat dibandingkan metode tradisional dan menawarkan fleksibilitas dalam menyesuaikan parameter model untuk memperoleh performa terbaik.

Kelima penelitian ini menggambarkan berbagai metodologi machine learning yang dapat diterapkan untuk menyelesaikan masalah klasifikasi produk smartphone sesuai dengan preferensi pengguna. Setiap metode menyajikan kelebihan tertentu yang bisa disesuaikan dengan sifat data dan tuntutan aplikasi. Contohnya, Random Forest dan GBM lebih baik dalam mengatasi data dengan banyak dimensi dan kompleksitas fitur, sedangkan SVM dan ANN handal dalam mengelola data yang memiliki pola non-linier yang rumit. KNN, walaupun sederhana, tetap penting untuk analisis yang bergantung pada kesamaan dengan memperhatikan efisiensi komputasi.

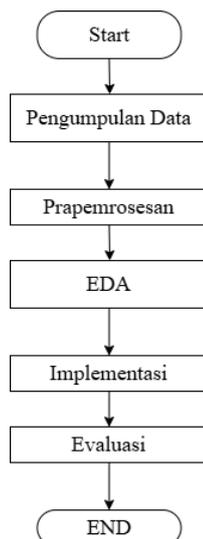
Dalam konteks penelitian klasifikasi produk iPhone, pemilihan algoritma yang tepat sangat krusial agar dapat menangkap preferensi pelanggan secara akurat dan mendukung pengambilan keputusan pemasaran yang efektif [11]. Pendekatan boosting yang digunakan oleh XGBoost, seperti yang dijelaskan dalam penelitian ini, merupakan pengembangan dari ide-ide pada GBM dengan optimasi yang lebih efisien dan pencegahan overfitting yang lebih baik [12]. Oleh karena itu, dengan mengadopsi XGBoost, penelitian ini diharapkan dapat memberikan solusi yang lebih handal dan akurat dalam klasifikasi produk iPhone berdasarkan data transaksi dan preferensi konsumen yang kompleks.

Dengan memperhatikan hasil dan metode dari kajian-kajian sebelumnya, penelitian ini juga memberikan kontribusi dalam memperkuat pemahaman mengenai implementasi algoritma boosting pada klasifikasi produk dengan data pasar yang besar dan beragam. Pendekatan ini tidak hanya bermanfaat untuk memperbaiki akurasi prediksi, tetapi juga mendukung segmentasi pasar yang lebih tepat, sehingga membantu produsen dan pemasar dalam strategi penetapan target pasar yang lebih efektif dan efisien. Dengan menggunakan algoritma XGBoost penelitian ini bertujuan untuk memberikan solusi efektif dalam klasifikasi produk Iphone berdasarkan preferensi pelanggan [13]. Dengan kemampuan XGBoost yang unggul dalam menangani data kompleks dan besar algoritma ini diharapkan mampu dalam meningkatkan akurasi klasifikasi serta meminimalisir kesalahan prediksi. Oleh karena itu, penelitian ini menekankan bahwa algoritma XGBoost adalah metode yang tepat untuk mengatasi permasalahan klasifikasi produk Iphone yang dihadapi.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini terdiri dari beberapa tahapan yang dilakukan secara sistematis mulai dari pengumpulan data hingga evaluasi hasil model machine learning. Gambar 1 menampilkan aliran data sebagai berikut



Gambar 1. Tahapan Penelitian

Pada tahap penelitian ini, start merupakan tahapan awal dalam proses penelitian. Dalam tahap ini sudah ada tujuan penelitian yang jelas kemudian di tahap pengumpulan data akan di kumpulkan data yang diperlukan untuk penelitian ini, data yang di kumpul berasal dari sumber sekunder berupa dataset. Data yang sudah di kumpulkan biasanya belum siap pakai dikarenakan kemungkinan adanya missing values, noise, atau format yang tidak sesuai. Pada tahap prapemrosesan dilakukan pembersihan data atau normalisasi dengan tujuan agar data lebih bersih dan siap untuk di gunakan. Tahap Explratory Data Analysis (EDA) bertujuan memahami data yang dilakukan dengan visualisasi data, statistic data dan mencari pola atau korelasi antara fitur pada data dan EDA akan meminimalisir masalah data lebih awal. Pada tahap implementasi dilakukan pembangunan model machine learning menggunakan algoritma XGBoost yang termasuk di dalam nya proses training menggunakan data yang sudah di proses dan tuning. Setelah model di buat tahap evaluasi akan menilai performa model dengan convolusional matrix seperti akurasi, presisi, recall, dan F1-Score.

2.2 Pengumpulan Data

Dimana data diperoleh dengan memanfaatkan dataset sekunder yang di peroleh dengan mengakses situs web Kaggle [<https://www.kaggle.com/datasets/gerhardien/iphone-transactions-indonesia-market>].

Dataset ini merupakan kumpulan data transaksi penjualan produk iPhone yang terjadi di pasar Indonesia. Setiap baris pada dataset merepresentasikan satu transaksi penjualan yang tercatat secara unik melalui atribut `transaction_id`, yang berfungsi sebagai identifikasi tunggal untuk masing-masing transaksi.

Setiap transaksi melibatkan sebuah produk iPhone yang diwakili oleh `product_id`, yaitu kode unik dari model iPhone yang di jual. Informasi `quantity` menunjukkan jumlah unit produk yang dibeli dalam satu transaksi, sementara `unit_price` merepresentasikan harga satuan dari produk tersebut dalam satuan rupiah.

Sebagai bagian dari strategi promosi atau penjualan, sejumlah transaksi mungkin menerima discount dalam bentuk persentase potongan harga. Nilai diskon ini mempengaruhi besarnya total, yaitu nilai akhir dari transaksi setelah diskon diterapkan. Kolom total ini mencerminkan pendapatan bersih diterima dari transaksi tersebut setelah pengurangan diskon.

Secara keseluruhan, dataset ini dapat digunakan untuk menganalisis performa penjualan produk iPhone di Indonesia, mengevaluasi efektivitas diskon terhadap volume pembelian, hingga melakukan segmentasi pelanggan atau prediksi penjualan di masa mendatang.

Adapun rangkuman data dapat dilihat di table berikut ini :

Tabel 1. Pengumpulan Data

transaction id	product id	quantity	unit price	discount	total
25218	12	3	7000000	15	17850000.0
78866	9	1	15000000	15	12750000.0
71024	1	3	11000000	15	28050000.0
45980	6	4	12000000	15	40800000.0
35845	4	3	17000000	15	43350000.0
....
35528	2	3	11000000	15	28050000.0
49627	8	4	7000000	15	23800000.0
18288	4	1	11000000	15	9350000.0
52222	7	1	11000000	15	9350000.0
72311	1	2	7000000	5	13300000.0

Contoh data dari tabel menunjukkan transaksi dengan berbagai produk, jumlah pembelian, harga, dan diskon yang berbeda-beda.

Dengan menggunakan data ini, penelitian dapat menganalisis pola transaksi, harga, dan diskon untuk membantu model machine learning dalam memahami perilaku pembelian di pasar iPhone Indonesia.

2.3 Prapemrosesan data

Pemrosesan data dilakukan agar tidak mempengaruhi hasil analisis, tahapan ini melakukan pemeriksaan pada missing value atau nilai yang hilang pada dataset serta penanganannya. Untuk masalah yang di hadapi dilakukan feature engineering yaitu memodifikasi atau membuat fitur lebih representative.

Dalam proses pengolahan data transaksi penjualan iPhone di Indonesia, terdapat beberapa tahapan penting yang dilakukan untuk memastikan kualitas dan kesiapan data sebelum digunakan dalam analisis atau pemodelan machine learning.

Pertama, dilakukan pengecekan terhadap missing value. Missing value merupakan data yang kosong atau tidak tercatat dalam dataset, yang dapat menyebabkan hasil analisis menjadi bias atau model prediktif menjadi kurang akurat. Untuk mengatasi hal ini, nilai yang hilang dapat ditangani dengan dua cara: diisi menggunakan metode tertentu seperti imputasi (mean, median, atau mode), atau dihapus jika jumlahnya terlalu kecil dan tidak signifikan terhadap keseluruhan data.

Tahapan kedua adalah feature engineering, yaitu proses memodifikasi, menggabungkan, atau menciptakan fitur-fitur baru dari data asli agar bisa lebih merepresentasikan informasi yang relevan. Misalnya, dari fitur unit_price, quantity, dan discount, dapat dihitung ulang total nilai transaksi, margin keuntungan, atau fitur turunan lainnya. Tujuan dari feature engineering ini adalah memberikan informasi tambahan yang lebih bermakna agar model dapat belajar dengan lebih optimal.

Selanjutnya, dilakukan normalisasi data menggunakan Min-Max Scaler. Karena nilai antar fitur dalam dataset memiliki rentang yang berbeda—seperti unit_price yang bisa mencapai jutaan rupiah, sementara quantity hanya berkisar satuan atau puluhan—maka diperlukan proses normalisasi. Dengan menggunakan Min-Max Scaler, setiap fitur akan diubah ke dalam rentang nilai yang seragam, biasanya antara 0 hingga 1. Proses ini penting agar model tidak memprioritaskan fitur tertentu hanya karena skala angkanya lebih besar.

Ketiga tahapan ini merupakan bagian esensial dalam tahap preprocessing data untuk memastikan bahwa data yang digunakan siap dan optimal dalam menghasilkan insight maupun prediksi yang akurat.

Hasil akhir analisis kemudian di normalisasi menggunakan metode Min-Max Scaler yang mengubah nilai fitur ke dalam rentang 0 sampai 1 dengan rumus:

$$X' = \frac{X_0 - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Penjelasan Rumus Min-Max Scaler

X' = nilai fitur yang sudah dinormalisasi (skala 0 sampai 1)

X_0 = nilai asli fitur sebelum normalisasi

X_{min} = nilai minimum dari fitur tersebut dalam dataset

X_{max} = nilai maksimum dari fitur tersebut dalam dataset

rumus ini "mengubah skala" nilai asli agar:

Nilai terkecil di fitur jadi 0

Nilai terbesar di fitur jadi 1

Nilai lainnya terletak di antara 0 dan 1 secara proporsional

2.4 Exploratory Data Analysis (EDA)

Yaitu melakukan analisis statistik deskriptif dengan tujuan memahami karakteristik data secara umum, pemeriksaan korelasi antar fitur untuk mengidentifikasi hubungan yang mungkin ada antar variabel, dan visualisasi data dengan berbagai plot untuk memudahkan interpretasi dan distribusi data.

Dalam konteks klasifikasi produk iPhone menggunakan algoritma XGBoost, EDA berperan penting dalam memahami distribusi data fitur-fitur produk seperti harga, spesifikasi teknis, dan kategori produk agar diketahui apakah data sudah seimbang atau memiliki bias tertentu. Selain itu, EDA juga membantu dalam memastikan kualitas data dengan memeriksa adanya missing value, duplikat, maupun outlier yang dapat memengaruhi performa model. Analisis korelasi antar fitur juga dilakukan untuk mengidentifikasi hubungan antar variabel yang dapat digunakan untuk pemilihan fitur agar model tidak mengalami overfitting. Visualisasi data seperti histogram, heatmap, dan boxplot mempermudah dalam mendeteksi pola, distribusi, dan anomali data. Selain itu, penting untuk mengevaluasi distribusi label kelas produk iPhone guna mengantisipasi masalah ketidakseimbangan kelas yang dapat mempengaruhi akurasi model XGBoost. Hasil dari EDA juga menjadi dasar dalam melakukan feature engineering, yaitu pembuatan atau transformasi fitur baru yang dapat meningkatkan kinerja klasifikasi produk iPhone dengan algoritma XGBoost.

2.5 Implementasi Algoritma

Implementasi Algoritma dilakukan dengan menerapkan algoritma machine learning pada kasus yang diteliti.

Implementasi algoritma XGBoost dalam klasifikasi produk iPhone bertujuan untuk mengelompokkan atau memprediksi jenis produk iPhone berdasarkan sejumlah fitur yang tersedia dalam dataset transaksi [14]. XGBoost, yang merupakan singkatan dari Extreme Gradient Boosting, adalah algoritma machine learning berbasis pohon keputusan (decision tree) yang sangat populer karena kecepatan, efisiensi, dan akurasinya yang tinggi, terutama dalam tugas klasifikasi dan regresi [15].

Dalam konteks klasifikasi produk iPhone, XGBoost dimanfaatkan untuk mempelajari pola dari fitur-fitur yang tersedia seperti harga satuan (unit_price), jumlah pembelian (quantity), besarnya diskon (discount), dan total transaksi (total) [16]. Model ini bekerja dengan cara membangun banyak pohon keputusan secara bertahap, di mana setiap pohon baru yang dibentuk akan mencoba memperbaiki kesalahan dari pohon-pohon sebelumnya melalui proses boosting. Sebelum model dilatih, data perlu diproses terlebih dahulu melalui tahapan pembersihan data, feature engineering, dan normalisasi agar semua fitur berada pada skala yang seragam dan model dapat belajar secara optimal. Setelah itu, data dibagi menjadi dua bagian yaitu data latih (training) dan data uji (testing). Model XGBoost kemudian dilatih menggunakan data latih, dan diuji akurasinya menggunakan data uji [17].

Keunggulan dari XGBoost dalam kasus ini adalah kemampuannya menangani data yang kompleks, bekerja dengan baik pada dataset yang besar, serta memiliki fitur regularisasi yang membantu mencegah overfitting [18]. Dengan menggunakan XGBoost, klasifikasi produk iPhone dapat dilakukan dengan akurasi tinggi, sehingga sangat membantu dalam memahami pola preferensi pembelian, pengelompokan produk berdasarkan karakteristik transaksi, serta mendukung strategi pemasaran dan penjualan yang lebih tepat sasaran [19]. Hasil pelatihan model kemudian dievaluasi menggunakan metrik akurasi dan confusion matrix untuk memastikan kualitas klasifikasi. Dengan akurasi yang tinggi serta kecepatan pelatihan yang optimal, XGBoost menjadi pilihan algoritma yang sangat efektif dalam mendukung proses klasifikasi produk iPhone secara otomatis dan cerdas [20].

Pseudocode Klasifikasi Produk iPhone dengan XGBoost

1. Mulai
2. Import library yang dibutuhkan
3. Load dataset
4. Lakukan preprocessing data:
 - a. Bersihkan data
 - b. Ubah data kategorikal ke format numerik
 - c. Bagi data menjadi fitur dan label kelas produk
 - d. Bagi dataset menjadi data latih dan data uji
5. Inisialisasi model XGBoost dengan parameter default atau yang sudah di-tuning
6. Latih model XGBoost dengan data latih
7. Lakukan prediksi pada data uji
8. Evaluasi performa model dengan metrik klasifikasi
9. Simpan model XGBoost yang terbaik untuk digunakan pada klasifikasi produk iPhone selanjutnya
10. Selesai

Pseudocode yang disajikan menggambarkan langkah-langkah sistematis dalam membangun model klasifikasi produk iPhone menggunakan algoritma XGBoost. Proses dimulai dengan pengimporan library yang diperlukan dan pemuatan dataset yang relevan. Selanjutnya, dilakukan preprocessing data yang meliputi pembersihan data, konversi variabel kategorikal menjadi format numerik, serta pemisahan data menjadi fitur dan label kelas produk. Dataset kemudian dibagi menjadi data latih dan data uji untuk keperluan pelatihan dan evaluasi model.

Model XGBoost diinisialisasi dengan parameter yang dapat berupa nilai default atau hasil tuning untuk mengoptimalkan performa. Setelah itu, model dilatih menggunakan data latih dan dilakukan prediksi terhadap data uji. Performa model dievaluasi menggunakan metrik klasifikasi yang sesuai untuk menilai akurasi dan efektivitas model dalam mengklasifikasikan produk iPhone. Terakhir, model terbaik disimpan untuk digunakan dalam klasifikasi produk iPhone di masa mendatang.

2.6 Evaluasi

Evaluasi dimana hasil dari model yang telah diimplementasikan diuji menggunakan metrik evaluasi yang sesuai dengan jenis machine learning yang digunakan. Evaluasi merupakan tahap penting dalam proses implementasi model machine learning, dimana hasil dari model yang telah diimplementasikan diuji menggunakan metrik evaluasi yang sesuai dengan jenis machine learning yang digunakan. Dalam konteks klasifikasi produk iPhone menggunakan algoritma XGBoost, evaluasi dilakukan untuk mengukur seberapa baik model dalam mengklasifikasikan produk ke dalam kategori yang benar.

Pada model klasifikasi, terdapat beberapa metrik evaluasi utama yang digunakan. Akurasi adalah metrik yang mengukur seberapa besar proporsi prediksi yang benar dibandingkan dengan keseluruhan data yang diuji. Metrik ini sering digunakan ketika distribusi data seimbang, karena memberikan gambaran umum mengenai performa model.

Selain itu, precision juga menjadi salah satu metrik penting, yaitu mengukur proporsi prediksi positif yang benar-benar tepat. Precision sangat berguna ketika kesalahan dalam prediksi positif memiliki dampak besar, seperti saat merekomendasikan produk premium kepada pelanggan tertentu. Kemudian ada recall, yaitu metrik yang menunjukkan seberapa besar proporsi data aktual positif yang berhasil ditemukan oleh model. Recall berguna untuk memastikan bahwa model tidak melewatkan terlalu banyak data yang seharusnya diprediksi sebagai positif.

Terakhir, digunakan pula F1-score, yaitu rata-rata harmonis antara precision dan recall. Metrik ini sangat berguna ketika diperlukan keseimbangan antara kedua aspek tersebut, terutama pada kondisi di mana data tidak seimbang atau terdapat ketimpangan antar kelas produk.

- a. Akurasi : Untuk mengukur seberapa banyak prediksi model yang benar dibandingkan dengan total keseluruhan prediksi.

$$Accuracy = \left[\frac{TP+TN}{TP+TN+FP+FN} \right] \quad (3)$$

Keterangan :

TP (True Positive) : prediksi benar positif

TN(True Negative) : prediksi benar negative

FP(False Positive) : prediksi salah positif

FN(False Negative) : prediksi salah negative

- b. Precision : Untuk menunjukkan proporsi prediksi positif yang benar

$$Precision = \left[\frac{TP}{TP+FP} \right] \quad (4)$$

- c. Recall : Untuk mengukur kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya

$$Recall = \left[\frac{TP}{TP+FN} \right] \quad (5)$$

- d. F1-Score : Adalah harmonic rata-rata dari precision dan recall dimana berguna saat data tidak seimbang

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Statistik Data :

Tabel 2. Data Statistik

	transaction id	Product id	Quantity	Unit price	Discount	Total
Count	200000	200000	200000	2.000000	200000	2.000000
Mean	49947.82	6.489240	2.500660	1.32494	9.176325	3.008285
Std	28867.93	3.451037	1.117075	3.000877	5.337469	1.548617
Min	2	1.000000	1.000000	7.000000	0.000000	5.950000
25%	24997.000000	3.000000	2.000000	1.200000	5.000000	1.615000
50%	49901.000000	6.000000	3.000000	1.300000	10.000000	2.850000
75%	74945.750000	9.000000	3.000000	1.500000	15.000000	4.180000
max	100000	12.000000	4.000000	1.800000	15.000000	7.200000

Berdasarkan hasil analisis statistik deskriptif terhadap dataset transaksi produk iPhone sebanyak 200.000 data, dapat disimpulkan bahwa dataset ini memuat 12 jenis produk yang berbeda. Rata-rata jumlah unit produk yang dibeli per transaksi adalah sekitar 2,5 unit, dengan harga per unit rata-rata sebesar Rp13.249.400. Diskon yang diberikan pada transaksi bervariasi dengan nilai rata-rata sebesar 9,18% dan diskon maksimum sebesar 15%. Nilai total transaksi menunjukkan variasi yang signifikan, mulai dari Rp5.950.000 hingga Rp72.000.000, yang dipengaruhi oleh jumlah unit pembelian, harga satuan, dan besaran diskon. Data tersebut telah dikelompokkan ke dalam dua cluster dengan distribusi yang relatif seimbang, yang merepresentasikan pola transaksi yang berbeda. Variasi nilai standar yang cukup besar pada beberapa fitur menunjukkan adanya heterogenitas yang signifikan dalam karakteristik transaksi, baik dari segi produk, kuantitas, harga, maupun diskon.

3.1.2 Korelasi :

Analisis korelasi pada dataset ini memberikan pemahaman mendalam mengenai hubungan antar variabel yang terlibat dalam transaksi penjualan produk iPhone. Dari hasil korelasi, diketahui bahwa variabel `transaction_id` memiliki korelasi sempurna (nilai 1) dengan dirinya sendiri, namun hampir tidak memiliki korelasi dengan variabel lainnya, yang ditunjukkan dengan nilai mendekati nol. Hal serupa juga terjadi pada `product_id`, yang hanya berkorelasi dengan dirinya sendiri dan tidak menunjukkan hubungan yang signifikan dengan fitur lain dalam dataset.

Variabel `quantity` menunjukkan korelasi positif yang sangat kuat dengan `total`, yaitu sebesar 0.87. Ini mengindikasikan bahwa semakin banyak unit produk yang dibeli, maka semakin besar pula nilai total transaksi. Namun demikian, `quantity` memiliki korelasi negatif yang cukup kuat terhadap `cluster` (-0.76), yang menandakan bahwa semakin banyak produk dibeli, cenderung masuk ke dalam cluster dengan nilai yang lebih rendah.

Sementara itu, unit_price memiliki korelasi positif sedang terhadap total sebesar 0.44. Artinya, meskipun harga satuan mempengaruhi nilai total transaksi, pengaruhnya tidak sebesar quantity. Selain itu, variabel ini juga berkorelasi negatif terhadap cluster dengan nilai -0.33.

Untuk variabel discount, ditemukan korelasi negatif kecil dengan total sebesar -0.12, yang berarti diskon memberikan sedikit penurunan terhadap nilai transaksi. Namun, discount juga memiliki korelasi positif kecil dengan cluster sebesar 0.09, menunjukkan pengaruh yang sangat lemah.

Variabel total sebagai hasil akhir dari transaksi menunjukkan korelasi positif yang kuat dengan quantity dan unit_price, serta korelasi negatif yang kuat dengan cluster sebesar -0.84. Ini menandakan bahwa transaksi dengan nilai besar cenderung masuk ke cluster yang lebih rendah secara numerik.

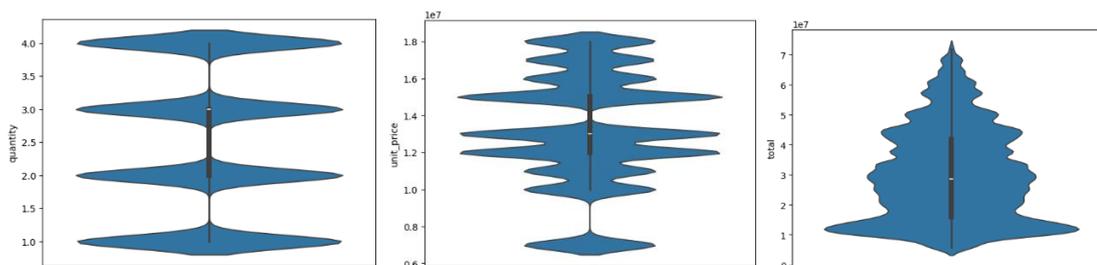
Terakhir, cluster memperlihatkan pola korelasi negatif yang kuat dengan quantity dan total, korelasi negatif sedang dengan unit_price, serta korelasi positif kecil dengan discount. Secara keseluruhan, analisis ini memberikan gambaran yang jelas tentang bagaimana setiap variabel saling berkaitan dan bagaimana pengaruhnya terhadap segmentasi atau pengelompokan data dalam analisis lebih lanjut.

	transaction_id	product_id	quantity	unit_price	discount	total	cluster
transaction_id	1.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00
product_id	-0.00	1.00	0.00	-0.00	-0.00	-0.00	0.00
quantity	-0.00	0.00	1.00	-0.00	0.00	0.87	-0.76
unit_price	0.00	-0.00	-0.00	1.00	-0.00	0.44	-0.33
discount	-0.00	-0.00	0.00	-0.00	1.00	-0.12	0.09
total	-0.00	-0.00	0.87	0.44	-0.12	1.00	-0.84
cluster	-0.00	0.00	-0.76	-0.33	0.09	-0.84	1.00

Gambar 2. Cluster Pola Korelasi

Dalam data ini, quantity dan unit_price adalah dua variabel utama yang memengaruhi total nilai transaksi secara positif. Diskon cenderung mengurangi total transaksi sedikit. Variabel cluster memiliki korelasi negatif dengan sebagian besar variabel utama transaksi (quantity, unit_price, dan total), yang mungkin menunjukkan bahwa klaster ini mewakili segmen dengan transaksi lebih kecil atau pola pembelian berbeda dibandingkan klaster lain.

3.1.3 Plot/violin quantity, unit_price, total:



Gambar 3. Violin Plot

Gambar pertama merupakan violin plot yang menunjukkan distribusi frekuensi dari jumlah unit iPhone (quantity) yang dibeli oleh pengguna. Violin plot ini menggabungkan visualisasi boxplot (tengahnya menunjukkan median dan kuartil) dengan density plot (lebar sisi menunjukkan kepadatan data).

Pada gambar pertama, distribusi quantity tampak diskret dan terpusat pada nilai-nilai 1 hingga 4. Ini menunjukkan bahwa sebagian besar pengguna hanya membeli satu hingga beberapa unit iPhone, dengan puncak yang merata di setiap nilai tersebut. Distribusi ini mencerminkan perilaku pembelian individu yang konsisten dan mengindikasikan bahwa sebagian besar pengguna kemungkinan membeli untuk keperluan pribadi, bukan untuk dijual kembali. Karakteristik ini penting dalam model klasifikasi karena menunjukkan preferensi berbasis volume pembelian yang sederhana namun konsisten.

Gambar kedua menggambarkan distribusi dari unit_price atau harga satuan iPhone. Violin plot ini memberikan informasi tentang seberapa bervariasinya harga produk yang dibeli pengguna.

Gambar kedua menunjukkan distribusi harga satuan iPhone yang lebih kompleks, dengan banyak puncak (multimodal). Ini menandakan bahwa pengguna membeli iPhone dari berbagai kelas harga—mulai dari versi yang lebih terjangkau hingga varian premium. Konsentrasi harga paling tinggi berada di kisaran 10 juta hingga 15 juta rupiah, yang kemungkinan besar merepresentasikan model iPhone kelas menengah hingga flagship. Variabilitas ini menjadi sinyal kuat dalam klasifikasi preferensi, karena pilihan harga sangat mencerminkan jenis produk yang diminati oleh pengguna (misalnya iPhone SE vs iPhone Pro Max).

Gambar ketiga adalah violin plot yang menunjukkan distribusi dari total pembayaran (total), yang dihitung sebagai hasil perkalian antara quantity dan unit_price.

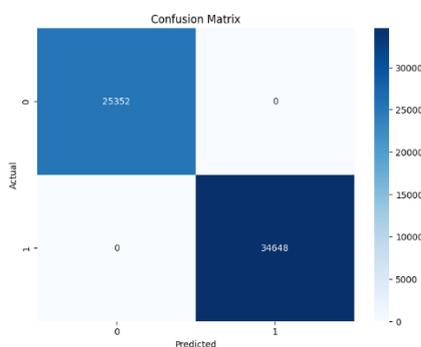
Pada gambar ketiga, distribusi total memiliki bentuk menyerupai segitiga simetris dengan kepadatan terbesar berada pada kisaran 2 juta hingga 4 juta rupiah. Ini menunjukkan bahwa sebagian besar pengguna melakukan pembelian dalam rentang harga tersebut, meskipun terdapat beberapa pembelian bernilai tinggi hingga di atas 6 juta rupiah. Hal ini menunjukkan bahwa meskipun ada pengguna premium, sebagian besar konsumen tetap berada pada segmentasi kelas menengah. Distribusi ini mencerminkan daya beli pengguna dan memperkuat segmentasi yang dapat dilakukan oleh model XGBoost.

3.1.4 Confusion Matrix :

Confusion matrix yang diberikan merupakan sebuah representasi visual dari hasil klasifikasi biner pada sebuah model prediksi. Matriks ini terdiri dari empat komponen utama yang mengukur kinerja model dalam membedakan dua kelas, yaitu kelas 0 dan kelas 1. Sumbu vertikal menunjukkan label aktual (actual), sedangkan sumbu horizontal menunjukkan label yang diprediksi (predicted).

Pada matriks ini, nilai 25352 berada pada sel diagonal kiri atas, yang menandakan jumlah kasus di mana model berhasil memprediksi kelas 0 dengan benar (True Negative). Artinya, sebanyak 25.352 data yang sebenarnya termasuk kelas 0 berhasil diprediksi dengan tepat oleh model. Selanjutnya, nilai 34648 pada diagonal kanan bawah adalah jumlah kasus di mana model memprediksi kelas 1 dengan benar (True Positive), menunjukkan bahwa sebanyak 34.648 data kelas 1 berhasil dikenali dengan akurat.

Tidak terdapat nilai pada sel diagonal kanan atas maupun kiri bawah, yang berarti model tidak menghasilkan prediksi salah sama sekali, baik False Positive (prediksi kelas 1 padahal kelas sebenarnya 0) maupun False Negative (prediksi kelas 0 padahal kelas sebenarnya 1). Kondisi ini menunjukkan bahwa model memiliki akurasi sempurna dalam memisahkan kedua kelas tersebut berdasarkan data uji yang diberikan.



Gambar 4. Confussion Matrix

Dari matriks ini dapat dihitung metrik kinerja utama seperti akurasi, presisi, recall, dan F1-score. Karena tidak ada prediksi yang salah, semua metrik tersebut akan bernilai 1 (atau 100%), menandakan performa model yang sangat baik dalam klasifikasi. Akurasi yang sempurna mengindikasikan bahwa model dapat diandalkan untuk mengklasifikasikan data pada dua kelas ini tanpa kesalahan.

Namun demikian, penting untuk mempertimbangkan bahwa hasil sempurna seperti ini perlu dianalisis lebih lanjut dalam konteks validasi silang dan data yang lebih beragam untuk memastikan bahwa model tidak mengalami overfitting dan benar-benar memiliki kemampuan generalisasi yang baik terhadap data baru. Evaluasi dengan data yang berbeda dan penggunaan metrik tambahan juga disarankan untuk memastikan ketahanan model.

Tabel 3. Evaluasi Model

Accuracy	Precision	Recall	F1-Score
1	1	1	1

Hasil evaluasi model klasifikasi yang ditampilkan dalam confusion matrix dan classification report menunjukkan performa yang sangat baik pada dataset yang diuji. Confusion matrix memperlihatkan bahwa model berhasil mengklasifikasikan seluruh data dengan benar tanpa kesalahan prediksi. Pada kelas 0, sebanyak 25.352 data berhasil diprediksi dengan tepat sebagai kelas 0 (True Negative), sedangkan pada kelas 1, sebanyak 34.648 data juga berhasil diprediksi dengan benar sebagai kelas 1 (True Positive). Tidak terdapat prediksi salah (False Positive maupun False Negative), yang mengindikasikan bahwa model ini memiliki kemampuan pemisahan kelas yang sangat efektif.

Classification report memberikan metrik evaluasi detail seperti precision, recall, dan F1-score untuk masing-masing kelas, yang semuanya bernilai sempurna yaitu 1.00 atau 100%. Precision yang mencapai 1.00 mengindikasikan bahwa setiap prediksi yang dibuat oleh model untuk kelas tertentu benar-benar akurat tanpa adanya kesalahan false positive. Recall yang juga 1.00 menunjukkan bahwa model berhasil mengidentifikasi semua instance aktual dari masing-masing kelas tanpa ada yang terlewat atau salah klasifikasi sebagai kelas lain. F1-score, sebagai rata-rata harmonis dari

precision dan recall, juga mencapai nilai maksimal yang mencerminkan keseimbangan yang sangat baik antara kedua metrik tersebut.

Secara keseluruhan, akurasi model juga tercatat sebagai 1.00, yang berarti 100% dari total 60.000 data uji berhasil diklasifikasikan dengan benar. Nilai makro rata-rata dan rata-rata tertimbang dari precision, recall, dan F1-score juga menunjukkan angka yang sempurna, menegaskan konsistensi performa model di seluruh kelas.

Namun demikian, meskipun hasil ini sangat positif, perlu diperhatikan bahwa akurasi dan metrik sempurna seperti ini perlu diuji lebih lanjut dalam konteks validasi silang, pengujian dengan dataset baru, serta analisis kemungkinan overfitting. Model dengan performa sempurna pada data uji tertentu berpotensi kehilangan kemampuan generalisasi jika data yang diuji kurang beragam atau terlalu mirip dengan data latih.

Dengan demikian, hasil evaluasi ini menunjukkan bahwa model klasifikasi yang diuji memiliki kemampuan prediksi yang sangat baik dan dapat diandalkan untuk klasifikasi data pada kondisi yang sama, namun evaluasi lanjutan tetap diperlukan untuk memastikan kestabilan dan keandalannya dalam aplikasi nyata.

3.2 Pembahasan

Berdasarkan keseluruhan hasil, dapat disimpulkan bahwa pada analisis perilaku transaksi konsumen terhadap produk iPhone yang dikaji secara kuantitatif melalui data statistik deskriptif, analisis korelasi antar variabel, visualisasi distribusi data dengan violin plot, serta evaluasi performa model klasifikasi berbasis XGBoost.

Dari sisi statistik deskriptif, dataset yang terdiri dari 200.000 transaksi memperlihatkan bahwa sebagian besar pembelian produk iPhone terjadi dalam volume kecil, rata-rata sekitar 2–3 unit per transaksi, dengan harga satuan sekitar Rp13 juta dan diskon rata-rata 9%. Total nilai transaksi pun menunjukkan rentang yang cukup lebar, mulai dari sekitar Rp5 juta hingga Rp72 juta, yang mengindikasikan adanya variasi signifikan antar pelanggan dalam hal volume pembelian dan jenis produk yang dipilih. Data tersebut juga telah dikelompokkan ke dalam dua kluster (Cluster 0 dan 1), yang tampaknya mewakili dua segmen konsumen berbeda, misalnya konsumen dengan daya beli rendah hingga sedang dan konsumen premium.

Analisis korelasi mendukung temuan tersebut dengan menunjukkan bahwa variabel yang paling berpengaruh terhadap total nilai transaksi adalah `quantity` dan `unit_price`, di mana `quantity` memiliki korelasi positif sangat tinggi dengan total (0.87). Sementara itu, variabel `discount` memiliki pengaruh negatif kecil terhadap total transaksi, dan cluster menunjukkan korelasi negatif terhadap hampir semua variabel transaksi utama. Artinya, cluster bisa merepresentasikan pembeda antara perilaku konsumen dalam hal belanja besar atau kecil.

Visualisasi melalui violin plot semakin memperjelas gambaran perilaku konsumen. Distribusi `quantity` cenderung terfokus pada nilai rendah (1–4 unit), menegaskan bahwa sebagian besar konsumen adalah pengguna akhir, bukan reseller. Harga satuan (`unit_price`) menunjukkan distribusi multimodal dengan konsentrasi antara Rp10 juta hingga Rp15 juta, yang mencerminkan keberagaman model iPhone yang dibeli. Sementara total pembayaran mayoritas berkisar di angka Rp20–40 juta, yang mengindikasikan kelompok pelanggan kelas menengah sebagai mayoritas.

Kemudian, hasil evaluasi model klasifikasi XGBoost memberikan gambaran performa luar biasa dari model dalam membedakan dua kluster yang ada. Dengan hasil confusion matrix sempurna yang menampilkan distribusi hasil prediksi model terhadap data aktual. Matriks ini terdiri dari empat komponen utama, yakni True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP). Dalam matriks tersebut, model berhasil mengklasifikasikan seluruh 25.352 instance kelas 0 dengan benar (TN) dan seluruh 34.648 instance kelas 1 dengan benar (TP). Tidak ada prediksi yang salah, sehingga nilai False Positive maupun False Negative adalah nol. Kondisi ini menunjukkan bahwa model mampu membedakan kedua kelas tanpa kesalahan sama sekali dalam dataset yang diuji, menandakan performa yang optimal pada tahap ini.

Classification report yang memberikan metrik evaluasi secara rinci, meliputi precision, recall, F1-score, dan support untuk masing-masing kelas. Precision sebesar 1.00 untuk kedua kelas menunjukkan bahwa seluruh prediksi model benar tanpa adanya prediksi positif palsu (false positive). Recall yang juga mencapai nilai sempurna 1.00 mengindikasikan bahwa model mampu menangkap semua instance yang benar dari masing-masing kelas tanpa ada yang terlewat (false negative). F1-score, yang merupakan harmonisasi antara precision dan recall, juga bernilai sempurna, memperkuat bahwa model tidak hanya akurat dalam prediksi positif tetapi juga lengkap dalam menangkap data aktual.

Dukungan (support) menunjukkan jumlah sampel di masing-masing kelas, dengan total 60.000 data uji, yang terbagi secara proporsional ke kelas 0 dan kelas 1.

Hasil ini secara keseluruhan menunjukkan bahwa model klasifikasi memiliki akurasi 100%, yang berarti seluruh data uji diklasifikasikan dengan benar tanpa kesalahan. Metrik makro dan weighted average yang juga bernilai sempurna menegaskan konsistensi dan stabilitas performa model di seluruh kelas.

4. KESIMPULAN

Penelitian ini membuktikan bahwa produk iPhone memiliki persepsi yang sangat positif di mata masyarakat, tidak hanya sebagai perangkat telekomunikasi canggih tetapi juga sebagai simbol status sosial yang prestisius. Keunggulan utama iPhone terletak pada sistem operasi iOS yang stabil, performa tinggi, keamanan data yang baik, serta integrasi antar lini produk Apple yang seamless. Selain itu, fitur-fitur seperti kamera berkualitas tinggi, desain premium, dan pembaruan perangkat lunak yang konsisten menjadi nilai tambah yang signifikan. Namun, keberagaman model, fitur, serta rentang

harga produk menyebabkan preferensi konsumen terhadap iPhone menjadi kompleks dan beragam. Oleh karena itu, diperlukan metode klasifikasi berbasis data untuk mengidentifikasi dan memahami segmentasi pasar dengan lebih akurat. Dalam konteks ini, algoritma XGBoost terbukti efektif dalam mengklasifikasikan produk iPhone berdasarkan data transaksi. Hasil implementasi menunjukkan bahwa model XGBoost memiliki performa klasifikasi yang sangat tinggi, dengan nilai akurasi, presisi, recall, dan F1-score yang mencapai 100%. Hasil tersebut menunjukkan kemampuan model dalam menangkap pola pembelian konsumen dan mengelompokkannya ke dalam segmen yang representatif. Meskipun hasil evaluasi sangat baik, validasi lebih lanjut perlu dilakukan dengan dataset yang lebih beragam untuk menghindari potensi overfitting. Penelitian ini memberikan kontribusi dalam penerapan machine learning untuk analisis pasar dan diharapkan menjadi dasar dalam perumusan strategi pemasaran produk iPhone yang lebih terarah dan berbasis data.

REFERENCES

- [1] M. Wolff, W. J. F. A. Tumbuan, and D. C. A. Lintong, "Pengaruh Gaya Hidup, Harga, Dan Citra Merek Terhadap Keputusan Pembelian Produk Smartphone Merek Iphone Pada Kaum Perempuan Milenial Di Kecamatan Tahuna," *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi*, vol. 10, no. 1, pp. 1671–1681, 2022.
- [2] A. F. Abdul Kadir, A. Habibi Lashkari, and M. Daghmechi Firoozjaci, "iPhone Operating System (iOS)," in *Understanding Cybersecurity on Smartphones: Challenges, Strategies, and Trends*, Springer, 2024, pp. 43–55.
- [3] S. Biswas, I. Ali, R. K. Chakraborty, H. H. Turan, S. Elsayah, and M. J. Ryan, "Dynamic modeling for product family evolution combined with artificial neural network based forecasting model: a study of iPhone evolution," *Technological Forecasting and Social Change*, vol. 178, p. 121549, 2022.
- [4] M. A. B. N. Hidayah, U. Soebiantoro, and Z. Zawawi, "Pengaruh Gaya Hidup, Citra Merek, dan Kualitas Produk terhadap Minat Beli Smartphone Iphone," *Al-Kharaj: Jurnal Ekonomi, Keuangan & Bisnis Syariah*, vol. 6, no. 2, pp. 1993–2003, 2024.
- [5] M. Wiens, A. Verone-Boyle, N. Henscheid, J. T. Podichetty, and J. Burton, "A tutorial and use case example of the eXtreme gradient boosting (XGBoost) artificial intelligence algorithm for drug development applications," *Clinical and Translational Science*, vol. 18, no. 3, p. e70172, 2025.
- [6] F. M. Hidayat, H. Sanjaya, and others, "ANALISIS SENTIMEN PUBLIK TERHADAP PENJUALAN IPHONE 16 DAN KEBIJAKAN TKDN DI INDONESIA," *INFOTECH journal*, vol. 11, no. 1, pp. 74–80, 2025.
- [7] A. Sayuti and others, "Perbandingan Evaluasi Kernel Support Vector Machine Untuk Klasifikasi Sentimen Apple Vision Pro Pada Sosial Media X," Universitas Muhammadiyah Malang, 2025.
- [8] D. Pradana, M. L. Alghifari, M. F. Juna, and D. Palaguna, "Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network," *Indonesian Journal of Data and Science*, vol. 3, no. 2, pp. 55–60, 2022.
- [9] P. Putra, A. M. H. Pardede, and S. Syahputra, "Analisis Metode K-Nearest Neighbour (Knn) Dalam Klasifikasi Data Iris Bunga," *JTIK (Jurnal Teknik Informatika Kaputama)*, vol. 6, no. 1, pp. 297–305, 2022.
- [10] N. N. Sari, T. T. Anisah, and R. Fitriani, "Implementasi Machine Learning Untuk Prediksi Harga Laptop Menggunakan Algoritma Regresi Linear Berganda," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 14, no. 2, pp. 162–177, 2024.
- [11] X. Zhang, F. Guo, T. Chen, L. Pan, G. Beliakov, and J. Wu, "A brief survey of machine learning and deep learning techniques for e-commerce research," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 18, no. 4, pp. 2188–2216, 2023.
- [12] B. Hamdikatama, "BEYOND ALGORITHMS: AN INTEGRATED APPROACH TO FAKE NEWS DETECTION USING MACHINE LEARNING TECHNIQUES".
- [13] R. K. Pratama and F. Piliang, "Rancang Bangun Aplikasi Penyewaan Lapangan Futsal Berbasis Web," *Jurnal Sistem Informasi dan Sains Teknologi*, vol. 1, no. 2, pp. 144–157, 2019, doi: 10.31326/sistek.v1i2.676.
- [14] T. P. P. Sinawang, "Klasifikasi Sentimen Mengenai Rekomendasi HP terbaik Awal Tahun 2024 menggunakan Ekstraksi Fitur TF-IDF dan Algoritma Naïve Bayes," 2024.
- [15] M. Al Faroby and H. Zajuli, "EXTREME GRADIENT BOOSTING UNTUK PENCARIAN PROTEIN YANG BERPENGARUH TERHADAP PRODUKSI INSULIN BERDASARKAN INTERAKSI PROTEIN-PROTEIN," Institut Teknologi Sepuluh Nopember, 2020.
- [16] T. N. Muthmainnah and A. Voutama, "Pendekatan Data Science Untuk Menemukan Customer Churn Pada Perusahaan Fashion Dengan Metode Machine Learning," *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, vol. 6, no. 2, pp. 463–471, 2023.
- [17] G. A. Prabowo, B. Rahmat, and H. E. Wahanani, "Aspect-Based Sentiment Analysis Iphone 14 Pro Menggunakan Algoritma XGBoost," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3947–3952, 2023.
- [18] R. G. Gunawan, E. S. Handika, and E. Ismanto, "Pendekatan Machine Learning Dengan Menggunakan Algoritma Xgboost (Extreme Gradient Boosting) Untuk Peningkatan Kinerja Klasifikasi Serangan Syn," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 453–463, 2022.
- [19] C. Hafidz Ardana, A. A. A. A. Khoyum, and M. Faisal, "Segmentasi pelanggan penjualan online menggunakan Metode K-Means Clustering," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 1, pp. 1–9, 2024.
- [20] A. M. K. Putri and A. F. Rozi, "IMPLEMENTASI CONVOLUTIONAL NEURAL NETWORK DALAM MENENTUKAN TINGKAT KEMATANGAN MENTIMUN DAN TOMAT BERDASARKAN WARNA KULIT," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 5, pp. 10388–10394, 2024.