

Analisis Perbandingan Metode Artificial Neural Network dan XGBoost untuk Prediksi Profit dari Data Transaksi Point of Sale

Panji Kurniawan, Putry Wahyu Setyaningsih*

Fakultas Teknologi Informasi, Sistem Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia

Email: ¹201210117@student.mercubuana-yogya.ac.id, ^{2,*}putryw@mercubuana-yogya.ac.id

Email Penulis Korespondensi: putryw@mercubuana-yogya.ac.id

Submitted: 19/05/2025; Accepted: 19/06/2025; Published: 20/06/2025

Abstrak—Dalam dunia bisnis, profit merupakan indikator kunci dari kesuksesan perusahaan, dan memprediksi profit di masa depan sangat penting untuk pengambilan keputusan strategis, seperti perencanaan inventaris, strategi harga, dan upaya pemasaran. Namun, fluktuasi pasar dan perilaku konsumen yang dinamis sering menjadikan prediksi profit sebagai tantangan yang signifikan. Seiring dengan kemajuan teknologi, penerapan data mining dalam menganalisis data semakin meningkat, salah satunya adalah penggunaan metode *Artificial Neural Network* (ANN) dan XGBoost. Penelitian ini secara eksplisit bertujuan untuk membandingkan kinerja metode ANN dan XGBoost dalam memprediksi profit berdasarkan data transaksi penjualan dari sistem *Point of Sale* (POS). ANN dipilih karena kemampuannya dalam mengenali pola data yang sangat kompleks dan non-linier, sedangkan XGBoost dikenal efisien dalam pemrosesan data dan menangani *overfitting* melalui teknik *boosting* dan regularisasi. Dataset yang digunakan terdiri dari 44.348 transaksi, dengan pembagian data sebesar 80% untuk pelatihan dan 20% untuk pengujian. Hasil penelitian menunjukkan bahwa model ANN mencapai R^2 sebesar 0,9996 dengan MAE sebesar 1.359, sedangkan XGBoost memperoleh R^2 sebesar 0,9978 dengan MAE sebesar 1.600. Perbedaan ini menunjukkan bahwa ANN secara signifikan lebih unggul dalam hal akurasi prediksi. Keunggulan ANN terletak pada kemampuannya dalam membentuk representasi internal yang kompleks dari data, sehingga lebih sensitif terhadap pola-pola halus dalam data transaksi. Hal ini memungkinkan ANN memberikan hasil prediksi yang lebih dekat dengan nilai aktual. Temuan ini menegaskan pentingnya pemilihan model yang sesuai dalam prediksi profit, serta dapat membantu perusahaan dalam mengambil keputusan strategis berbasis data secara lebih akurat dan optimal.

Kata Kunci: Artificial Neural Network; Data Mining; Point of Sale; Prediksi Profit; XGBoost

Abstract—In the business world, profit is a key indicator of a company's success, and predicting future profit is essential for strategic decision-making, such as inventory planning, pricing strategies, and marketing efforts. However, market fluctuations and dynamic consumer behavior often make profit prediction a significant challenge. With technological advancements, data mining methods have become increasingly utilized for analyzing such complex datasets, including Artificial Neural Networks (ANN) and XGBoost. This study explicitly aims to compare the performance of ANN and XGBoost in predicting profit based on transactional data from a Point of Sale (POS) system. ANN was selected for its ability to learn intricate and non-linear patterns in data, while XGBoost is known for its efficiency in processing large datasets and preventing overfitting through boosting and regularization techniques. The dataset consists of 44,348 transactions, with 80% used for training and 20% for testing. Results show that the ANN model achieved an R^2 of 0.9996 and a MAE of 1,359, outperforming the XGBoost model, which obtained an R^2 of 0.9978 and a MAE of 1,600. This significant difference indicates that ANN delivers more accurate predictions. ANN's advantage lies in its capacity to develop complex internal representations of data, making it more responsive to subtle patterns in transactional behavior. These findings highlight the importance of choosing the appropriate model for profit prediction and demonstrate that ANN provides superior predictive accuracy, supporting more precise and data-driven strategic decisions for financial and sales management.

Keywords: Artificial Neural Network; Data Mining; Point of Sale; Profit Prediction; XGBoost

1. PENDAHULUAN

Dalam dunia bisnis saat ini, perusahaan retail telah mengalami transformasi yang sangat signifikan untuk menunjang proses bisnisnya termasuk perkembangan dan pemanfaatan *Point of Sale* (POS) secara efektif. Meskipun data POS menyediakan banyak informasi berharga, aktualnya masih banyak perusahaan yang mengalami kendala dalam memilih dan menerapkan metode analisis yang tepat untuk menghasilkan prediksi yang akurat. Perusahaan masih kesulitan dalam mengolah data tersebut menjadi informasi yang berguna untuk menjadi kunci dalam pengambilan keputusan. Hal ini disebabkan oleh kompleksitas data transaksi dan dipengaruhi oleh berbagai faktor eksternal seperti promosi, musim, dan perilaku konsumen yang dinamis[1]. Dalam konteks ritel, prediksi profit yang akurat sangat krusial karena menyetuh langsung pada perencanaan stok, strategi harga, manajemen promosi, dan alokasi sumber daya. Prediksi yang tidak akurat dapat mengakibatkan *overstocking* atau *stockout*, strategi harga yang tidak kompetitif, serta kampanye pemasaran yang tidak tepat sasaran yang semuanya berdampak pada kerugian finansial dan menurunnya kepuasan pelanggan.

Beberapa tahun terakhir, perkembangan teknologi informasi dan komunikasi telah membawa transformasi pesat dalam operasi bisnis termasuk dalam mengambil keputusan strategis[2]. Perusahaan kini memiliki akses ke volume data yang sangat besar dan beragam. Pemanfaatan data ini secara optimal melalui analisis yang akurat dapat menjadikan perusahaan memperoleh wawasan strategis yang memberikan keunggulan kompetitif bagi perusahaan[3]. Data transaksi POS yang mengandung pola hubungan *non-linear* dan variabel yang beragam memerlukan metode analisis yang canggih untuk dapat menghasilkan prediksi profit yang akurat dan berguna.

Teknik *Machine Learning*(ML) sudah banyak digunakan untuk menganalisis data bisnis karena kemampuannya dalam menangkap hubungan *non-linear* dan kompleks antar variabel[4]. Dua metode ML yang cukup populer dan sering digunakan untuk tugas prediktif adalah *Artificial Neural Network*(ANN) dan *Extreme Gradient Boosting* (XGBoost). ANN dikenal efektif dalam menangani data dengan pola kompleks[5]. ANN dirancang untuk meniru cara kerja *neuron* dalam otak manusia. Metode ini dipilih dalam penelitian karena memiliki sejumlah keunggulan terutama dalam mengatasi permasalahan yang kompleks. ANN mampu bekerja secara efektif pada data berskala besar, dengan jumlah variabel yang banyak, serta mampu memodelkan hubungan yang bersifat *non-linear* antar variabel[6]. Sementara XGBoost adalah algoritma *machine learning* berbasis pohon keputusan yang dikembangkan untuk meningkatkan performa dan efisiensi untuk data terstruktur[7]. Algoritma ini memproses data dengan cara mengembangkan model secara bertahap (*iteratif*), di mana setiap model baru terus memperbaiki kekeliruan dari model sebelumnya[8]. XGBoost dikenal karena kecepatan pelatihan yang tinggi, kemampuan regulasi (*regularization*) yang kuat untuk mencegah *overfitting*, serta akurasi prediksi yang tinggi, terutama pada data tabular dan kompleks[9].

Penelitian yang dilakukan Wibiksana pada tahun 2023 dengan judul “Prediksi Harga Saham pada Portofolio Investor dengan Analisis Time Series Harga Saham menggunakan ANN” bertujuan untuk mengevaluasi keakuratan ANN dalam memprediksi harga saham BNI, BCA, dan Mandiri. Penelitian ini mengindikasikan bahwa ANN memiliki akurasi tinggi, terutama pada data harian, dengan tingkat akurasi masing-masing 97,6927% (BNI), 97,9754% (BCA), dan 97,6275% (Mandiri). Dari data mingguan akurasi sedikit lebih rendah, sementara akurasi bulanan paling rendah dibandingkan lainnya. Dari segi investasi, penelitian ini menemukan bahwa portofolio yang mencakup ketiga saham memberikan keuntungan tambahan sebesar 18,19% dibandingkan investasi pada satu saham saja. Dengan demikian, ANN terbukti sebagai metode yang efektif dalam prediksi saham, dan diversifikasi portofolio untuk peningkatan keuntungan investasi secara signifikan[10]. Penelitian oleh Awalia & Titik Kristanti di tahun 2023, dengan judul yaitu “Prediksi *financial distress* pada Perusahaan Perbankan di BEI periode 2017–2021 Menggunakan Metode Artificial Neural Network (ANN)”. Penilaian *financial distress* didasarkan pada nilai Earnings Per Share (EPS), dengan empat rasio keuangan sebagai input: *current ratio*, *return on asset*, *debt to asset ratio*, dan *total asset turnover*. Model ANN dengan arsitektur 20-5-1 mencapai akurasi 87%, menunjukkan efektivitas ANN dalam mendeteksi potensi krisis keuangan di sektor perbankan[11]. Hendra Wijaya melakukan penelitian terkait prediksi penjualan di sektor ritel menunjukkan bahwa pendekatan konvensional berbasis intuisi semakin tidak relevan seiring dengan meningkatnya kompleksitas data. Menanggapi tantangan tersebut, studi oleh peneliti pada perusahaan ritel XYZ memanfaatkan algoritma XGBoost yang dioptimalkan menggunakan teknik *Random Search*. Sebelum optimasi, model XGBoost mengukuhkan hasil yang memuaskan dengan nilai koefisien determinasi (R^2) sebesar 96,50%. Setelah dilakukan tuning *hyperparameter* menggunakan *Random Search*, nilai R^2 meningkat menjadi 97,21%, menunjukkan peningkatan akurasi prediksi sebesar 0,71%. Dengan prediksi yang lebih akurat, perusahaan ritel dapat mengelola persediaan lebih efisien, menyusun rencana produksi yang lebih tepat, dan mengalokasikan sumber daya secara optimal[12]. Septia et al dalam kajiannya dengan judul “Prediksi Harga Saham pada Bank BCA Tbk dengan Teknik XGBoost” bertujuan untuk memprediksi harga saham Bank BCA Tbk menggunakan teknik XGBoost berdasarkan data historis periode 1 Januari 2019 hingga 31 Desember 2023. Dari penelitian ini diperoleh data prediksi yang menunjukkan bahwa meskipun terdapat fluktuasi, tren harga saham Bank BCA secara umum mengalami peningkatan yang stabil. Faktor eksternal seperti pandemi COVID-19 sempat memicu penurunan harga. Model XGBoost memberikan hasil MAPE sebesar 2,48%, yang menunjukkan tingkat akurasi sangat tinggi dalam meramalkan pergerakan harga saham. Temuan ini menegaskan bahwa pemanfaatan *machine learning* berbasis data historis dapat menjadi alat prediksi yang bernilai tinggi bagi investor dalam pengambilan keputusan investasi, khususnya pada saham perbankan[13]. Ngakan Nyoman et al melakukan penelitian terhadap data kecelakaan tahunan di Provinsi Bali dari Badan Pusat Statistik dengan rentang tahun 1996–2019. Adapun untuk penelitian ini menggunakan metode XGBoost. Hasil peramalan dievaluasi menggunakan RMSE, menunjukkan performa baik pada kategori jumlah orang meninggal (RMSE 4,92) dan luka berat (RMSE 4,11). Sementara itu, kategori jumlah kejadian kecelakaan dan luka ringan memiliki RMSE lebih tinggi, masing-masing 21,69 dan 77,24[13]. Dari referensi sebelumnya dengan judul “Sistem Estimasi Pencapaian Target Profit Menggunakan Model Regresi Berbasis Machine Learning”. Pada Perusahaan Gerai Mustika, belum memanfaatkan *big data* secara optimal untuk memperkirakan pencapaian profit. Untuk mengatasi permasalahan tersebut, penelitian dilakukan dengan membangun sistem estimasi pencapaian target profit menggunakan model regresi berbasis *machine learning*. Model regresi yang digunakan mampu memberikan hasil akurasi yang baik dalam menganalisis data profit perusahaan. Hasil evaluasi menunjukkan performa model mencapai akurasi sebesar 91% dengan rasio data pelatihan dan pengujian sebesar 80:20. Hasil ini menunjukkan bahwa pendekatan *machine learning* dapat digunakan sebagai solusi dalam proses estimasi profit perusahaan secara objektif, cepat, dan akurat. Pendekatan[14]. Raihan Winurputra melakukan penelitian yang bertujuan untuk meramalkan penjualan menggunakan algoritma XGBoost yang diintegrasikan dengan kerangka kerja CRISP-DM (*Cross-Industry Process for Data Mining*). Data historis penjualan rentang Januari 2020 hingga Juni 2023. Model terbaik menunjukkan nilai MAE sebesar 11,58 dan RMSE sebesar 29,19, yang mencerminkan kemampuan peramalan yang baik dan relevan untuk mendukung pengambilan keputusan manajemen persediaan di UB Mart[15].

Namun demikian, terdapat celah penting yang belum banyak dieksplorasi, yaitu perbandingan langsung antara ANN dan XGBoost dalam konteks prediksi profit berbasis data POS yang memiliki karakteristik sangat dinamis dan kompleks. Sebagian besar penelitian sebelumnya hanya fokus pada satu metode tertentu atau pada domain data selain

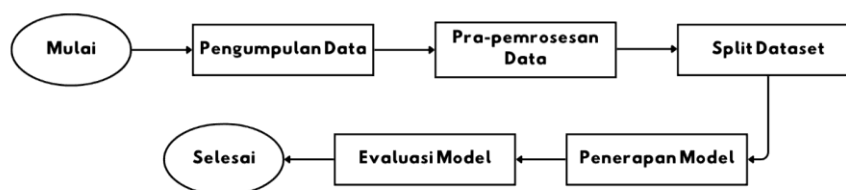
POS, seperti saham atau data makroekonomi. Padahal, data transaksi ritel dari POS menyimpan pola non-linear yang unik dan dipengaruhi oleh banyak variabel kontekstual yang saling terkait.

Topik ini dipilih karena relevansinya dengan kebutuhan industri ritel untuk meningkatkan akurasi prediksi profit melalui pemanfaatan data transaksi POS. Dengan membandingkan dua metode machine learning yaitu ANN dan XGBoost, penelitian ini bertujuan untuk mengidentifikasi pendekatan yang paling efektif dalam konteks data transaksi ritel. Harapan dari penelitian ini bisa memberikan kontribusi signifikan baik dari sisi akademik maupun praktis. Dari segi akademik, penelitian ini menyuguhkan wawasan yang lebih spesifik mengenai efektivitas metode ANN dan XGBoost dalam memprediksi profit berdasarkan data transaksi dari sistem *Point of Sale* (POS). Sementara itu, secara praktis, hasil penelitian ini dapat dijadikan panduan untuk pelaku usaha ritel dalam memilih metode analisis data yang tepat untuk meningkatkan akurasi prediksi profit mereka. Dengan demikian, penelitian ini diharapkan mampu menjembatani kebutuhan antara dunia akademik dan dunia industri dalam mengoptimalkan pemanfaatan data transaksi untuk peningkatan profitabilitas usaha.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan dari penelitian ini diharapkan dapat menghasilkan sistem prediksi yang akurat diantaranya yaitu Pengumpulan data, Pra-pemrosesan data, split dataset, penerapan model, evaluasi model, dan analisis hasil [16].



Gambar 1 Tahapan Penelitian

Gambar 1 merupakan tahapan-tahapan dalam proses penelitian ini. Setiap langkah dalam diagram menunjukkan alur sistematis yang dimulai dari pengumpulan data hingga evaluasi performa model. Penjelasan tiap tahap adalah sebagai berikut:

2.1.1 Pengumpulan Data

Awal dari penelitian ini adalah pengumpulan data transaksi penjualan yang berasal dari ekspor data ke excel dari sistem *Point of Sales* (POS) toko perlengkapan bayi yang sudah menggunakan sistem ERP dengan rentan waktu 31 Mei 2024 s/d 31 Maret 2025. Data yang dikumpulkan mencakup variabel-variabel penting seperti tanggal transaksi, order *reference*, hari, total sales, dan profit.

2.1.2 Pra-pemrosesan Data

Tahapan berikutnya, data mentah dari sistem POS diolah menjadi bentuk yang dapat digunakan oleh model *machine learning*. Prosesnya terdiri dari enam langkah utama, yaitu:

- Date Conversion*
Kolom *Date* diubah ke format *datetime* untuk memungkinkan ekstraksi informasi waktu seperti hari, bulan, dan tahun. Informasi ini penting karena profit sering kali dipengaruhi oleh faktor musiman dan pola mingguan [17].
- Perhitungan Harga Pokok Penjualan (HPP)
HPP adalah biaya yang dikeluarkan untuk mendapatkan atau membuat barang yang dijual ke pelanggan. Biaya ini bisa termasuk harga beli barang, ongkos kirim, atau biaya produksi sampai barang siap dijual [18]. Penambahan kolom HPP_{total} (selisih antara Total Sales dan Profit) dalam penelitian ini bertujuan untuk memberikan informasi yang lebih lengkap mengenai struktur biaya, sehingga membantu model memahami faktor pembentuk profit secara lebih akurat. Kolom ini menjadi fitur penting karena mencerminkan biaya pokok penjualan, yang secara langsung memengaruhi profit. Dengan memasukkan HPP_{total}, model seperti ANN dan XGBoost dapat membangun hubungan yang lebih tepat antara input dan target serta meningkatkan akurasi prediksi.
- One-Hot Encoding*
One-hot encoding merujuk pada metode yang digunakan untuk memodifikasi data berbentuk kategori (seperti nama atau label) menjadi bentuk angka, yaitu vektor biner. Setiap nilai kategori akan diwakili oleh deretan angka 0 dan 1, di mana hanya satu angka 1 yang muncul untuk menunjukkan posisi kategori tersebut, dan sisanya 0. Panjang deretan angka ini sesuai dengan jumlah total kategori yang ada [19]. Fitur *Day_of_week* dan *Month*, yang bersifat kategorikal, dikonversi ke format numerik menggunakan teknik *one-hot encoding*. Dengan representasi biner ini, model seperti ANN dan XGBoost dapat mengenali dan mempelajari pola hubungan antara hari atau bulan tertentu dengan profit tanpa mengasumsikan adanya urutan atau hubungan numerik antar kategori. Hal ini

membantu meningkatkan akurasi prediksi serta menghindari bias yang mungkin timbul dari pemberian nilai numerik langsung pada variabel kategorikal dan model dapat mengolah data secara optimal.

d. Pembersihan Data

Proses pembersihan dilakukan untuk memastikan kualitas data dan mencegah kesalahan pada saat pelatihan model serta validitas hasil penelitian[20].

e. Normalisasi Data

Semua fitur numerik dinormalisasi menggunakan *StandardScaler* untuk mengatur skala fitur menjadi distribusi dengan *mean* 0 dan standar deviasi 1[21]. Ini bertujuan agar model ANN dan XGBoost dapat bekerja lebih efisien dan stabil dalam proses pembelajaran.

2.1.3 Split Dataset

Setelah proses preprocessing selesai, langkah selanjutnya adalah membagi dataset menjadi dua segmen, yaitu data pelatihan (*training set*) sebanyak 80% dan data pengujian (*testing set*) sebanyak 20%. Segmentasi ini digunakan untuk memisahkan data yang digunakan dalam proses pelatihan model dengan data yang digunakan untuk menguji performa model[22].

2.1.4 Penerapan Model

Tahap selanjutnya, dilakukan implementasi dua model pembelajaran mesin, yaitu ANN dan XGBoost, untuk melakukan prediksi nilai *profit* dari data transaksi POS.

a. ANN

Model ANN dibangun menggunakan pustaka *Keras* dari *TensorFlow* dengan arsitektur berlapis[23]. ANN adalah sebuah model komputasi yang meniru cara kerja otak manusia. ANN terdiri dari sejumlah unit kecil yang disebut neuron, yang saling terhubung satu sama lain. Setiap neuron menerima informasi, memrosesnya, lalu meneruskannya ke neuron lain melalui sambungan yang memiliki bobot tertentu (*weight*). Semakin penting informasi yang lewat, semakin besar bobotnya. ANN bekerja dengan cara belajar dari data yang diberikan, sehingga bisa mengenali pola dan memberikan hasil atau prediksi yang konsisten. Setiap neuron bisa menerima banyak masukan dan menghasilkan satu keluaran, yang bisa menjadi hasil akhir atau masukan untuk neuron lainnya. Dengan pelatihan yang cukup, ANN bisa melakukan tugas seperti manusia dalam mengenali pola atau membuat keputusan berdasarkan data[16]. Terdapat tiga lapisan tersembunyi yang masing-masing terdiri dari 128 *neuron* pada lapisan pertama, 64 *neuron* di lapisan kedua, dan 32 *neuron* di lapisan ketiga. setiap lapisan menggunakan fungsi aktivasi ReLU. Untuk meningkatkan generalisasi dan mencegah overfitting, digunakan teknik regularisasi L2 (Ridge) dan Dropout sebesar 20%. Model dikompilasi menggunakan optimizer Adam dan fungsi kerugian *Mean Squared Error* (MSE). Untuk menghindari pelatihan berlebih (*overfitting*), diterapkan juga *Early Stopping*, yaitu menghentikan pelatihan jika *validation loss* tidak membaik dalam 10 epoch berturut-turut. Model ini dilatih selama maksimal 50 epoch dengan *batch size* sebesar 32 dan *validation split* sebesar 20%.

b. XGBoost

Model XGBoost merupakan algoritma *boosting* berbasis pohon keputusan yang sangat efisien dan akurat[24]. Sejak pengembangannya, XGBoost mampu unggul di berbagai kompetisi *machine learning* dan pilihan solusi yang kuat, efisien dan berguna untuk mengatasi berbagai masalah klasifikasi[25]. Model ini memiliki fleksibilitas yang besar dan dapat menyesuaikan modelnya dengan beragam kondisi data, seperti yang terlihat dalam dataset retail. Oleh karena itu, penyetelan parameter menjadi kunci untuk peningkatan kinerjanya. Komponen untuk membangun *xgboost* diantaranya yaitu *n_estimator=200*, *learning_rate=0.1*, *max_depth=10*, dan *random_state=42*. Model dilatih dengan data yang sudah dinormalisasi dan kemudian diuji untuk menghasilkan prediksi nilai profit.

2.1.5 Evaluasi Model

Dalam penelitian ini, performa dari dua model yang digunakan diukur dengan empat metrik evaluasi, diantaranya yaitu :

a. *Mean Absolute Error* (MAE)

MAE digunakan untuk menghitung seberapa besar rata-rata kekeliruan antara nilai yang diprediksi oleh model dan nilai sebenarnya. MAE mengambil selisih absolut antara keduanya, lalu dirata-ratakan. Nilai MAE yang lebih kecil menunjukkan bahwa prediksi model semakin mendekati nilai aktual. Karena hanya menghitung nilai absolut dari selisih, MAE memberikan gambaran yang mudah dipahami dalam satuan asli data (misalnya rupiah, jika yang diprediksi adalah transaksi penjualan).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

Dalam rumus MAE, *n* merupakan jumlah data atau observasi, *y_i* adalah nilai aktual pada observasi ke-*i*, dan *ŷ_i* adalah nilai prediksi untuk observasi ke-*i*. Simbol $|y_i - \hat{y}_i|$ menunjukkan nilai absolut dari selisih antara nilai aktual dan nilai prediksi.

b. *Mean Squared Error* (MSE)



$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{2}$$

MSE juga mengukur selisih antara nilai prediksi dan nilai aktual, namun kesalahan tersebut dikuadratkan terlebih dahulu sebelum dirata-ratakan. Karena kesalahan dikuadratkan, MSE memberikan penalti yang lebih besar terhadap prediksi yang jauh dari nilai sebenarnya. Hal ini menjadikannya metrik yang sensitif terhadap outlier (nilai ekstrem). Nilai MSE yang lebih kecil menandakan performa model yang lebih baik.

c. *Root Mean Squared Error* (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{3}$$

RMSE adalah akar dari MSE. Karena satuannya sama dengan satuan dari data asli, RMSE lebih mudah diinterpretasikan dibandingkan MSE. RMSE tetap memberikan penalti yang lebih besar pada kesalahan besar seperti MSE, namun menyajikannya dalam skala yang sama dengan data. Semakin kecil nilai RMSE, semakin akurat model dalam melakukan prediksi..

d. *R-squared* (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{4}$$

R^2 atau koefisien determinasi mengukur seberapa besar proporsi variansi (keragaman) dalam data target (nilai aktual) yang bisa dideklarasikan oleh model. Nilai R^2 berada pada rentang 0 hingga 1. Semakin mendekati 1, artinya model semakin baik dalam mendeklarasikan variasi data. Jika $R^2 = 1$, berarti model dapat memprediksi nilai target secara sempurna.

Hasil dari metrik-metrik ini digunakan untuk membandingkan performa kedua model dan menentukan model mana yang lebih unggul dalam memprediksi *profit* dari data transaksi POS.

3. HASIL DAN PEMBAHASAN

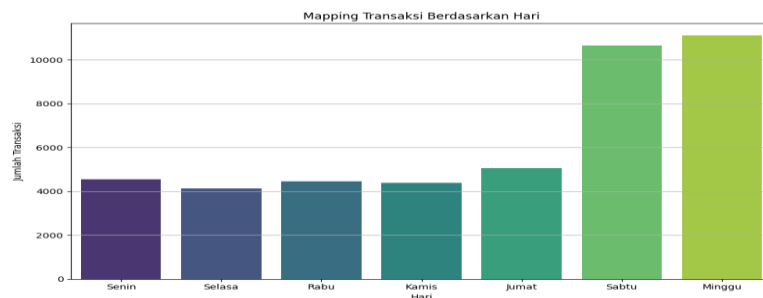
3.1 Pengumpulan Data

Penelitian ini menggunakan data yang diperoleh dari sistem transaksi *Point of Sale* (POS) milik perusahaan yang beroperasi dalam sektor ritel. Rentang waktu pengambilan data dimulai dari tanggal 30 Mei 2024 hingga 31 Maret 2025, yang mencakup hampir sepuluh bulan aktivitas penjualan. Selama periode tersebut, tercatat sebanyak 44.348 transaksi berhasil dihimpun. Seluruh data ini kemudian diolah dan dianalisis lebih lanjut untuk keperluan prediksi profit menggunakan metode ANN dan XGBoost.

Tabel 1. Sample Dataset

Paid Date	Order Ref	Date	Day	Total Sales	Profit
5/30/2024 8:54	POS 2/0025	5/30/2024 8:54	Thursday	239870	115099
5/30/2024 8:58	POS 2/0026	5/30/2024 8:58	Thursday	49000	5105
5/30/2024 9:00	POS-3/0041	5/30/2024 9:00	Thursday	156028	28200
5/30/2024 9:04	POS-3/0042	5/30/2024 9:04	Thursday	75000	6000
5/30/2024 9:32	POS-3/0044	5/30/2024 9:32	Thursday	99600	9089
5/30/2024 9:36	POS-3/0045	5/30/2024 9:36	Thursday	22000	4501
5/30/2024 9:44	POS-3/0046	5/30/2024 9:44	Thursday	123265	33705.8
5/30/2024 9:54	POS-3/0047	5/30/2024 9:54	Thursday	358300	152045.4
5/30/2024 10:15	POS 2/0029	5/30/2024 10:15	Thursday	264500	34338

Pada sample dataset di Tabel 1 Setiap baris dalam tabel merepresentasikan satu transaksi yang mencakup beberapa atribut penting, yaitu waktu pembayaran (*Paid Date*), referensi pesanan (*Order Ref*), tanggal transaksi (*Date*), hari dalam minggu saat transaksi terjadi (*Day*), total penjualan yang diperoleh dari transaksi tersebut (*Total Sales*), serta keuntungan bersih yang dihasilkan (*Profit*). Data ini memberikan gambaran awal mengenai aktivitas penjualan harian.



Gambar 2 Mapping transaksi berdasarkan hari

Gambar 2 memperlihatkan visualisasi jumlah transaksi berdasarkan hari dalam seminggu, yang bertujuan untuk mengidentifikasi pola perilaku konsumen guna menunjang prediksi profit. Dari grafik terlihat bahwa jumlah transaksi cenderung rendah selama hari kerja (Senin hingga Jumat), dengan rata-rata transaksi harian sekitar 4.000–5.000. Namun, terjadi lonjakan signifikan pada akhir pekan, terutama hari Sabtu dan Minggu yang masing-masing mencatat lebih dari 10.000 transaksi. Informasi ini penting karena menunjukkan bahwa hari dalam seminggu merupakan salah satu fitur yang relevan dan potensial dalam model prediksi profit. Dengan memahami tren ini, model dapat memanfaatkan variabel hari sebagai indikator kuat untuk memperkirakan pendapatan dan keuntungan yang lebih akurat berdasarkan pola perilaku pelanggan.

3.2 Pra-pemrosesan Data

3.2.1 Date Conversion

Untuk mendukung proses analisis dan prediksi profit, dilakukan konversi pada kolom tanggal menjadi format waktu yang terstruktur. Hasil dari konversi tersebut ditunjukkan pada tabel dibawah ini yang memuat informasi tambahan berupa hari dalam seminggu, bulan, dan tahun dari masing-masing entri data.

Tabel 2 Data hasil date conversion

No	Date	Day of week	Month	Year
1	2024-05-30	3	5	2024
2	2024-06-01	5	6	2024
3	2024-06-03	0	6	2024
4	2025-03-30	6	3	2025
5	2025-03-31	0	3	2025

Tabel 2 menampilkan hasil konversi dari kolom *Date* menjadi format *datetime* yang memungkinkan ekstraksi komponen waktu seperti hari dalam seminggu (*Day of week*), bulan (*Month*), dan tahun (*Year*). Konversi ini dilakukan agar model prediksi profit dapat memanfaatkan informasi temporal yang berpengaruh terhadap pola penjualan. Contoh, angka pada kolom *Day_of_week* menunjukkan representasi hari dengan format numerik, di mana 0 berarti Senin dan 6 berarti Minggu. Pola ini penting untuk menangkap potensi lonjakan transaksi pada akhir pekan atau pergeseran tren musiman yang berkaitan dengan bulan atau tahun tertentu. Dengan demikian, penambahan fitur-fitur waktu ini diharapkan mampu meningkatkan akurasi model dalam memprediksi profit secara lebih kontekstual.

3.2.2 Perhitungan Harga Pokok Penjualan

Penambahan kolom baru bernama *HPP_total* yang diambil dari selisih antara total sales dengan profit. Pada Tabel 3 ini dapat digunakan untuk menjelaskan bahwa pada tahap prapemrosesan, dilakukan penambahan fitur turunan yang relevan untuk mendukung prediksi profit, seperti *HPP_total* sebagai indikator utama beban biaya produk.

Tabel 3 Penambahan kolom *HPP_total*

Total Sales	Profit	HPP total
239870	115099	124771
49000	5105	43895
156028	28200	127828
922000	623443	298557
721200	622800	98400

3.2.3 One-Hot Encoding

Dalam retail, pola transaksi bisa berbeda-beda tergantung hari. Misalnya, transaksi mungkin ramai di akhir pekan (Sabtu, Minggu). Agar model bisa belajar pola ini, informasi hari perlu diubah ke bentuk angka yang bisa dimengerti mesin inilah fungsinya *One-Hot Encoding*.

Tabel 4 Data *after one-hot encoding*

Senin	Selasa	Rabu	Kamis	Jumat	Sabtu	Minggu
0	0	0	0	1	0	0
0	0	0	0	0	1	0
0	0	0	0	0	0	1

Seperti dipaparkan pada Tabel 4, tiap baris hanya mempunyai satu nilai 1, dan sisanya 0. Artinya, satu transaksi hanya terjadi pada satu hari sehingga model akan mempelajari pengaruh masing-masing hari terhadap profit penjualan.

3.2.4 Pembersihan Data

Pembersihan dilakukan dengan menghapus seluruh baris yang mengandung nilai NaN. Hal ini dilakukan dengan menggunakan fungsi `dropna()` dari pustaka *pandas*. Pendekatan ini dipilih untuk menjaga integritas data yang digunakan oleh model, serta memastikan bahwa setiap baris data memiliki informasi lengkap pada seluruh kolom fitur maupun target.

Namun demikian, data yang digunakan dalam penelitian ini merupakan data transaksi harian dari sistem POS, yang secara operasional mencatat setiap aktivitas penjualan secara real-time. Oleh karena itu, dataset yang diperoleh telah memiliki kualitas yang baik dan tidak mengandung nilai kosong. Proses pembersihan data dengan `dropna()` bersifat sebagai langkah verifikasi tambahan untuk memastikan konsistensi dan reliabilitas data.

3.2.5 Normalisasi Data

Normalisasi dilakukan dengan menggunakan metode *StandardScaler*, yang mengubah data sehingga mempunyai nilai rata-rata (mean) sebesar 0 dan standar deviasi sebesar 1. Transformasi ini diterapkan terlebih dahulu pada data pelatihan (X_{train}) dengan metode `fit_transform`, dan kemudian pada data pengujian (X_{test}) menggunakan `transform`, dengan parameter yang sama untuk menghindari data leakage. Dengan data yang telah dinormalisasi, proses pelatihan menjadi lebih stabil, cepat, dan hasil prediksi model menjadi lebih akurat.

Tabel 5 Sample data sebelum normalisasi

Total Sales	HPP total
295000	227910.36
223900	176516.14
201500	156201.73

Tabel 5 menampilkan sampel data asli sebelum proses normalisasi, di mana nilai pada kolom Total Sales dan HPP_total masih dalam skala besar dan tidak seragam. Nilai-nilai ini mencerminkan skala asli dari data numerik yang bisa sangat bervariasi antar fitur. Skala yang besar dan tidak konsisten dapat memengaruhi performa model machine learning.

Tabel 6 Sample data setelah normalisasi

Total Sales	HPP total
0.150643	0.412287
-0.183424	0.050319
-0.288671	-0.092755

Tabel 6 menunjukkan data yang sama setelah dilakukan proses normalisasi menggunakan metode *StandardScaler*. Setelah normalisasi, nilai-nilai pada masing-masing kolom telah ditransformasikan sehingga memiliki rata-rata mendekati 0 dan standar deviasi mendekati 1. Proses ini penting untuk memastikan bahwa semua fitur memiliki kontribusi yang seimbang dalam proses pelatihan model. Dengan data yang telah dinormalisasi, algoritma dapat belajar lebih efisien dan hasil prediksi menjadi lebih akurat serta stabil.

3.3 Split Dataset

Dalam penelitian ini, dataset sebanyak 44.348 displit dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian yang artinya jumlah data latih sebanyak 35.478 dan 8.870 untuk data uji. Pengujian menggunakan fungsi `train_test_split` dari pustaka *scikit-learn*. Parameter `random_state=42` dipakai untuk memvalidasi bahwa pembagian data bersifat konsisten dan dapat direproduksi. Dengan cara ini, model dapat belajar dari sebagian besar data yang tersedia, sementara sisa data digunakan untuk mengevaluasi kemampuannya dalam melakukan prediksi terhadap data baru yang belum pernah dilihat sebelumnya.

3.4 Pembahasan Model

3.4.1 ANN

Proses perancangan dimulai dengan membangun arsitektur jaringan syaraf tiruan menggunakan modul *Sequential* dari *Keras*. Model terdiri dari tiga lapisan tersembunyi (*hidden layers*) dengan jumlah neuron yang berbeda, yakni 128,

64, dan 32 neuron. Setiap lapisan menggunakan fungsi aktivasi ReLU (*Rectified Linear Unit*) dan diberikan regularisasi L2 dengan nilai 0.01 untuk mengurangi risiko *overfitting*. Selain itu, ditambahkan juga lapisan *Dropout* sebesar 20% setelah dua lapisan pertama untuk menghilangkan beberapa neuron secara acak selama pelatihan, sehingga meningkatkan generalisasi model.

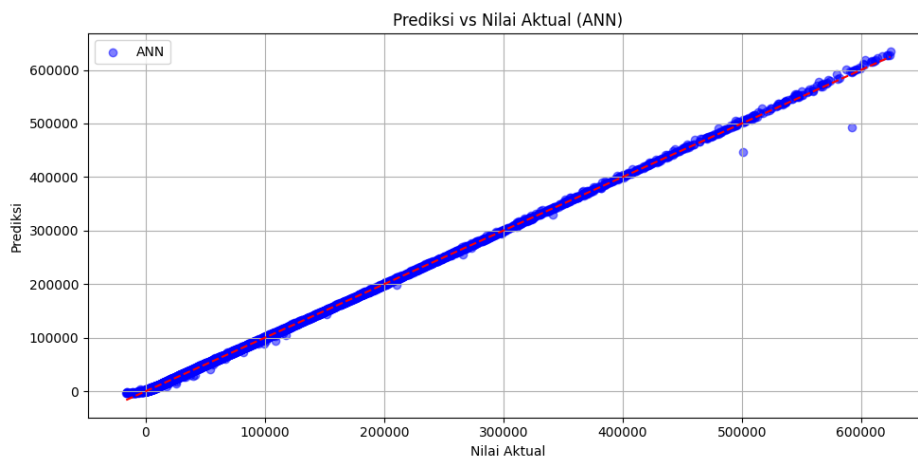
Lapisan output hanya terdiri dari satu neuron karena target yang diprediksi adalah nilai numerik profit (regresi). Model kemudian disusun menggunakan optimizer Adam, yang dikenal efisien dalam pelatihan jaringan saraf, dengan fungsi kerugian *Mean Squared Error* (MSE) yang sesuai untuk permasalahan regresi. Untuk mencegah *overfitting*, digunakan *Early Stopping*, yaitu teknik yang secara otomatis pelatihan berhenti jika tidak ada peningkatan pada nilai validasi loss selama 10 epoch berturut-turut, sekaligus membentuk bobot terbaik yang pernah dicapai selama pelatihan.

Model kemudian dilatih selama maksimal 50 epoch dengan pembagian data pelatihan sebanyak 35.478 dan 8.870 dari data pelatihan digunakan sebagai data validasi. Ukuran batch ditetapkan sebesar 32. Pelatihan selesai di proses, model dipakai untuk melakukan prediksi terhadap data pengujian (*X_test_scaled*), sehingga menghasilkan output prediksi nilai profit (*y_pred*) yang akan digunakan pada tahap evaluasi model. Pendekatan ini memastikan bahwa model ANN yang dirancang tidak hanya kompleks, tetapi juga memiliki kesanggupan generalisasi yang efektif terhadap data baru.



Gambar 3 Grafik training loss vs validation loss

Gambar 3 diatas memperlihatkan bahwa model ANN berhasil belajar secara efektif dari data. Pada awal pelatihan, nilai loss menurun tajam, menandakan proses pembelajaran yang aktif. Setelah sekitar epoch ke-10, baik nilai training loss maupun validation loss menunjukkan tren yang stabil dengan selisih yang kecil, mengindikasikan bahwa model tidak mengalami *overfitting* dan mampu melakukan generalisasi dengan baik. Hal ini juga menunjukkan bahwa penggunaan regularisasi dan strategi *early stopping* telah berhasil mengoptimalkan kinerja model selama proses pelatihan.



Gambar 4 Visualisasi grafik prediksi vs nilai aktual (ANN)

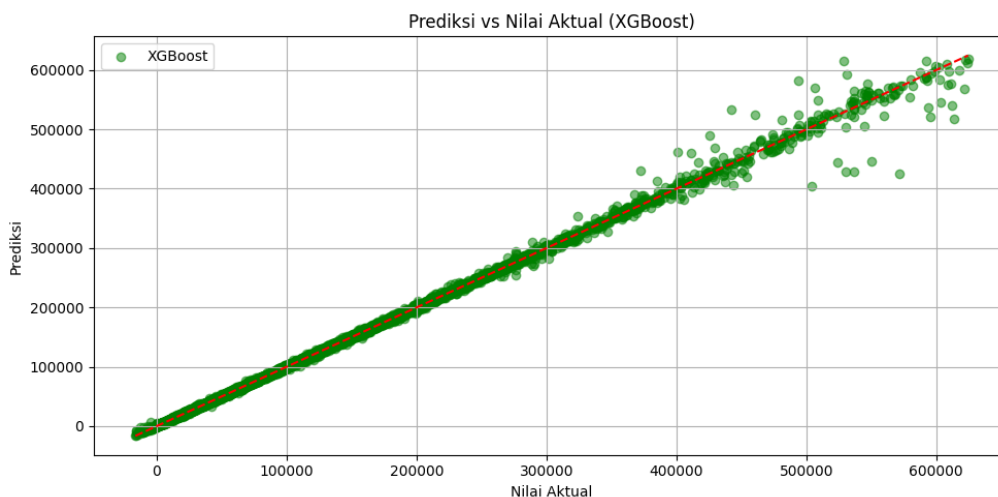
Dari Gambar 4 menunjukkan bahwa titik-titik prediksi (berwarna biru) tersebar sangat dekat dengan garis diagonal merah yang merepresentasikan garis idealnya. Kedekatan titik-titik ini terhadap garis referensi menandakan bahwa hasil prediksi model sangat akurat dan mendekati nilai aktual profit dari transaksi POS. Hanya sedikit titik yang tampak menyimpang secara signifikan, yang memperlihatkan bahwa model memiliki keunggulan prediksi yang sangat efektif secara keseluruhan. Ini mengindikasikan bahwa ANN mampu mempelajari pola dari data dengan sangat efektif.

3.4.2 XGBoost

Pada tahap ini, model XGBoost diterapkan untuk melakukan prediksi nilai profit dari data transaksi POS. Model dikonfigurasi menggunakan parameter utama seperti `n_estimators=200` yang menunjukkan jumlah pohon keputusan (trees) yang digunakan, `learning_rate=0.1` untuk mengatur kontribusi setiap pohon terhadap prediksi akhir, serta `max_depth=10` yang mengontrol kedalaman maksimum tiap pohon dalam model. Untuk menjaga konsistensi hasil, `random_state=42` digunakan sebagai nilai seed acak.

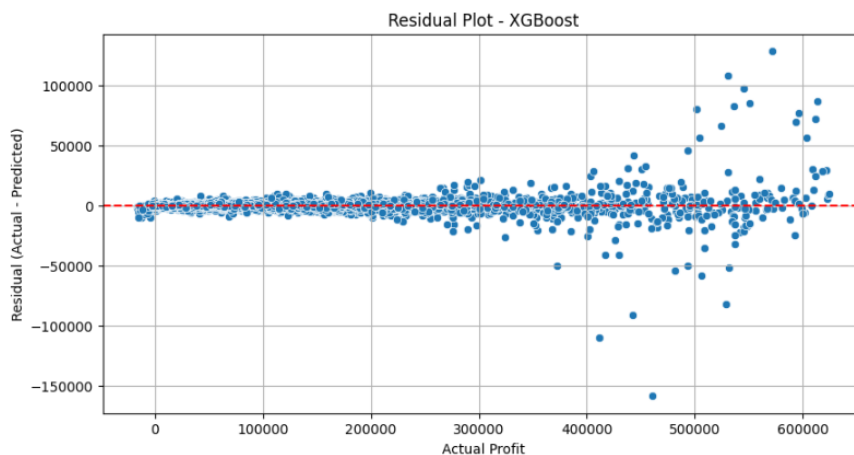
Model kemudian dilatih memakai data pelatihan yang telah dinormalisasi (`X_train_scaled` dan `y_train`) melalui fungsi `.fit()`. Setelah proses pelatihan selesai, model dipakai untuk memberikan prediksi pada data pengujian (`X_test_scaled`) melalui fungsi `.predict()`, dan hasilnya disimpan ke dalam variabel `xgb_pred`.

Model XGBoost dipilih karena kemampuannya yang unggul dalam menangani data numerik, ketahanan terhadap overfitting, serta kecepatan komputasi yang tinggi dibandingkan metode boosting lainnya, menjadikannya cocok untuk memprediksi nilai profit dengan akurasi yang tinggi.



Gambar 5 Visualisasi grafik prediksi vs aktual (XGBoost)

Berdasarkan Gambar 5 menunjukkan hubungan antara nilai prediksi dan nilai aktual dari model XGBoost dalam memprediksi profit penjualan. Titik-titik hijau mewakili hasil prediksi terhadap nilai aktual, sedangkan garis merah merepresentasikan garis ideal di mana prediksi dan nilai aktual bernilai sama ($y = x$). Sebagian besar titik data berada cukup dekat dengan garis merah, menandakan bahwa model mampu memberikan prediksi yang cukup akurat. Namun, terlihat beberapa titik yang menyimpang, khususnya pada nilai profit yang tinggi, mengindikasikan bahwa model masih memiliki sedikit kesalahan dalam memprediksi transaksi dengan nominal besar.



Gambar 6 Residual XGBoost

Gambar 6 adalah *residual plot* dari model XGBoost, yang menunjukkan selisih antara nilai aktual dan nilai prediksi (residual) terhadap nilai aktual profit. Titik-titik biru merepresentasikan masing-masing data, sementara garis merah putus-putus di tengah menunjukkan garis nol sebagai referensi residual ideal. Sebagian besar titik berada di sekitar garis nol, menandakan bahwa prediksi model cukup akurat. Namun, terdapat penyebaran residual yang mulai membesar pada nilai profit yang tinggi, yang berarti model mulai kesulitan dalam memprediksi data dengan nilai besar secara akurat. Pola penyebaran ini menunjukkan bahwa meskipun model cukup baik secara umum, masih terdapat potensi peningkatan, khususnya dalam menangani outlier atau nilai transaksi yang ekstrem

3.5 Evaluasi Model

Evaluasi dilakukan terhadap dua model regresi yaitu Artificial Neural Network (ANN) dan XGBoost dengan menggunakan metrik evaluasi berupa *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (*R-squared* atau R^2). Berdasarkan hasil yang diperoleh, model ANN menunjukkan kinerja yang lebih baik dari XGBoost dalam memprediksi nilai profit dari data transaksi ritel seperti pada tabel 7 dibawah ini.

Tabel 7 Hasil evaluasi model

Metrik Evaluasi	(ANN)	XGBoost
MAE	1,359.02	1,600.18
MSE	4,635,848.60	26,218,337.66
RMSE	2,153.10	5,120.38
R^2	0.9996	0.9977

Berdasarkan hasil evaluasi model pada tabel 7, diketahui bahwa model ANN memiliki performa yang lebih baik dibandingkan dengan model XGBoost. Hal ini terlihat dari nilai MAE, MSE, dan RMSE yang lebih kecil pada ANN, serta nilai R^2 yang lebih tinggi. Model ANN mencatat nilai MAE sebesar 1.359,02, yang berarti rata-rata kesalahan prediksi hanya sekitar 1.359 satuan dari nilai sebenarnya. Sementara itu, XGBoost memiliki MAE sebesar 1.600,18, yang menunjukkan bahwa prediksinya sedikit lebih meleset. Begitu juga pada nilai MSE dan RMSE, yang mengukur kesalahan kuadrat dan akar kesalahan, ANN memiliki nilai yang jauh lebih rendah dibandingkan XGBoost, menandakan bahwa prediksi ANN lebih konsisten dan akurat. Selain itu, nilai R^2 dari ANN adalah 0,9996, mendekati angka sempurna 1, yang berarti model ini sangat mampu menjelaskan variasi data target. Sebagai perbandingan, XGBoost juga memiliki R^2 yang tinggi yaitu 0,9977, namun masih sedikit di bawah ANN.

Keunggulan utama dari ANN adalah kemampuannya dalam menangkap pola yang kompleks dan hubungan non-linier dalam data. Ini menjadikan ANN sangat cocok digunakan pada data yang rumit dan memiliki banyak variabel yang saling berinteraksi. Namun, ANN juga memerlukan proses pelatihan yang lebih lama dan harus melalui proses normalisasi serta pengaturan arsitektur jaringan yang tepat. Di sisi lain, XGBoost lebih cepat dan mudah diterapkan untuk data tabular, serta tidak terlalu bergantung pada proses normalisasi data. Meskipun hasilnya tidak seakurat ANN dalam studi ini, XGBoost tetap menunjukkan performa yang sangat baik dan stabil.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan Model ANN menghasilkan nilai MAE sebesar 1,359.02, MSE sebesar 4,635,848.60, dan RMSE sebesar 2,153.10, dengan nilai R^2 mencapai 0.9996. Nilai-nilai tersebut menunjukkan bahwa prediksi yang dilakukan oleh model ANN sangat mendekati nilai aktual dengan tingkat kesalahan rata-rata yang kecil. Nilai R^2 yang hampir sempurna (mendekati 1) mengindikasikan bahwa hampir seluruh variasi pada data target dapat dijelaskan oleh model ini, sehingga dapat dikatakan bahwa ANN memiliki *fit* yang sangat baik terhadap data. Sebaliknya, model XGBoost menghasilkan nilai MAE sebesar 1,600.18, MSE sebesar 26,218,337.66, dan RMSE sebesar 5,120.38, dengan R^2 sebesar 0.9977. Meskipun nilai R^2 XGBoost juga cukup tinggi dan berada dalam kategori sangat baik, performanya masih berada di bawah ANN. Selisih nilai RMSE dan MSE yang jauh lebih besar pada XGBoost menunjukkan bahwa model ini cenderung membuat kesalahan yang lebih besar dalam memprediksi nilai profit dibandingkan ANN. Perbedaan performa ini dapat disebabkan oleh beberapa faktor seperti kompleksitas model, kemampuan dalam menangani data non-linear, dan teknik regularisasi. ANN yang memiliki arsitektur multi-layer dan fungsi aktivasi nonlinear lebih unggul dalam menangkap pola kompleks dalam data, sedangkan XGBoost yang berbasis pada ensemble tree dapat mengalami kesulitan dalam memodelkan pola kontinu yang sangat presisi seperti pada kasus ini. Secara keseluruhan, penggunaan model prediktif berbasis machine learning terbukti efektif dalam menganalisis data transaksi retail dan memberikan estimasi penjualan yang akurat, yang dapat digunakan sebagai pedoman pengambilan keputusan operasional dan strategis dalam bisnis retail.

REFERENCES

- [1] R. Putri, D. Prasetya, R. N. Azizah, J. Bayu, W. Halwa, and R. H. Nugroho, "Implementasi Penggunaan Data Analytics untuk Mengoptimalkan Pengambilan Keputusan Bisnis di Era Digital," *jbkd*, vol. 2, no. 2, pp. 1–12, 2025, doi: 10.47134/jbkd.v2i2.3459.
- [2] I. A. M. Basid, N. Islamiyah, R. A. Zuleika, A. Inka, S. Andriani, and I. R. Kusumasari, "Peran Teknologi Informasi dalam



- Perencanaan dan Pengembangan Bisnis di Era Digital: Tantangan dan Peluang,” *Econ. Bus. Manag. J.*, vol. 3, no. 1, pp. 1–6, 2024, <https://www.ejournal-rmg.org/index.php/EBMJ/article/download/201/243>
- [3] J. Manajemen and U. S. Sulistyawati, “Decoding Big Data : Mengubah Data Menjadi Keunggulan Kompetitif dalam Pengambilan Keputusan Bisnis Abstrak,” *jmt*, vol. 1, no. 2, pp. 58–71, 2024, doi: 10.63447/jmt.v1i2.1114.
- [4] F. A. M. Nurlaelatul, “Journal of Artificial Intelligence and Digital Business (RIGGS) Prediksi Harga Properti Di Indonesia Menggunakan Algoritma Random,” *RIGGS*, vol. 4, no. 1, pp. 43–49, 2025, doi: 10.31004/riggs.v4i1.367.
- [5] A. I. Sakti *et al.*, “Implementasi Artificial Neural Network (ANN) dalam Memprediksi Nilai Tukar Rupiah terhadap Dolar Amerika Implementasi Artificial Neural Network (ANN) dalam Memprediksi Nilai Tukar Rupiah terhadap Dolar Amerika,” *EULER*, vol. 12, no. 2, pp. 124–130, 2024, doi: 10.37905/euler.v12i2.26654.
- [6] I. Gunawan, “Optimasi Model Artificial Neural Network untuk Klasifikasi Paket Jaringan,” *Simetris*, vol. 14, no. 2, pp. 1–5, 2020, doi: 10.51901/simetris.v14i2.135.
- [7] A. Haris *et al.*, “Improving Prediction Accuracy Of Imported Goods Prices Using,” vol. 12, no. 2, 2025, doi: 10.25126/jtiik.2025129419.
- [8] J. Fang, H. Wang, F. Yang, K. Yin, X. Lin, and M. Zhang, “A failure prediction method of power distribution network based on PSO and XGBoost,” *Aust. J. Electr. Electron. Eng.*, vol. 19, no. 4, pp. 371–378, 2022, doi: 10.1080/1448837X.2022.2072447.
- [9] W. Kurniawan and U. Indahyanti, “Prediksi Angka Harapan Hidup Penduduk Menggunakan Metode XGBoost,” *Indones. J. Appl. Technol.*, vol. 1, no. 2, p. 18, 2024, doi: 10.47134/ijat.v1i2.3045.
- [10] H. Wibiksana, “Prediksi Harga Saham pada Portofolio Investor dengan Analisis Time Series Harga Saham menggunakan ANN,” *Bull. Comput. Sci. Res.*, vol. 4, no. 1, pp. 10–17, 2023, doi: 10.47065/bulletincsr.v4i1.298.
- [11] R. Awalia and F. Titik Kristanti, “Prediksi Financial Distress Menggunakan Artificial Neural Network (Ann) Pada Perusahaan Perbankan Yang Terdaftar Di Bursa Efek Indonesia (Bei) Periode 2017-2021,” *J. Kaji. Pendidik. Ekon. dan Ilmu Ekon.*, vol. 7, no. 2, pp. 2549–2284, 2023, doi: 10.23969/oikos.v7i2.7924.
- [12] H. Wijaya, D. P. Hostiadi, and E. Triandini, “Meningkatkan Prediksi Penjualan Retail Xyz Dengan Teknik Optimasi Random Search Pada Model Xgboost,” *SPINTER (Prosiding Semin. Has. Penelit. Inform. dan Komputer)*, vol. 1, no. 2, pp. 829–833, 2024, [Online]. Available: <https://spinter.stikom-bali.ac.id/index.php/spinter/article/view/226>
- [13] S. T. Nurmaya, B. Soesilo, and D. A. Fatah, “Prediksi Harga Saham Pada Bank Bca Tbk Dengan Teknik Xgboost (Extreme Gradient Boosting),” *JURSIMA*, 2023, doi: 10.47024/js.v1i2i2.925.
- [14] S. Suparmadi and A. Ramadhani, “Sistem Estimasi Pencapaian Target Profit Menggunakan Model Regresi Berbasis Machine Learning,” *J. Sci. Soc. Res.*, vol. 5, no. 3, p. 703, 2022, doi: 10.54314/jssr.v5i3.1042.
- [15] R. Winurputra and D. E. Ratnawati, “Peramalan Penjualan Produk Menggunakan Extreme Gradient Boosting (Xgboost) Dan Kerangka Kerja Crisp-Dm Untuk Pengoptimalan Manajemen Persediaan (Studi Kasus : Ub Mart),” *JTIK*, vol. 12, no. 2, pp. 417–428, 2025, doi: 10.25126/jtiik.2025129451.
- [16] D. Pradana, M. Luthfi Alghifari, M. Farhan Juna, and D. Palaguna, “Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network,” *Indones. J. Data Sci.*, vol. 3, no. 2, pp. 55–60, 2022, doi: 10.56705/ijodas.v3i2.35.
- [17] A. Khumaidi, R. Raafi’udin, and I. P. Solihin, “Pengujiann Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung,” *J. Telemat.*, vol. 15, no. 1, pp. 13–18, 2020, doi: 10.61769/telematika.v15i1.340.
- [18] D. Satriani and V. V. Kusuma, “Perhitungan harga pokok produksi dan harga pokok penjualan terhadap laba penjualan,” *J. Ilm. MEA (Manajemen, Ekon. dan Akuntansi)*, vol. 4, no. 2, p. 438, 2020, [Online]. Available: <https://journal.universitassuryadarma.ac.id/index.php/jtin/article/view/645>
- [19] C. Herdian, A. Kamila, and I. G. Agung Musa Budidarma, “Studi Kasus Feature Engineering Untuk Data Teks: Perbandingan Label Encoding dan One-Hot Encoding Pada Metode Linear Regresi,” *Technol. J. Ilm.*, vol. 15, no. 1, p. 93, 2024, doi: 10.31602/tji.v15i1.13457.
- [20] J. Elektronika and D. A. N. Komputer, “Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI,” *ELKOM*, vol. 17, no. 2, pp. 657–666, 2024, doi: 10.51903/elkom.v17i2.2311.
- [21] F. W. Atmojo *et al.*, “Analisis Pemanfaatan Machine Learning Guna Prediksi Indeks,” *Simtek J. Sist. Inf. Dan Tek. Komput.*, vol. 9, no. 2, 2024, doi: 10.51876/simtek.v9i2.390.
- [22] A. H. Nasrullah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris,” *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021, doi: 10.35329/jiik.v7i2.203.
- [23] C. Sandi, M. B. Adityawan, D. Harlan, M. Farid, and N. Nadeak, “Artificial Neural Network dan Pemodelan Numerik untuk Prediksi Parameter Aliran akibat Dam Break,” *J. Tek. Sumber Daya Air*, vol. 2, no. 2, pp. 129–140, 2022, doi: 10.56860/jtsda.v2i2.50.
- [24] G. Chairunisa *et al.*, “Life Expectancy Prediction Using Decision Tree, Random Forest, Gradient Boosting, and XGBoost Regressions,” *J. Sintak*, vol. 2, no. 2, pp. 71–82, 2024, doi: 10.62375/jsintak.v2i2.249.
- [25] P. Septiana Rizky, R. Haiban Hirzi, and U. Hidayaturrohman, “Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang,” *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 15, no. 2, pp. 228–236, 2022, doi: 10.36456/jstat.vol15.no2.a5548.