

Pengembangan Model Prediksi Penjualan Mingguan pada Perusahaan Ritel Walmart Berbasis Artificial Neural Network

(Development of a Weekly Sales Prediction Model at Walmart Retail Company Based on Artificial Neural Network)

Gieska N. Salamena^{1*}, Vyarlita I. Fataruba², dan Shela Tappa³

^{1,2,3}Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Pattimura
Jl. Ir. M. Putuhena, Ambon, 97233, Indonesia

*Corresponding author's e-mail: * gieskasalamena03@gmail.com

Manuscript submitted:
17th February 2025

Manuscript revision:
20th March 2025

Accepted for publication:
29th May 2025

Abstract

Sales prediction is a crucial element in the modern retail industry, as it directly influences decision-making in inventory management, marketing strategy planning, and product distribution. This study aims to develop a weekly sales prediction model for Walmart stores using the Artificial Neural Network approach. The dataset comprises 5,508 historical records that have been cleaned from outliers and includes several variables such as temperature, fuel price, Consumer Price Index, unemployment rate, and holiday indicators. The model is designed using a Sequential architecture with hyperparameter optimization involving the number of neurons and learning rate. Model performance is evaluated using the Mean Squared Error metric.

Keywords: Artificial Neural Network; Deep Learning; Mean Squared Error; Sales Prediction; Walmart.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

How to cite this article:

G. Salamena, V. Fataruba, and S. Tappa, "Pengembangan Model Prediksi Penjualan Mingguan Pada Perusahaan Ritel Walmart Berbasis Artificial Neural Network", *algorithm*, vol. 1, no. 1, pp. 1-8, May 2025.

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong pemanfaatan data besar (big data) dalam pengambilan keputusan bisnis. Dalam sektor ritel, akurasi dalam memprediksi permintaan pasar menjadi elemen penting. Walmart, sebagai salah satu pengecer terbesar di dunia, memiliki data historis penjualan yang kompleks dan membutuhkan pendekatan non-linear seperti Artificial Neural Network (ANN) untuk memberikan prediksi yang lebih tepat. Kemajuan teknologi digital telah memberikan pengaruh besar terhadap berbagai aspek kehidupan, termasuk sektor perdagangan dan keuangan. Otomatisasi dan pemanfaatan data memungkinkan organisasi untuk menghimpun serta memanfaatkan informasi dalam skala besar, termasuk data operasional dan preferensi pelanggan. Di era data masif ini, pemrosesan informasi menjadi elemen krusial dalam merumuskan keputusan yang bersifat strategis. Salah satu permasalahan utama dalam dunia perdagangan eceran adalah fluktuasi permintaan pasar. Keakuratan dalam memproyeksikan permintaan serta angka penjualan sangat dibutuhkan, karena kesalahan dalam estimasi dapat menimbulkan konsekuensi seperti stok yang berlebih, kekurangan produk, hingga kerugian secara ekonomi [1]. Ketepatan dalam estimasi penjualan juga berpengaruh terhadap keberhasilan strategi promosi, distribusi barang, dan efisiensi manajemen perusahaan.

Walmart sebagai jaringan ritel berskala global, merupakan contoh perusahaan yang memiliki catatan transaksi historis yang sangat besar dan beragam. Informasi tersebut mencakup aspek seperti harga energi, tingkat ketenagakerjaan, hari besar nasional, serta indeks ekonomi seperti Consumer Price Index. Kompleksitas serta sifat non-linier dari data tersebut menyebabkan pendekatan statistik konvensional seperti model regresi linier seringkali kurang efektif dalam menangkap pola mendalam secara tepat [2].

Dengan meningkatnya variasi serta volume informasi, diperlukan pendekatan yang responsif dan fleksibel, salah satunya melalui teknik pembelajaran mesin (machine learning). Salah satu algoritma yang menunjukkan kinerja baik dalam memetakan relasi kompleks antar fitur adalah Artificial Neural Network. JST dapat merepresentasikan pola non-linier dan mampu melakukan prediksi terhadap data baru yang belum pernah dipelajari sebelumnya, menjadikannya relevan untuk skenario prediksi dalam bisnis yang terus berubah [3] [4].

Selain itu, model berbasis pembelajaran mendalam juga telah menunjukkan performa tinggi dalam mengolah data berskala besar, khususnya untuk tugas seperti estimasi permintaan dan optimasi logistik [5]. Namun demikian, ANN masih menjadi pendekatan yang cocok dalam situasi di mana peramalan sangat dipengaruhi oleh variabel ekonomi eksternal serta struktur data yang khas milik perusahaan [6].

Berdasarkan kondisi tersebut, penelitian ini menitikberatkan pada penggunaan *Artificial Neural Network* (ANN) dalam melakukan estimasi penjualan mingguan menggunakan data dari Walmart. Penelitian ini bertujuan untuk mengukur seberapa baik model ANN dapat menghasilkan prediksi yang akurat, serta membandingkan kinerjanya dengan metode konvensional. Dengan memanfaatkan data historis yang telah melalui proses pembersihan dari nilai pencilan, diharapkan model yang dibangun dapat memberikan wawasan yang bermanfaat dalam perencanaan dan strategi bisnis ritel.

Penelitian ini memiliki tujuan untuk merancang sebuah model prediksi terhadap penjualan mingguan pada perusahaan ritel Walmart dengan menerapkan pendekatan *Artificial Neural Network* (ANN). Penelitian ini secara khusus ditujukan untuk menganalisis kemampuan ANN dalam memetakan pola non-linier yang kompleks dari hubungan antara berbagai variabel seperti suhu, harga bahan bakar, indeks harga konsumen (CPI), tingkat pengangguran, serta hari libur terhadap total penjualan. Selain itu, penelitian ini juga mencakup proses optimasi *hyperparameter* pada model guna memperoleh hasil prediksi yang optimal, sekaligus membandingkannya dengan performa model sebelumnya yang memanfaatkan algoritma Random Forest.

Dengan menggunakan pendekatan ini, penelitian diharapkan mampu memberikan kontribusi signifikan dalam meningkatkan ketepatan peramalan penjualan yang sangat penting untuk mendukung keputusan bisnis strategis, seperti pengelolaan stok, aktivitas promosi, dan perencanaan distribusi. Lebih jauh lagi, hasil dari penelitian ini juga dapat dijadikan referensi bagi akademisi dan praktisi dalam mengaplikasikan metode *deep learning* pada data ritel yang bersifat kompleks dan dinamis, serta memperkuat pemanfaatan kecerdasan buatan sebagai solusi inovatif dalam menghadapi tantangan era *big data*.

2. METODE PENELITIAN

Pada bagian ini dijelaskan tahapan-tahapan yang dilakukan dalam proses penelitian guna membangun model prediksi penjualan mingguan pada toko-toko Walmart menggunakan algoritma Artificial Neural Network (ANN). Penelitian ini terdiri atas proses pembersihan data, eksplorasi fitur, pemilihan model, serta evaluasi kinerja melalui metrik kesalahan prediksi.

2.1. Pengumpulan dan Persiapan Data

Dataset yang digunakan merupakan data penjualan mingguan historis dari 45 gerai Walmart di Amerika Serikat, bersumber dari platform Kaggle. Data ini mencakup 5.508 entri yang mencerminkan laporan penjualan mingguan, serta dilengkapi dengan berbagai variabel eksternal seperti suhu, harga bahan bakar, indeks harga konsumen (*Consumer Price Index/CPI*), tingkat pengangguran, dan indikator hari libur.

Sebelum dilakukan analisis, data telah melalui tahap pembersihan dengan menghapus nilai-nilai pencilan (*outlier*) yang berpotensi mengurangi akurasi model. Setiap entri dalam dataset merepresentasikan situasi ekonomi dan operasional toko pada periode mingguan tertentu.

2.2. Eksplorasi Data dan Analisis Korelasi

Analisis korelasi dilakukan guna mengidentifikasi hubungan antara variabel-variabel yang ada. Hasil analisis menunjukkan bahwa variabel *store* memiliki korelasi negatif sedang terhadap *weekly_sales* ($r = -0.32$). Sementara itu, *temperature*, *cpi*, dan *unemployment* memiliki korelasi yang lemah terhadap variabel target. Korelasi tertinggi antar fitur ditemukan antara *temperature* dan *fuel_price* ($r = 0.12$), serta antara *cpi* dan temperatur ($r = 0.22$), yang mengindikasikan bahwa hubungan linier antar fitur eksternal tergolong terbatas.

2.3. Perancangan Model ANN

Model dikembangkan dengan pendekatan *Sequential Neural Network*, di mana arsitekturnya terdiri dari lapisan masukan yang menerima delapan fitur numerik sebagai input, diikuti oleh satu atau dua lapisan tersembunyi (*hidden layer*) dengan jumlah neuron yang bervariasi, serta diakhiri dengan satu lapisan keluaran (*output layer*) yang bertugas memprediksi nilai penjualan mingguan. Untuk mendapatkan performa terbaik, dilakukan serangkaian eksperimen dengan menguji berbagai kombinasi konfigurasi *hyperparameter*. Jumlah neuron yang digunakan dalam *hidden layer* diuji dengan nilai 32, 64, dan 128. Proses pelatihan dilakukan dengan dua pilihan *learning rate*, yaitu 0.001 dan 0.01, sambil mempertahankan penggunaan fungsi aktivasi ReLU pada setiap lapisan tersembunyi. Model dioptimasi menggunakan algoritma Adam, dan evaluasi performa dilakukan dengan menggunakan fungsi kerugian Mean Squared Error (MSE).

2.4. Proses Pelatihan Model

Dataset dibagi ke dalam dua bagian utama, yaitu data pelatihan dan data pengujian, dengan proporsi sebesar 80:20. Model kemudian dilatih selama sejumlah *epoch*, di mana proses pelatihan disertai dengan validasi secara simultan guna meminimalkan risiko terjadinya *overfitting* pada model. Berdasarkan hasil eksperimen yang telah dilakukan, diperoleh bahwa konfigurasi model

terbaik dicapai ketika menggunakan 128 neuron dengan nilai learning rate sebesar 0.01. Pada konfigurasi tersebut, model berhasil menghasilkan nilai *Mean Squared Error* (MSE) sebesar 232.283.171.177,98.

2.5. Evaluasi Kinerja Model

Evaluasi performa model dilakukan dengan menggunakan metrik *Mean Squared Error* (MSE) serta divisualisasikan dengan grafik *loss function*. Berdasarkan grafik yang dihasilkan, terlihat bahwa nilai *loss* mengalami penurunan secara konsisten dan stabil sepanjang proses pelatihan. Tidak ditemukan indikasi *overfitting*, yang menandakan bahwa model memiliki kemampuan generalisasi yang baik terhadap data baru di luar data pelatihan. Pelatihan model dilakukan selama 100 *epoch*, sehingga memberikan kesempatan yang cukup bagi model untuk belajar secara optimal tanpa kehilangan kemampuan generalisasinya.

Tabel 1. Konfigurasi Model ANN dan Hasil Evaluasi MSE

Jumlah Neuron	Learning rate	Mean Squared Error (MSE)
32	0.001	633.008.611.099,81
32	0.01	254.465.622.799,16
64	0.001	306.179.407.286,24
64	0.01	241.817.246.805,62
128	0.001	280.296.744.740,36
128	0.01	232.283.171.177,98 (terbaik)

3. HASIL DAN PEMBAHASAN

Bagian ini secara khusus menyajikan hasil yang diperoleh dari proses pelatihan dan evaluasi model *Artificial Neural Network* (ANN) yang digunakan untuk memprediksi nilai penjualan mingguan. Hasil yang ditampilkan mencakup performa model selama pelatihan serta kemampuannya dalam memprediksi data yang belum pernah dilihat sebelumnya.

3.1. Pemrosesan Data

Penelitian ini memiliki tujuan utama untuk melakukan prediksi terhadap nilai penjualan mingguan (*weekly sales*) pada toko-toko Walmart dengan memanfaatkan sejumlah variabel eksternal yang dapat mempengaruhi tingkat penjualan. Variabel-variabel eksternal tersebut meliputi temperatur lingkungan (*temperature*), harga bahan bakar (*fuel price*), indeks harga konsumen atau *Consumer Price Index* (CPI), serta tingkat pengangguran (*unemployment*). Dalam proses analisis, digunakan dataset yang terdiri dari sebanyak 5508 entri data yang sebelumnya telah melalui proses pembersihan untuk menghilangkan nilai-nilai outlier yang dapat mengganggu keakuratan model prediksi.

Dataset ini mencakup periode waktu dari tahun 2010 hingga 2012, yang merepresentasikan catatan penjualan mingguan di berbagai cabang Walmart selama rentang waktu tersebut. Setiap baris data pada dataset ini merepresentasikan catatan penjualan mingguan di salah satu cabang atau toko Walmart tertentu, yang juga disertai dengan informasi waktu serta kondisi-kondisi eksternal saat penjualan tersebut terjadi. Dengan demikian, masing-masing data tidak hanya menunjukkan seberapa besar *weekly sales* yang terjadi, namun juga mencerminkan situasi ekonomi dan lingkungan pada periode tersebut. Penjelasan lebih rinci mengenai setiap fitur dalam dataset disajikan pada Tabel 2.

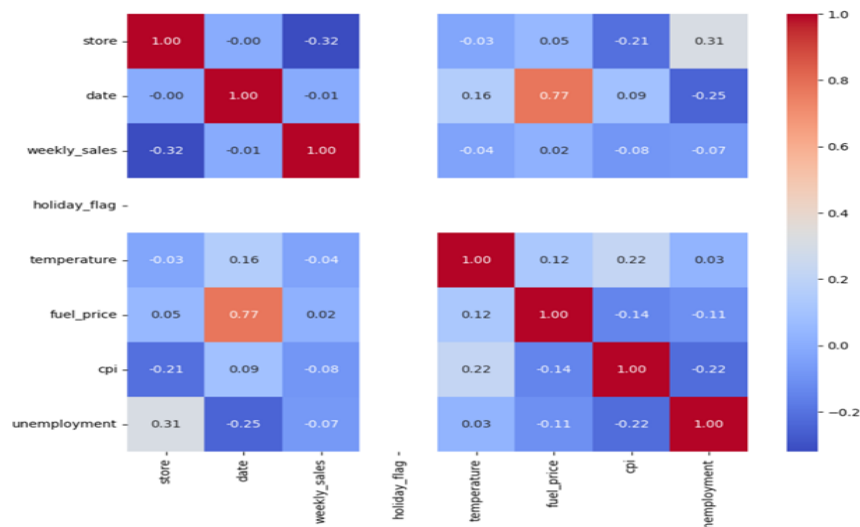
Tabel 2. Rincian Fitur pada Dataset

<i>Fitur</i>	<i>Deskripsi</i>
<i>store</i>	Kode identifikasi numerik dari setiap toko Walmart, berkisar dari toko nomor 1 hingga toko nomor 45.
<i>date</i>	Tanggal dilakukannya pencatatan penjualan mingguan.
<i>weekly_sales</i>	Total nilai penjualan yang tercatat dalam periode satu minggu.
<i>holiday_flag</i>	Indikator biner yang menyatakan apakah minggu tersebut mengandung hari libur nasional (1) atau tidak (0).
<i>temperature</i>	Suhu rata-rata lingkungan sekitar pada minggu tersebut, dinyatakan dalam satuan Fahrenheit.
<i>fuel_price</i>	Harga bahan bakar yang berlaku pada saat itu.
<i>cpi</i>	Nilai indeks harga konsumen sebagai indikator ekonomi makro.
<i>unemployment</i>	Persentase tingkat pengangguran pada waktu tertentu.

Adapun untuk variabel target, yaitu *weekly_sales*, tercatat memiliki variasi nilai yang cukup besar, yakni mulai dari sekitar 209.986 dolar hingga mencapai nilai maksimum sebesar 2.678.206 dolar. Rata-rata penjualan mingguan secara keseluruhan mencapai angka sekitar 1.035.645 dolar, yang menunjukkan besarnya skala operasional dan potensi keuntungan dari masing-masing cabang Walmart yang diamati.

3.2. Analisis Korelasi

Analisis korelasi antar fitur menunjukkan bahwa variabel *store* memiliki korelasi negatif sedang terhadap *weekly_sales* dengan nilai korelasi sebesar -0.32, sedangkan fitur *temperature*, *cpi*, dan *unemployment* menunjukkan hubungan yang tergolong lemah terhadap *weekly_sales*, dan korelasi tertinggi antar fitur ditemukan antara *temperature* dan *fuel_price* sebesar 0.12 serta antara *cpi* dan *temperature* sebesar 0.22, yang seluruhnya divisualisasikan melalui matriks korelasi, dapat dilihat pada Gambar 1.

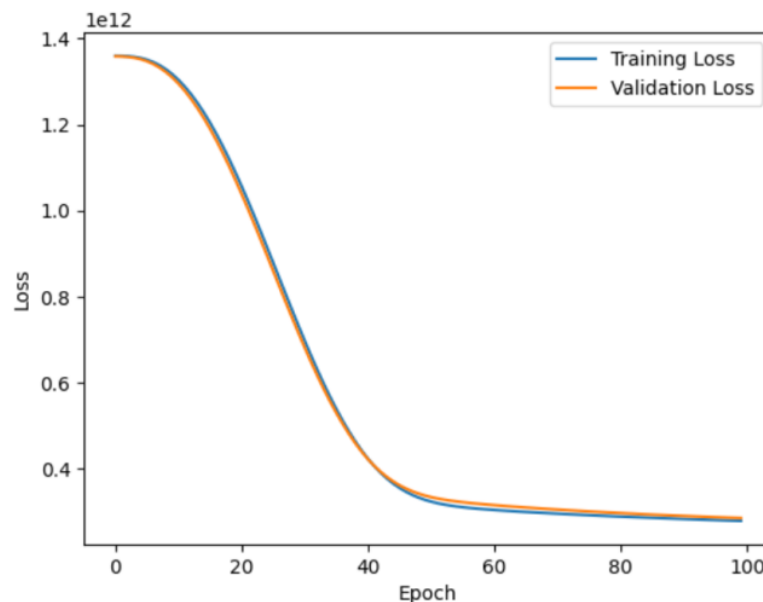
**Gambar 1.** Heatmap Korelasi Antar Variabel

3.3. Training Model

Model ini dilatih menggunakan dataset *training* yang diambil dari keseluruhan data. Dataset dibagi menjadi dua bagian, yaitu 80% untuk data pelatihan (*training set*) dan 20% untuk data

pengujian (testing set), guna memastikan bahwa model dapat diuji terhadap data yang tidak pernah dilihat sebelumnya. Hasil pelatihan menegaskan bahwa konfigurasi terbaik diperoleh ketika menggunakan 128 *neuron* dengan *learning rate* sebesar 0.01, menghasilkan nilai MSE sebesar 232.283.171.177,98. Nilai tersebut mengindikasikan kemampuan model dalam memprediksi penjualan mingguan dengan tingkat akurasi yang relatif tinggi, terutama mengingat bahwa nilai penjualan dapat mencapai jutaan dolar per minggu. Perbandingan hasil dari beberapa konfigurasi lainnya juga telah diujikan, antara lain konfigurasi dengan 32 *neuron* dan *learning rate* 0.001 memberikan MSE sebesar 633.008.611.099,81, sedangkan untuk konfigurasi dengan 32 *neuron* dan *learning rate* 0.01, nilai MSE menurun menjadi 254.465.622.799,16. Untuk konfigurasi dengan 64 *neuron*, ketika menggunakan *learning rate* 0.001 diperoleh nilai MSE sebesar 306.179.407.286,24, sedangkan dengan *learning rate* 0.01, nilai MSE mencapai 241.817.246.805,62. Selain itu, konfigurasi yang menggunakan 128 *neuron* namun dengan *learning rate* 0.001 menghasilkan nilai MSE sebesar 280.296.744.740,36. Berdasarkan perbandingan tersebut, konfigurasi dengan 128 *neuron* dan *learning rate* 0.01 terbukti memberikan error kuadrat rata-rata yang paling rendah jika dibandingkan dengan konfigurasi-konfigurasi lainnya.

Model dengan performa terbaik ini diharapkan mampu membantu manajemen dalam melakukan perencanaan persediaan dan strategi pemasaran yang lebih tepat berdasarkan perkiraan penjualan. Untuk memperjelas performa model selama proses pelatihan, grafik penurunan nilai *loss* terhadap jumlah *epoch* dapat dilihat pada Gambar 2.



Gambar 2. Grafik Nilai *Loss* Terhadap Jumlah *Epoch*

Berdasarkan grafik yang ditampilkan, dapat diamati bahwa kurva *training loss* maupun *validation loss* mengalami penurunan secara bertahap dan konsisten, yang menandakan proses pelatihan berjalan dengan stabil. Kedua kurva mulai mencapai titik konvergensi sekitar *epoch* ke-50, yang menunjukkan bahwa model telah berhasil menemukan pola data yang relevan tanpa menunjukkan gejala *overfitting*. Ketiadaan jarak signifikan antara kurva *training* dan *validation* memperkuat indikasi bahwa model memiliki kemampuan generalisasi yang baik terhadap data

yang belum pernah dilihat sebelumnya, serta mampu melakukan prediksi secara akurat dan efisien di luar data pelatihan.

3.4. Interpretasi dan Implikasi

Hasil penelitian yang diperoleh menunjukkan bahwa penggunaan Artificial Neural Network (ANN) terbukti menjadi pendekatan yang sangat tepat sekaligus terbukti menjadi pendekatan yang sangat tepat sekaligus efektif dalam menangani permasalahan regresi non-linier penjualan. Pendekatan ini sangat relevan dalam konteks prediksi penjualan mingguan yang membutuhkan pemodelan hubungan kompleks antar variabel. Kemampuan model ANN dalam peta pola dalam peta pola-pola hubungan yang tidak linier di antara berbagai variabel memungkinkan tercapainya hasil prediksi yang jauh lebih akurat dan andal dibandingkan dan andal dibandingkan metode konvensional. Dengan keunggulan tersebut, model ini diharapkan mampu memberikan kontribusi yang signifikan dalam mendukung proses pengambilan dalam mendukung proses pengambilan keputusan manajerial, terutama berkaitan dengan aspek pengelolaan stok barang agar tetap optimal, serta menyusun perencanaan strategi promosi yang lebih yang lebih tepat sasaran, efektif, dan efisien. Pemanfaatan ANN dalam konteks ini membuka peluang besar bagi perusahaan untuk meningkatkan kinerja operasional dan daya saing melalui prediksi yang berbasis data yang kuat.

4. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk merancang dan membangun model prediksi penjualan mingguan pada toko-toko Walmart dengan menggunakan algoritma Artificial Neural Network (ANN). Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa model ANN memiliki kemampuan yang efektif dalam merepresentasikan hubungan non-linier antar fitur, terutama dalam memetakan pengaruh dari variabel-variabel ekonomi dan operasional seperti suhu, harga bahan bakar, indeks harga konsumen (CPI), tingkat pengangguran, serta hari libur terhadap total penjualan mingguan. Konfigurasi terbaik untuk model ini ditemukan pada penggunaan 128 *neuron* pada lapisan tersembunyi dengan nilai *learning rate* sebesar 0.01, yang menghasilkan *Mean Squared Error* (MSE) sebesar 232.283.171.177,98. Evaluasi lebih lanjut terhadap grafik *loss function* menunjukkan bahwa proses pelatihan berjalan secara stabil, dengan tren penurunan kesalahan yang konsisten baik pada data pelatihan maupun validasi, tanpa adanya indikasi terjadinya *overfitting*.

Walaupun hasil penelitian menunjukkan performa model yang cukup memuaskan, terdapat beberapa keterbatasan penting yang perlu diperhatikan. Model ANN yang digunakan bersifat black-box, sehingga tidak mampu memberikan interpretasi secara langsung mengenai kontribusi masing-masing fitur terhadap hasil prediksi yang dihasilkan. Selain itu, model ini belum menerapkan teknik validasi silang untuk meningkatkan keandalan hasil, serta belum mempertimbangkan dinamika musiman dan tren waktu yang kemungkinan besar memiliki pengaruh signifikan terhadap fluktuasi penjualan mingguan.

Sebagai arah pengembangan lebih lanjut, disarankan agar penelitian mendatang mengeksplorasi arsitektur model yang lebih cocok untuk data deret waktu, seperti *Recurrent Neural Network* (RNN) atau *Temporal Convolutional Network* (TCN). Selain itu, integrasi metode interpretabilitas model, seperti SHAP atau LIME, juga perlu dipertimbangkan untuk memberikan pemahaman yang lebih transparan dan mendalam terhadap keputusan prediktif yang dihasilkan. Dengan demikian, hasil prediksi yang diperoleh tidak hanya akurat secara kuantitatif, tetapi juga dapat menjadi landasan yang lebih transparan dan meyakinkan dalam mendukung pengambilan keputusan strategis di sektor ritel.

5. UCAPAN TERIMAKASIH

Penulis menyampaikan terima kasih kepada dosen pembimbing atas segala bimbingan, arahan, serta dukungan yang telah diberikan selama proses penelitian berlangsung. Ucapan terima kasih turut disampaikan kepada rekan-rekan satu tim atas kerja sama dan kontribusi yang sangat berarti, khususnya dalam pengolahan data dan penyusunan laporan ini. Selain itu, apresiasi juga diberikan kepada pihak penyedia dataset yang telah mendukung kelancaran proses penelitian ini.

REFERENSI

- [1] N. A. W. a. N. A. R. Jusoh, ""Sales forecasting in retail: A systematic literature review," *Journal of Business Research*, vol. 110, p. 420–437, 2020.
- [2] Md. S. Haque. "Retail Demand Forecasting Using Neural Networks and Macroeconomic Variables," *Journal of Mathematics and Statistic Studies*, vol. 4, no. 3, pages 1-6. DOI: <https://doi.org/10.32996/jmss.2023.4.3.1>.
- [3] B. E. P. a. M. Y. H. G. P. Zhang, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, Vol. 14, no. 1, no. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii>, pp. 35-62, 2001.
- [4] P. R. K. a. V. Ravi, " "Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review,"," *European Journal of Operational Research*, Vols. vol. 180, no. 1, no. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0377221706008769> , pp. 1-28, 2007.
- [5] A. Borovykh, " "Conditional time series forecasting with convolutional neural networks,"," *arXiv preprint*, vol. arXiv:1703.04691, 2017.
- [6] F. Chollet, "Deep Learning with Python," CT: Manning Publications, 2018.