

Meningkatkan Prediksi Penjualan Retail Xyz Dengan Teknik Optimasi Random Search Pada Model Xgboost

Hendra Wijaya ¹, Dandy Pramana Hostiadi ², Evi Triandini ³

Program Studi Magister Sistem Informasi
Institut Teknologi dan Bisnis STIKOM Bali
Denpasar, Indonesia

e-mail: 222012007@stikom-bali.ac.id ¹, dandy@stikom-bali.ac.id ², evi@stikom-bali.ac.id ³

Abstrak

Dalam dunia ritel konvensional, prediksi penjualan sering kali bergantung pada naluri dan intuisi, namun dengan meningkatnya kompleksitas data, pendekatan ini menjadi kurang relevan. Karena itu, *machine learning* menjadi kunci dalam mengatasi tantangan yang dihadapi oleh peritel saat ini. Penelitian ini fokus pada meningkatkan akurasi prediksi penjualan perusahaan ritel XYZ dengan mengembangkan dan mengoptimalkan model XGBoost menggunakan teknik Random Search. Awalnya, model XGBoost memiliki koefisien determinasi (R^2) sebesar 96,50%, menunjukkan kemampuannya dalam menjelaskan variasi dalam data penjualan. Namun, setelah dilakukan optimasi menggunakan teknik Random Search, terjadi peningkatan signifikan pada R^2 menjadi 97,21%, mencerminkan peningkatan akurasi model. Selisih antara keduanya adalah sebesar 0,71%, yang menunjukkan peningkatan yang substansial setelah proses optimasi. Dengan prediksi yang lebih tepat, perusahaan dapat mengoptimalkan manajemen persediaan, meningkatkan keakuratan perencanaan produksi, dan mengalokasikan sumber daya secara optimal. Temuan penelitian menegaskan efektivitas teknik optimasi ini dalam menghasilkan prediksi penjualan yang lebih akurat untuk ritel XYZ.

Kata kunci: Xgboost, retail, Random Search, grid search, bayesian optimization.

1. Pendahuluan

Di masa lampau, manusia bergantung pada naluri dan intuisi untuk mengelola bisnis mereka. Dengan perkembangan teknologi dewasa ini, kecerdasan buatan kini memainkan peran vital dalam berbagai sektor, termasuk di industri ritel. Penjualan dalam perusahaan ritel menjadi krusial, bahkan menentukan kesuksesan atau kegagalan perusahaan tersebut [1]. Namun, kebanyakan peritel, termasuk perusahaan retail XYZ, sering menghadapi kesulitan dalam membuat prediksi penjualan yang tepat. Beruntungnya, kemunculan *machine learning* telah membuka pintu bagi solusi baru dalam hal ini. [2].

Machine learning bertujuan untuk menciptakan model yang universal berdasarkan data yang spesifik. Dengan pengembangan algoritma baru dan ketersediaan data yang melimpah, diharapkan *machine learning* dapat meningkatkan prediksi penjualan secara lebih efektif dibandingkan dengan metode tradisional, serta mengatasi korelasi yang kompleks dalam industri ritel. [3].

XGBoost, sebuah model *machine learning* yang handal, telah digunakan secara luas untuk keperluan prediksi dalam berbagai sektor, termasuk namun tidak terbatas pada properti, geografi, dan bidang medis. [4][5][6][7][8]. Contoh-contoh sektor tersebut menunjukkan bahwa model XGBoost memiliki fleksibilitas untuk menyesuaikan diri dengan beragam situasi dan kondisi, termasuk di sektor ritel. [9]. Sektor ritel ditandai dengan tingginya jumlah pekerjaan manusia dan margin keuntungan yang kecil, sesuai untuk menerapkan teknologi kecerdasan buatan dan *machine learning* [10]. Pelaku industri ini semakin memahami pentingnya pemanfaatan model *machine learning* dalam meramalkan penjualan produk mereka [11].

Penelitian yang dipimpin oleh Li dan rekannya [12], Mencoba meramalkan penjualan di masa depan, peneliti menggunakan model XGBoost dan menganalisis data penjualan dari tahun 2013 hingga 2017 di beberapa toko di Favorita, Ekuador. Hasilnya menunjukkan prediksi yang sangat akurat.

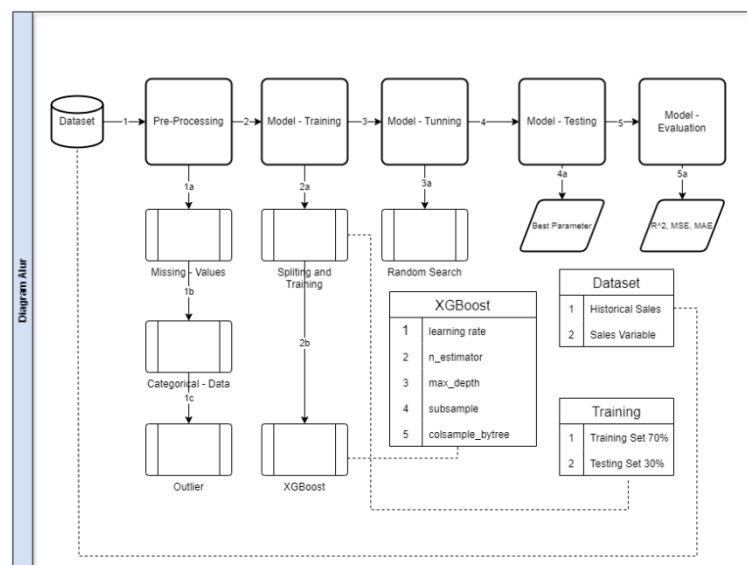
Selanjutnya Pavlyshenko dkk [13] Menerapkan algoritma XGBoost dalam pengembangan model mereka untuk memprediksi seri waktu penjualan. Hasil penelitian menunjukkan bahwa penggunaan model tersebut menghasilkan hasil yang lebih baik daripada menggunakan metode time series, terutama saat pola penjualan tidak mengikuti tren historis yang jelas. Akande dan timnya melakukan penelitian yang serupa. [14]. Mereka memanfaatkan XGBoost dalam meramalkan penjualan dengan menggunakan data dari 45 toko

retail. Hasil analisis menunjukkan bahwa *XGBoost* efektif dalam meramalkan penjualan dan dapat memberikan kontribusi dalam pengambilan keputusan harga oleh manajer penjualan.

Model *XGBoost* telah terbukti mampu melakukan prediksi data regresi dan klasifikasi dengan baik dan akurat. Namun, masih ada potensi untuk meningkatkan akurasi model tersebut dengan menggunakan berbagai teknik *tuning* parameter. Model ini memiliki fleksibilitas yang luas dan dapat menyesuaikan diri dengan beragam kondisi data, seperti yang terlihat dalam dataset retail. Oleh karena itu, penyetelan parameter menjadi kunci untuk meningkatkan kinerjanya. Penelitian ini bertujuan untuk mengeksplorasi dan mengoptimalkan model *XGBoost* dengan menggunakan teknik *Random Search*. Tujuannya adalah untuk meningkatkan akurasi prediksi, yang dapat membawa dampak positif pada efisiensi rantai pasokan, manajemen stok, dan perencanaan strategi penjualan di industri ritel, terutama bagi perusahaan Retail XYZ. Diharapkan bahwa optimasi model *XGBoost* akan membantu perusahaan menghadapi ketidakpastian pasar dengan lebih efektif.

2. Metode Penelitian

Berikut ini adalah diagram alir yang kami usulkan dalam penelitian ini :



Gambar 1. Diagram Alir

Tahap pertama dalam penelitian ini adalah melakukan *preprocessing* (1) pada *dataset*. Hal ini meliputi penanganan data yang hilang (1a), pengonversian data kategorikal (1b), serta eliminasi *outlier* (1c). Setelahnya, model utama, yaitu *XGBoost*, akan dilatih (2) dengan menggunakan skema validasi 30-70 (2a). Proses selanjutnya terdiri dari *tuning* model (3) menggunakan metode *tuning*, yaitu *Random Search* (3a). Setelah *tuning*, model akan diuji / testing (4, 4a) dan hasilnya akan dievaluasi (5) dengan metrik R^2 , MAE (*Mean Absolute Error*), dan MSE (*Mean Square Error*) (5a) untuk mengukur performa dan akurasi model prediksi penjualan. Tujuan evaluasi ini adalah untuk memverifikasi bahwa model *XGBoost* telah dioptimalkan secara efisien dan mampu memberikan prediksi penjualan yang akurat. Dengan menggunakan ketiga metode tersebut, penelitian ini tidak hanya membandingkan hasil dari ketiga metode *tuning*, tetapi juga memastikan bahwa parameter terbaik dipilih untuk meningkatkan akurasi prediksi penjualan dalam konteks industri *ritel*, terutama pada perusahaan *retail XYZ*.

Dalam pembahasan deskripsi data terkait *dataset* [17] penjualan ini merupakan *public dataset* dari situs *kaggle*, akan dijelaskan karakteristik dasar dari data penjualan yang terkumpul.

Tabel 1. Deskripsi Dataset

No	Features		Data Type
	Feature Name	Description	
1	Store	Merepresentasikan nomor atau identifikasi unik dari toko retail	Integer
2	Date	Merupakan tanggal penjualan yang tercatat dalam format tahun-bulan-tanggal (YYYY-MM-DD)	String
	Type	Merupakan Type Store dengan nilai (A, B, C)	String
	Size	Merupakan Size dari store	Integer
3	Weekly_Sales	Merupakan total sales mingguan dari toko yang bersangkutan.	Float
4	IsHoliday	Merupakan variabel biner (1 atau 0) yang menandakan apakah hari penjualan jatuh pada hari libur atau bukan.	Integer
5	Temperature	Merupakan suhu rata-rata pada hari penjualan di lokasi toko retail tersebut	Float
6	Fuel_Price	Merupakan harga rata-rata bahan bakar pada hari penjualan	Float
7	CPI	(Consumer Price Index): Merupakan indeks harga konsumen yang mengukur tingkat inflasi atau perubahan harga rata-rata barang dan jasa yang dikonsumsi oleh konsumen	Float
8	Unemployment	Merupakan tingkat pengangguran pada hari penjualan.	Float
9	Year	Merupakan tahun penjualan.	Integer
10	Month	Merupakan bulan penjualan.	Integer
11	Week	Merupakan minggu (ke) penjualan.	Integer
12	Min	Merupakan nilai minimum dari data penjualan mingguan.	Float
13	Max	Merupakan nilai maksimum dari data penjualan mingguan.	Float
14	Mean	Merupakan nilai rata-rata dari data penjualan mingguan.	Float
15	STD	Merupakan nilai standar deviasi dari data penjualan mingguan	Float
16	Total_MarkDown	Merupakan nilai diskon total dari data penjualan mingguan.	Float

Fitur-fitur dalam dataset tersebut terdiri dari 16 jenis yang mencakup beragam data. Enam fitur menggunakan tipe data integer, yang mencakup informasi penting seperti informasi toko, ukuran, apakah hari tersebut merupakan hari libur atau tidak, serta informasi tentang tahun, bulan, dan minggu. Sedangkan, dua fitur menggunakan tipe data string untuk merepresentasikan informasi tanggal dan tipe. Sementara itu, delapan fitur lainnya menggunakan tipe data *float* dan termasuk parameter penting dalam analisis penjualan, seperti penjualan mingguan, suhu, harga bahan bakar, indeks harga konsumen CPI (*Consumer Price Index*), tingkat pengangguran, serta statistik *mark down* (minimum, maksimum, rata-rata, dan standar deviasi), dan *total mark down*. Kualitas dataset ini membuatnya sangat relevan untuk prediksi penjualan, terutama dengan adanya fitur yang merepresentasikan nilai minimum, maksimum, dan rata-rata penjualan mingguan. Pentingnya fitur eksternal seperti CPI (*Consumer Price Index*), suhu, dan apakah hari tersebut merupakan hari libur menjadi jelas, memungkinkan analisis terhadap faktor-faktor tersebut dalam memprediksi perilaku penjualan. Sebagai contoh, fitur apakah hari tersebut merupakan hari libur memberikan wawasan apakah penjualan memiliki kecenderungan naik atau turun selama periode libur, memberikan pemahaman yang lebih mendalam tentang dinamika penjualan dalam konteks waktu.

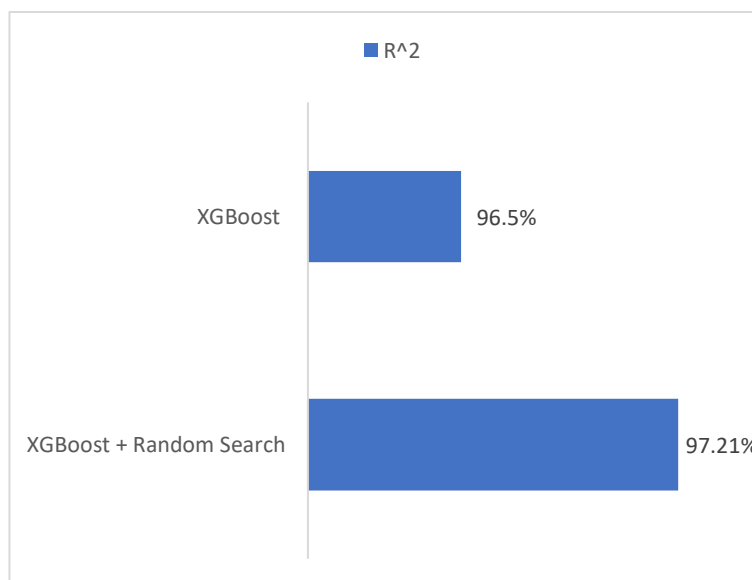
3. Hasil dan Pembahasan

Tabel 2. Hasil Pengujian Awal

Model	R^2	MSE	MAE
XGBoost	96.50	3786648.40	1153.51
XGBoost + <i>Random Search</i>	97.21	3019287.87	1025.57

Dalam konteks penelitian mengenai kinerja model prediksi penjualan, hasil penelitian menunjukkan bahwa Model *XGBoost* yang telah melalui proses *tunning* / optimasi menggunakan teknik *Random Search* mengalami peningkatan yang cukup signifikan dalam semua metrik evaluasi dibandingkan dengan versi awalnya. Terutama, terjadi peningkatan yang mencolok dalam nilai R^2 , dari 96,50% menjadi 97,21%, hal ini menggambarkan kemampuan model yang telah dioptimalkan untuk memahami variasi data penjualan dengan lebih baik.

Selain itu, penurunan yang signifikan dalam nilai MSE (*Mean Square Error*) dan MAE (*Mean Absolute Error*) menandakan bahwa model yang telah dioptimalkan mampu menghasilkan prediksi yang lebih akurat, dengan tingkat kesalahan yang lebih rendah dibandingkan dengan model dasar. Hasil ini menggarisbawahi kesuksesan proses optimasi menggunakan teknik *Random Search* dalam meningkatkan kemampuan prediktif model *XGBoost*, terutama dalam konteks memprediksi penjualan untuk perusahaan ritel XYZ.

Gambar 2. Komparasi Grafik R^2

Dari grafik R^2 di atas, terlihat adanya kemajuan yang signifikan dalam evaluasi kinerja model prediksi penjualan melalui serangkaian langkah optimasi. Awalnya, model *XGBoost* memiliki koefisien determinasi (R^2) sebesar 96,50%, mencerminkan kemampuannya untuk menjelaskan variasi dalam data penjualan. Setelah melakukan optimasi menggunakan teknik *Random Search*, terjadi peningkatan R^2 menjadi 97,21%, yang menggambarkan peningkatan akurasi model. Peningkatan ini mengkonfirmasi bahwa langkah optimasi, memberikan kontribusi positif yang signifikan terhadap akurasi dan keandalan model dalam melakukan prediksi penjualan.

4. Kesimpulan

Dalam konteks prediksi penjualan, model *XGBoost* yang telah dioptimasi menggunakan teknik *Random Search* menunjukkan performa yang lebih baik dibandingkan dengan metode dasar. Secara garis besar, temuan ini menegaskan pentingnya pengembangan lebih lanjut dalam berbagai aspek penelitian, termasuk pemilihan model alternatif, pengembangan fitur, seleksi variabel, penggunaan teknik *outlier*, dan variasi metode optimasi. Namun, hasil penelitian ini juga memiliki keterbatasan, terutama dalam hal penggunaan model *machine learning* yang eksklusif, teknik optimasi yang hanya pada *Random Search* dan dataset yang terbatas pada industri ritel. Oleh karena itu, penelitian selanjutnya harus melakukan eksplorasi lebih lanjut untuk meningkatkan validitas dan generalisasi hasil penelitian.

Daftar Pustaka

- [1] X. Dairu and Z. Shilong, "Machine Learning Model for Sales Forecasting by Using XGBoost," *2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021*, pp. 480–483, Jan. 2021, doi: 10.1109/ICCECE51280.2021.9342304.
- [2] A. Schmidt, M. W. U. Kabir, and M. T. Hoque, "Machine Learning Based Restaurant Sales Forecasting," *Mach Learn Knowl Extr*, vol. 4, no. 1, 2022, doi: 10.3390/make4010006.
- [3] A. Mitra, A. Jain, A. Kishore, and P. Kumar, "A Comparative Study of Demand Forecasting Models for a Multi-Channel Retail Company: A Novel Hybrid Machine Learning Approach," *Operations Research Forum*, vol. 3, no. 4, Dec. 2022, doi: 10.1007/S43069-022-00166-4.
- [4] K. Xu, "Predicting housing prices and analyzing real estate markets in the Chicago suburbs using machine learning," *Journal of Student Research*, vol. 11, no. 3, p. undefined-undefined, Aug. 2022, doi: 10.47611/JSRHS.V11I3.3459.
- [5] C. Çilgin and H. Gökçen, "Machine learning methods for prediction real estate sales prices in Turkey," *Revista de la Construcción*, vol. 22, no. 1, pp. 163–177, 2023, doi: 10.7764/RDLC.22.1.163.
- [6] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," *Geoscience Frontiers*, vol. 12, no. 1, pp. 469–477, Jan. 2021, doi: 10.1016/J.GSF.2020.03.007/PREDICTION_OF_UNDRAINED_SHEAR_STRENGTH_USING_EXTREME_GRADIENT_BOOSTING_AND_RANDOM_FOREST_BASED_ON_BAYESIAN_OPTIMIZATION.PDF.
- [7] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIsci Remote Sens*, vol. 57, no. 1, pp. 1–20, Jan. 2020, doi: 10.1080/15481603.2019.1650447.
- [8] N. Hou *et al.*, "Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost," *J Transl Med*, vol. 18, no. 1, Dec. 2020, doi: 10.1186/S12967-020-02620-5.
- [9] K. Matuszelański and K. Kopczewska, "Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 1, 2022, doi: 10.3390/jtaer17010009.
- [10] F. Weber and R. Schütte, "A domain-oriented analysis of the impact of machine learning—the case of retailing," *Big Data and Cognitive Computing*, vol. 3, no. 1, 2019, doi: 10.3390/bdcc3010011.
- [11] E. Martins and N. V. Galeale, "RETAIL SALES FORECASTING INFORMATION SYSTEMS: COMPARISON BETWEEN TRADITIONAL METHODS AND MACHINE LEARNING ALGORITHMS," in *Proceedings of the 15th IADIS International Conference Information Systems 2022, IS 2022*, 2022. doi: 10.33965/is2022_2022011004.
- [12] K. Li, "A Sales Prediction Method Based on XGBoost Algorithm Model," *BCP Business & Management*, vol. 36, pp. 367–371, Jan. 2023, doi: 10.54691/BCPBM.V36I.3487.
- [13] B. M. Pavlyshenko, "Machine-learning models for sales time series forecasting," *Data (Basel)*, vol. 4, no. 1, 2019, doi: 10.3390/data4010015.
- [14] Y. F. Akande, J. Idowu, A. Misra, S. Misra, O. N. Akande, and R. Ahuja, "Application of XGBoost Algorithm for Sales Forecasting Using Walmart Dataset," *Lecture Notes in Electrical Engineering*, vol. 881, pp. 147–159, 2022, doi: 10.1007/978-981-19-1111-8_13.
- [15] "Optimasi hyperparameter XGBoost-studi kasus prediksi klaim asuransi = Hyperparameter optimization in XGBoost-case study of insurance claim prediction." Accessed: Jan. 09, 2024. [Online]. Available: <https://lib.ui.ac.id/detail?id=20509606&lokasi=lokal>
- [16] M. Ryan Afrizal, R. Adi Nugroho, D. Kartini, R. Herteno, J. Ahmad Yani Km, and K. Selatan, "XGBOOST DENGAN RANDOM SEARCH HYPER-PARAMETER TUNING UNTUK KLASIFIKASI SITUS PHISING".
- [17] "Amazon uk SalesForecasting 2019-2021 | Kaggle." Accessed: Jun. 10, 2023. [Online]. Available: <https://www.kaggle.com/datasets/revanthkrishnakomali/amazon-uk-salesforecasting-20192021>