E-ISSN: 3025-0994
Halaman 1605-1619
Volume 4 Nomor 4 Tahun 2026

Prediksi Penjualan XGBoost untuk Rekomendasi Diskon Produk Kuliner Berbasis Web

R. Q. Putra[1], L. O. M. G. Jaya[2], A. H. Wibowo[3]

[1,2,3] Jurusan Informatika, Fakultas Teknik, Universitas Halu Oleo, Kendari, Indonesia
e-mail korespondensi: qubrarama@gmail.com, laodemgj@uho.ac.id, asa.hari@uho.ac.id

Abstrak

Penentuan diskon produk kuliner pada PT Bonton ATB Sejahtera masih mengandalkan pengamatan persediaan dan pengalaman pengelola sehingga informasi historis penjualan belum dimanfaatkan secara optimal. Penelitian ini bertujuan mengembangkan aplikasi pendukung keputusan berbasis web yang mengintegrasikan prediksi jumlah penjualan dengan rekomendasi diskon berdasarkan kondisi stok serta masa kedaluwarsa produk. Penelitian menggunakan 936 catatan penjualan selama 2023-2025 yang mencakup 13 produk dalam empat kategori. Sebanyak 624 catatan periode 2023-2024 digunakan untuk pelatihan, sedangkan 312 catatan tahun 2025 digunakan untuk pengujian temporal. Setelah prapemrosesan diperoleh 22 fitur untuk model Extreme Gradient Boosting. Hasil prediksi diproses dengan aturan bisnis berprioritas sehingga menghasilkan rekomendasi diskon sebesar 0%, 10%, 20%, atau 30%. Aplikasi dikembangkan menggunakan Django dan diuji melalui evaluasi galat prediksi serta pengujian jalur program. Nilai kesalahan absolut rata-rata mencapai 1,894050 produk dan akar kesalahan kuadrat rata-rata sebesar 3,179004 produk pada seluruh data uji. Kesalahan persentase absolut rata-rata sebesar 9,887142% dihitung pada 211 catatan dengan penjualan aktual positif. Enam skenario pengujian fungsi rekomendasi menunjukkan keluaran sesuai aturan implementasi. Penelitian menghasilkan prototipe fungsional untuk mendukung keputusan diskon, tetapi belum membuktikan peningkatan laba atau pengurangan limbah. Aturan untuk produk yang telah mencapai masa kedaluwarsa masih memerlukan perbaikan sebelum penerapan operasional. Hasil ini belum menjadi dasar klaim efektivitas kebijakan harga dalam kondisi pasar yang sesungguhnya.

Kata Kunci: Django, Prediksi Penjualan, Produk Kuliner, Rekomendasi Diskon, XGBoost

Abstract

Discount decisions for culinary products at PT Bonton ATB Sejahtera have relied largely on inventory observations and managerial experience, limiting the practical use of historical sales records. This study aimed to develop a web-based decision-support application that combined sales-quantity forecasting with discount recommendations informed by stock conditions and remaining product shelf life. The research used 936 sales records collected between 2023 and

2025, covering 13 products in four categories. A total of 624 observations from 2023-2024 were assigned to model training, while 312 observations from 2025 were reserved for temporal testing. Data preprocessing yielded 22 predictors for an Extreme Gradient Boosting regression model. Forecasts were subsequently evaluated through prioritized business rules to generate proposed discounts of 0%, 10%, 20%, or 30%. The Django application was examined using forecasting-error measures and program-path testing. The model achieved a mean absolute error of 1.894050 produks and a root mean squared error of 3.179004 produks on all test observations. Its mean absolute percentage error was 9.887142%, calculated exclusively for 211 observations with positive actual sales. Six recommendation-function test scenarios produced outputs consistent with the implemented rules. The findings support the feasibility of an integrated forecasting and recommendation prototype, but they do not establish improvements in profit or food-waste reduction. Furthermore, the rule concerning products that have reached their expiration date requires correction before operational deployment.

Keywords: Culinary Products, Django, Discount Recommendation, Sales Forecasting, XGBoost

LATAR BELAKANG

Digitalisasi usaha kuliner tidak cukup berhenti pada pencatatan transaksi. Data penjualan perlu diolah menjadi informasi yang membantu penentuan persediaan dan promosi. Studi terkait digitalisasi usaha mikro, kecil, dan menengah mengidentifikasi hambatan dalam penggunaan teknologi untuk pengambilan keputusan (Dafitri & Daryani, 2025); sistem informasi penjualan dapat memperbaiki pengelolaan data usaha kuliner (Nurjanah et al., 2026). Perubahan proses bisnis berbasis data juga dibahas pada penelitian digitalisasi UMKM Indonesia (Meilariza et al., 2024), sedangkan orientasi pasar dan inovasi produk berhubungan dengan kinerja pemasaran sektor kuliner (Hardilawati et al., 2023).

PT Bonton ATB Sejahtera merupakan perusahaan yang memproduksi serta memasarkan produk kuliner di Kendari. Perusahaan sudah mencatat transaksi secara digital, tetapi keputusan pemberian diskon masih banyak dipengaruhi kondisi stok dan pertimbangan pengelola. Produk dengan umur simpan terbatas menimbulkan risiko apabila permintaan lebih rendah dibandingkan jumlah stok. Diskon dapat digunakan sebagai alternatif penanganan produk yang masih layak jual, tetapi pemberiannya harus mempertimbangkan kemungkinan penjualan, persediaan, dan waktu sebelum kedaluwarsa.

Data penelitian terdiri atas 936 catatan transaksi tahun 2023-2025, yang mencakup 13 produk dalam kategori kacang, kerupuk, kue, dan sambal. Informasi penjualan tersebut memuat tanggal, harga, stok, jumlah terjual, kategori, serta masa kedaluwarsa. Penelitian ini mengembangkan aplikasi berbasis Django yang memprediksi jumlah penjualan menggunakan XGBoost. Hasil prediksi kemudian diproses bersama indikator stok dan sisa masa simpan melalui aturan bisnis untuk menyusun rekomendasi diskon.

Kesenjangan yang ingin dijawab bukan sekadar akurasi model prediksi, melainkan bagaimana keluarannya dapat ditampilkan sebagai bantuan keputusan diskon dalam alur kerja aplikasi. Tujuan penelitian meliputi pembangunan sistem prediksi penjualan, perumusan

rekomendasi diskon berdasarkan kondisi produk, serta pengujian performa model dan logika program. Persentase diskon tidak dipelajari langsung oleh XGBoost, melainkan ditentukan oleh aturan yang dirancang peneliti. Karena belum ada uji lapangan setelah penerapan diskon, peningkatan keuntungan maupun penurunan limbah tidak dijadikan klaim hasil penelitian.

Pemanfaatan data historis sebagai dasar kebijakan promosi memerlukan pemisahan dua persoalan: memperkirakan berapa banyak produk berpotensi terjual dan memilih tindakan harga berdasarkan hasil perkiraan tersebut. Kinerja prediksi saja tidak cukup untuk menyimpulkan bahwa kebijakan diskon meningkatkan pendapatan perusahaan. Pada barang dengan umur simpan terbatas, pengelola juga perlu mengetahui apakah produk aman dijual, berapa stok yang masih tersedia, dan kapan masa simpan berakhir. Penelitian ini menempatkan prediksi sebagai masukan bagi keputusan, bukan sebagai sistem optimasi harga otomatis. Perbedaan fungsi ini penting supaya pembacaan hasil eksperimen tetap sesuai dengan jenis bukti yang dikumpulkan.

Prediksi penjualan memperkirakan jumlah barang yang mungkin terjual dengan memanfaatkan pola historis. Model statistik maupun pembelajaran mesin telah digunakan untuk tujuan tersebut. Falatouri dkk. membandingkan pendekatan SARIMA dan LSTM untuk prediksi permintaan ritel (Falatouri et al., 2022), sedangkan Shahar dkk. menerapkan metode pembelajaran mesin pada toko daring (Shahar et al., 2024). Jin membandingkan SVM, Random Forest, dan XGBoost (Jin, 2024); Dankorpho menggunakan XGBoost untuk peramalan penjualan bisnis ritel (Dankorpho, 2024). Hasil penelitian terdahulu menunjukkan relevansi pendekatan tersebut, tetapi tidak membuktikan bahwa XGBoost pasti unggul pada seluruh himpunan data.

Pengembangan model gabungan RF-XGBoost untuk penjualan produk pertanian berumur simpan pendek telah diteliti oleh J. Li dkk. (J. Li et al., 2024). Di bidang perdagangan digital, Duan menggunakan pembelajaran mesin untuk memprediksi pembelian ulang (Duan, 2024), sedangkan Ruscikasani dkk. mengembangkan prediksi pembelian berdasarkan perilaku sesi dengan XGBoost (Ruscikasani et al., 2025). Tang juga menerapkan model XGBoost pada prediksi kinerja penjualan tiket film (Tang, 2024). Studi-studi itu memperlihatkan fleksibilitas model prediktif, tetapi pemetaan prediksi jumlah penjualan menjadi rekomendasi diskon tetap memerlukan mekanisme keputusan tersendiri.

Penetapan diskon untuk produk mudah rusak berkaitan dengan persediaan, sisa masa simpan, harga, dan potensi pemborosan. Sanders membahas pengaruh kebijakan harga dinamis serta insentif pengurangan limbah pangan (Sanders, 2024). Lin dan Januardi menganalisis penetapan harga dua tahap dengan upaya menjaga kualitas dan mengurangi limbah (Lin & Januardi, 2023), sedangkan Haijema dkk. mengkaji diskon dinamis berdasarkan tanggal kedaluwarsa (Haijema et al., 2026). Ketiga penelitian tersebut memberi landasan konseptual penggunaan umur simpan, tetapi tidak memvalidasi ambang diskon 10%, 20%, atau 30% untuk perusahaan yang menjadi objek penelitian ini.

Pengujian white-box berbasis jalur digunakan untuk menelusuri cabang keputusan pada fungsi program melalui kompleksitas siklomatis. Pratiwi dan Widianti memberikan contoh penerapan teknik basis path pada pengujian fungsi sistem informasi (Pratiwi & Widianti, 2025). Pengujian semacam ini digunakan untuk memeriksa kesesuaian keluaran program dengan aturan yang dibuat, bukan untuk membuktikan efektivitas bisnis aturan tersebut.

Penggunaan umur simpan sebagai indikator sederhana memiliki keterbatasan karena tanggal kedaluwarsa tidak selalu menggambarkan seluruh dimensi kualitas produk. Hassini dkk. membahas keputusan markdown yang dipicu pengukuran kualitas pada produk pangan (Hassini et al., 2026). Oleh sebab itu, rekomendasi diskon pada penelitian ini harus dipandang sebagai bantuan keputusan awal dan tetap memerlukan pemeriksaan kelayakan produk oleh pengelola.

Kebutuhan pembandingan dengan pendekatan yang lebih sederhana tampak dalam penelitian Wang dkk., yang mengembangkan gradient boosting dengan informasi spasial dan temporal untuk prediksi permintaan ritel (Wang et al., 2024). Long dkk. juga mengevaluasi penggunaan gradient boosted trees dalam sistem peramalan berskala besar, termasuk tantangan deret penjualan yang tidak selalu memiliki permintaan positif (Long et al., 2025). Kedua penelitian tersebut tidak menjadi bukti keunggulan model pada data PT Bonton ATB Sejahtera, tetapi memperjelas pentingnya menilai kemampuan generalisasi, kestabilan antarkategori, dan ketepatan pemilihan pembanding ketika model diterapkan pada perusahaan tertentu.

Dalam penentuan harga barang mudah rusak, H. Li dkk. menyajikan pendekatan peramalan serta keputusan persediaan dan harga sayuran berdasarkan kesegaran dan profitabilitas (H. Li et al., 2024). Chehrazi dkk. menunjukkan bahwa ketidaklengkapan informasi stok dan umur simpan dapat membatasi kemampuan pengecer menentukan harga secara dinamis (Chehrazi et al., 2025). Temuan tersebut relevan sebagai landasan kehati-hatian ketika aplikasi hanya menyediakan jumlah stok dan tanggal kedaluwarsa tanpa pemantauan kualitas aktual. Akibatnya, rekomendasi yang dihasilkan penelitian ini harus dinilai sebagai bantuan keputusan awal, sementara validitas pengukuran stok dan keselamatan pangan memerlukan kontrol tambahan.

METODE

Penelitian dilakukan pada PT Bonton ATB Sejahtera menggunakan data penjualan historis selama 2023-2025. Produk pengamatan berupa catatan penjualan produk, bukan transaksi pelanggan individual. Dari 936 pengamatan, sebanyak 624 catatan tahun 2023-2024 dipakai untuk data latih dan 312 catatan tahun 2025 untuk data uji. Pembagian berdasarkan urutan waktu digunakan untuk mengurangi risiko tercampurnya data periode lebih baru ke tahap pelatihan.

Data dikumpulkan dari pencatatan penjualan internal perusahaan dan digunakan pada tingkat catatan penjualan produk. Identitas pelanggan individual tidak menjadi produk analisis. Pemisahan pelatihan dan pengujian mengikuti kalender tahun agar seluruh catatan pengujian berasal dari periode yang lebih baru. Strategi ini dimaksudkan untuk mendekati penggunaan model pada waktu mendatang dan menghindari pengacakan yang dapat mencampurkan pola dari masa depan dengan data latih. Namun, penetapan satu periode uji tidak identik dengan validasi berulang terhadap perubahan pola musiman dan kondisi usaha.

Table 1. Karakteristik data dan pembagian pengujian

Komponen	Keterangan
Periode data	2023-2025
Jumlah catatan	936
Produk/kategori	13 / 4
Data latih	624 (2023-2024)
Data uji	312 (2025)
Jumlah fitur	22
Target	Jumlah terjual

Seperti pada Tabel 1, data uji sepenuhnya berasal dari periode setelah data latih. Variabel yang tersedia mencakup tanggal transaksi, nama produk, kategori, harga, stok, jumlah terjual, dan tanggal kedaluwarsa. Atribut sumber penjualan terdapat dalam dataset, tetapi tidak digunakan sebagai fitur model.

Pembagian temporal dipilih karena pemakaian model dalam aplikasi bertujuan memperkirakan kondisi setelah periode data historis. Apabila data tahun 2025 dibagikan secara acak ke data latih, informasi dari periode uji dapat ikut dipelajari dan menghasilkan evaluasi yang terlalu optimistis. Meskipun begitu, penggunaan tahun 2025 sebagai satu-satunya periode uji tidak cukup untuk memastikan kestabilan akurasi pada berbagai kondisi musim maupun perubahan permintaan di masa depan. Pengujian ulang secara bergulir, ketika jumlah data baru bertambah, layak menjadi tahap lanjutan. Penelitian ini tidak melaporkan validasi silang berulang atau pemilihan parameter melalui pencarian otomatis; karena itu, parameter yang digunakan harus dipandang sebagai konfigurasi implementasi yang diuji, bukan konfigurasi paling optimal.

Tanggal penjualan dan kedaluwarsa dikonversi ke format tanggal untuk membentuk fitur tahun, bulan, dan sisa hari kedaluwarsa. Harga serta stok dipertahankan sebagai fitur numerik; nama produk dan kategori diubah menjadi variabel dummy melalui one-hot encoding. Seluruh proses membentuk 22 fitur. Pada proses prediksi, susunan kolom masukan diselaraskan dengan fitur model yang dilatih.

Regresi XGBoost diimplementasikan menggunakan XGBRegressor dengan n_estimators = 100, learning_rate = 0,1, max_depth = 5, dan random_state = 42. Evaluasi dilakukan menggunakan satu pengujian holdout temporal pada 2025, bukan rerata validasi silang. Akurasi diukur melalui MAE dan RMSE pada seluruh 312 catatan. MAPE hanya dihitung pada data dengan nilai penjualan aktual lebih besar daripada nol untuk menghindari pembagian dengan nol.

Untuk membaca hasil evaluasi secara tepat, MAE dan RMSE diperlakukan sebagai ukuran galat dalam satuan jumlah terjual, sedangkan MAPE menyatakan galat relatif pada pengamatan yang memenuhi syarat nilai aktual positif. Karena distribusi penjualan memuat sejumlah nilai nol, ketiga ukuran tersebut tidak memakai dasar pengamatan yang identik. Oleh sebab itu, MAPE tidak boleh dilaporkan tanpa menyebutkan jumlah data positif yang terlibat. Penilaian komparatif terhadap metode lain juga memerlukan pembagian data dan definisi fitur yang sama agar hasilnya dapat dibandingkan secara adil.

Pengkodean produk dan kategori menghasilkan fitur biner yang menunjukkan identitas pengamatan. Proses ini memungkinkan model menggunakan perbedaan pola antarproduk tanpa mengasumsikan adanya urutan numerik pada nama produk. Fitur tahun dan bulan mewakili aspek waktu, sedangkan harga, stok, dan hari menuju kedaluwarsa mewakili kondisi yang terkait dengan penjualan. Namun, seluruh fitur masukan harus tersedia pada saat prediksi benar-benar diminta. Khusus variabel stok, diperlukan pemeriksaan apakah nilai dalam data historis merupakan stok sebelum penjualan atau setelah transaksi. Apabila yang dipakai adalah stok setelah penjualan, model berpotensi mempelajari informasi yang seharusnya belum tersedia pada waktu prediksi. Audit pencatatan waktu dan definisi kolom menjadi lebih penting daripada sekadar menaikkan angka akurasi.

Fitur numerik disusun dari harga, stok, tahun, bulan, serta selisih hari antara tanggal kedaluwarsa dan tanggal transaksi. Fitur kategorikal dibentuk melalui pengkodean identitas produk dan kategori. Hasil pengkodean dipertahankan konsisten pada masukan prediksi melalui penyelarasan kolom terhadap fitur pelatihan. Sebanyak 22 fitur yang dilaporkan merupakan total setelah pembentukan seluruh variabel. Penelitian ini tidak melaporkan pencarian parameter sistematis, sehingga pengaturan jumlah pohon, laju pembelajaran, dan kedalaman pohon diperlakukan sebagai konfigurasi yang dievaluasi, bukan hasil optimasi global.

Model menghasilkan jumlah penjualan prediksi yang kemudian dibulatkan menjadi P. Jika S adalah stok saat prediksi, estimasi stok tersisa dihitung dengan $R = \max(0, S - P)$. Untuk $P > 0$, rasio stok tersisa didefinisikan $r = R/P$. Nilai r serta jumlah hari menuju kedaluwarsa (E) diperiksa melalui aturan berprioritas sebagaimana Tabel 2. Aturan ini menghasilkan rekomendasi, bukan perubahan harga otomatis.

Model regresi tidak menghitung secara langsung persentase diskon. Setelah perkiraan jumlah produk terjual diperoleh, nilai tersebut dibulatkan sebelum masuk ke mekanisme rekomendasi. Pada tahap ini, estimasi stok tersisa dibandingkan terhadap penjualan prediksi, kemudian dikombinasikan dengan waktu menjelang kedaluwarsa. Urutan pemeriksaan disusun agar sebuah observasi yang memenuhi beberapa kondisi sekaligus hanya memperoleh satu keputusan. Logika ini transparan dan dapat ditelusuri melalui penjelasan dalam aplikasi, meskipun tidak mengoptimalkan margin, elastisitas permintaan, atau besaran kerugian akibat produk tidak terjual.

Table 2. Aturan rekomendasi diskon dalam implementasi

Prioritas	Kondisi	Diskon
1	$P \leq 0$	0%
2	$E \leq 0$	30%
3	$E \leq 30$ dan $r \geq 1{,}0$	30%
4	$E \leq 60$ dan $r \geq 0{,}5$	20%
5	$r \geq 0{,}5$	10%
6	Kondisi lainnya	0%

Tabel 2 merepresentasikan implementasi yang diuji, bukan kebijakan harga baru yang telah dioptimasi. Harga setelah rekomendasi diperoleh dari harga awal dikalikan (1 − persentase diskon/100). Keputusan pemberian diskon tetap berada pada administrator. Perlu dicatat bahwa cabang E ≤ 0 saat ini masih menghasilkan 30%, sehingga belum aman diterapkan pada barang yang sudah tidak layak dijual.

Urutan prioritas aturan diperlukan karena sebuah produk dapat memenuhi lebih dari satu kondisi secara bersamaan. Misalnya, produk dengan rasio stok tinggi dan sisa masa simpan singkat dapat memenuhi aturan diskon 10% sekaligus 30%; sistem menggunakan kondisi yang diperiksa lebih dahulu. Dengan demikian, besaran rekomendasi merupakan konsekuensi langsung dari urutan pemeriksaan serta ambang r dan E. Pemilihan ambang masih dilakukan secara heuristik. Belum terdapat data eksperimen tentang kenaikan permintaan akibat perubahan harga yang dapat digunakan untuk menghitung elastisitas atau keuntungan maksimum. Evaluasi pada penelitian ini karena itu dibatasi pada konsistensi hasil perhitungan model dan aturan program.

Pengembangan aplikasi mengikuti tahapan Rational Unified Process (RUP), yang meliputi identifikasi kebutuhan, perancangan, implementasi, dan pengujian. Aplikasi dibuat dengan Python serta kerangka kerja Django dan menyediakan pengelolaan data produk, unggah berkas CSV historis, pelatihan model, prediksi, rekomendasi diskon, serta riwayat hasil. Pada fungsi train_model() dan rekomendasi_diskon() dilakukan pengujian white-box basis path dengan skenario yang memeriksa jalur keputusan program.

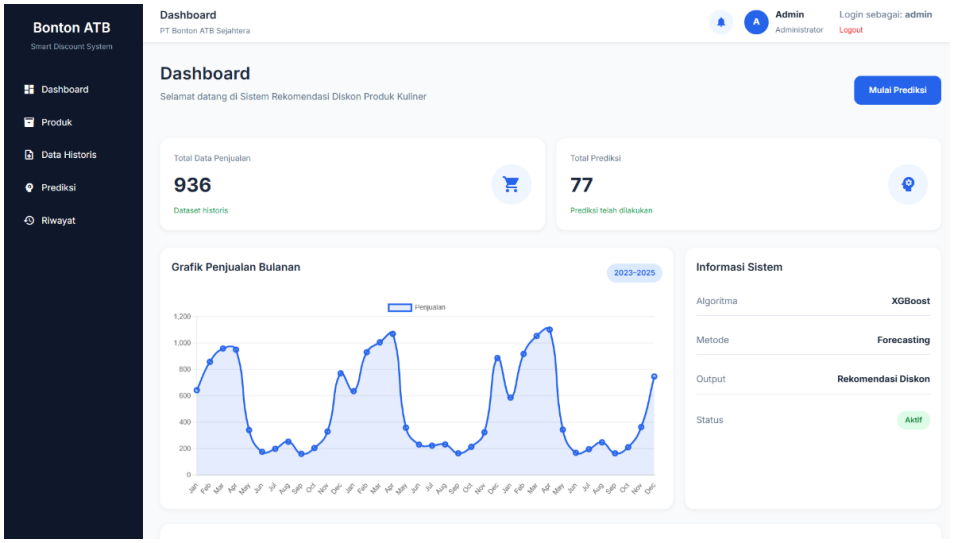
Pengujian perangkat lunak menilai logika fungsi dan validasi jalur, sementara evaluasi prediksi menilai selisih antara target penjualan aktual dengan keluaran regresi. Kedua pengujian tersebut mempunyai indikator keberhasilan yang berbeda. Performa prediksi dicatat dalam metrik galat, sedangkan hasil pengujian jalur dicatat berdasarkan kesesuaian keluaran fungsi terhadap ketentuan percabangan. Penggunaan keduanya memberikan informasi tentang fungsi sistem yang bekerja, tetapi tidak menunjukkan adanya peningkatan permintaan setelah diskon diberikan.

HASIL DAN PEMBAHASAN

Aplikasi yang dihasilkan mencakup halaman dasbor, pengelolaan produk, unggah dataset dan pelatihan, formulir prediksi, serta riwayat rekomendasi. Administrator memasukkan data produk, harga, stok, dan tanggal yang berkaitan dengan umur simpan. Setelah model memperkirakan jumlah terjual, sistem menampilkan stok tersisa, tingkat diskon yang disarankan, perkiraan harga setelah diskon, serta penjelasan aturan yang terpicu. Fitur tersebut menunjukkan integrasi antara pemodelan prediktif dan dukungan keputusan.

Halaman dasbor digunakan untuk mengawasi ringkasan data yang tersedia dan menyediakan akses ke fungsi prediksi. Tampilan ini memperlihatkan jumlah catatan historis, informasi prediksi yang tersimpan, grafik ringkasan, serta navigasi ke pengelolaan produk. Kemampuan ini memudahkan administrator berpindah dari peninjauan data ke permintaan prediksi, meskipun tidak menunjukkan pengukuran dampak ekonomi dari rekomendasi diskon.

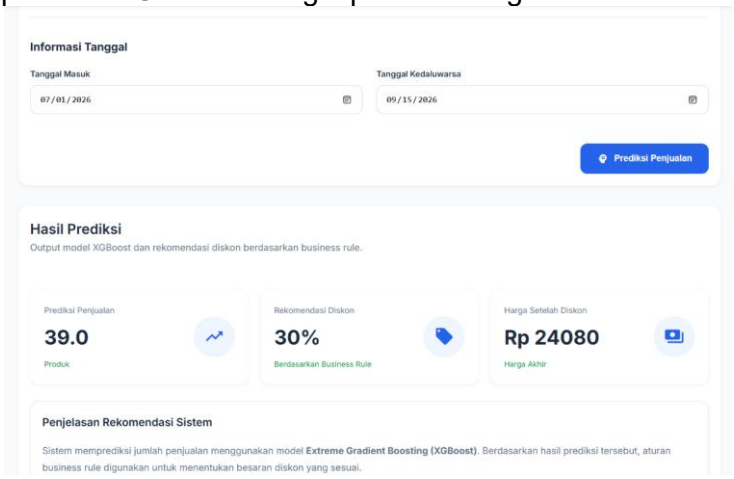
E-ISSN: 3025-0994
Halaman 1605-1619
Volume 4 Nomor 4 Tahun 2026



Gambar 1. Tampilan dasbor aplikasi

Gambar 1 menggambarkan alur pemantauan pada dasbor sistem. Jumlah data yang tersaji merupakan informasi operasional dan tidak boleh diartikan sebagai jumlah pengamatan baru di luar dataset yang digunakan dalam evaluasi model.

Pemisahan aktivitas pada antarmuka memungkinkan data produk dikelola sebelum pengguna meminta prediksi. Ketika dataset historis diperbarui, proses pelatihan ulang dapat dijalankan dari menu yang tersedia dan hasilnya disimpan untuk penggunaan berikutnya. Pada tahap prediksi, sistem membaca parameter produk, membentuk masukan yang sesuai struktur fitur model, kemudian menampilkan nilai penjualan dan rekomendasi harga sebagai dua keluaran yang berbeda. Pemisahan ini memberi ruang bagi perubahan aturan bisnis tanpa harus menganggap model XGBoost sebagai penentu harga.



Gambar 2. Tampilan hasil prediksi dan rekomendasi diskon

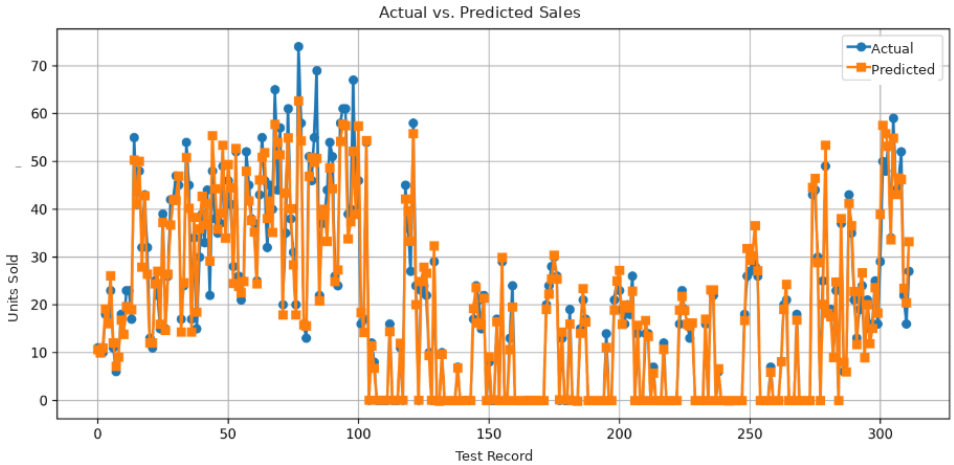
Gambar 2 menampilkan contoh keluaran saat administrator meminta prediksi: jumlah penjualan yang diperkirakan, tingkat diskon, harga setelah diskon, dan keterangan sistem. Contoh antarmuka ini berasal dari dokumentasi implementasi skripsi dan bukan bukti bahwa pemberian diskon meningkatkan pendapatan. Penilaian kualitas hasil tetap harus dilakukan melalui data uji serta evaluasi aturan keputusan.

Pengujian pada 312 catatan tahun 2025 menghasilkan MAE sebesar 1,894050 produk dan RMSE 3,179004 produk. Artinya, selisih mutlak rata-rata mencapai sekitar 1,89 produk penjualan per catatan. MAPE sebesar 9,887142% dihitung hanya pada 211 catatan yang nilai aktualnya positif; sebanyak 101 catatan bernilai aktual nol tidak dimasukkan dalam perhitungan MAPE. Seluruh catatan tetap digunakan untuk MAE dan RMSE.

Table 3. Hasil evaluasi model pada data uji tahun 2025

Metrik	Nilai	Jumlah data
MAE	1,894050	312
RMSE	3,179004	312
MAPE	9,887142%	211

Tabel 3 menunjukkan model telah dievaluasi secara kuantitatif, tetapi belum dibandingkan dengan metode dasar seperti rata-rata historis atau prediksi musiman. Karena itu, penelitian ini tidak menyimpulkan bahwa XGBoost paling akurat dibandingkan seluruh algoritma alternatif.



Gambar 3. Grafik perbandingan penjualan aktual dan prediksi

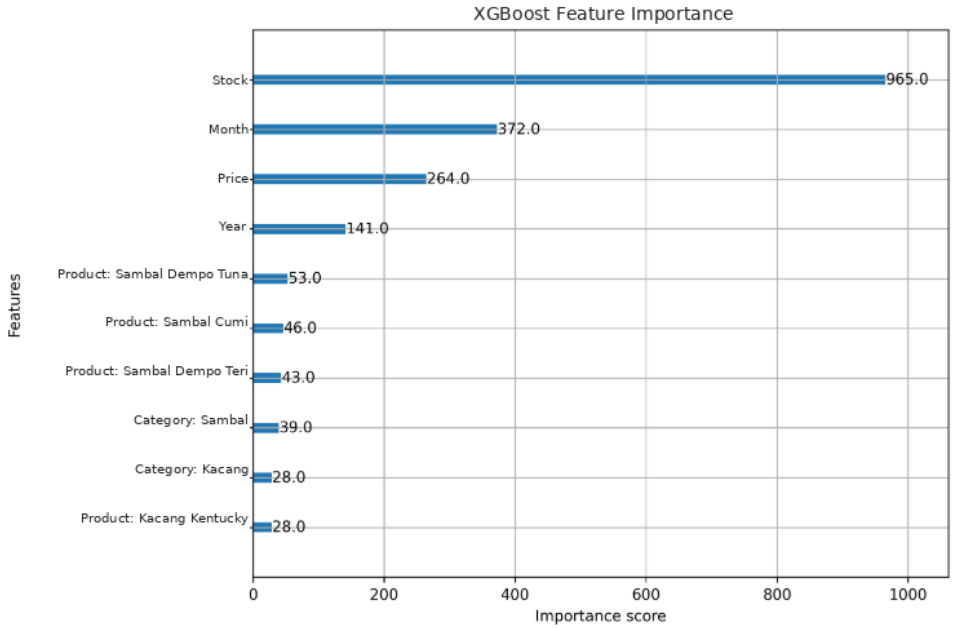
Gambar 3 menunjukkan bahwa sebagian perubahan penjualan dapat diikuti model, meskipun terdapat perbedaan pada sejumlah titik, terutama ketika jumlah terjual berubah tajam. Kurva ini merupakan pelengkap interpretasi visual, sedangkan penilaian kuantitatif

mengacu pada Tabel 3. Keberhasilan pada periode uji 2025 belum membuktikan ketahanan model pada periode yang lebih panjang.

Penggunaan MAPE pada catatan dengan penjualan positif perlu diberi perhatian khusus ketika data berisi banyak angka nol. Dalam himpunan uji, 101 dari 312 pengamatan memiliki penjualan aktual nol, sehingga sekitar sepertiga catatan tidak masuk denominator perhitungan MAPE. Artinya, persentase 9,887142% tidak boleh ditafsirkan sebagai rata-rata kesalahan relatif seluruh data uji tanpa pengecualian. Pada situasi stok produk yang tidak terjual, justru kemampuan memprediksi nilai nol atau nilai rendah penting untuk keputusan pemberian diskon. Oleh sebab itu, pengujian berikutnya sebaiknya juga melaporkan galat per produk dan frekuensi kesalahan pada kelompok penjualan nol agar pengelola mengetahui kapan rekomendasi memerlukan kehati-hatian.

Perbedaan cakupan pengamatan antarmetrik merupakan persoalan pelaporan yang penting. MAE dan RMSE dihitung pada semua 312 baris sehingga mencerminkan kesalahan absolut dan penalti atas deviasi besar dalam seluruh data uji. Sebaliknya, MAPE tidak dapat didefinisikan secara biasa pada target aktual nol dan dalam penelitian ini diterapkan hanya pada 211 baris positif. Pembaca tidak boleh menggunakan nilai MAPE untuk menyimpulkan akurasi pada 101 baris bernilai nol. Kelompok nol seharusnya dianalisis terpisah karena produk yang tidak laku dapat menjadi masukan penting bagi keputusan diskon dan persediaan.

Pengukuran tingkat kepentingan fitur menggunakan importance_type = 'weight', yaitu frekuensi suatu fitur dipakai untuk memecah simpul pohon. Stok menjadi fitur dengan frekuensi tertinggi (965), diikuti bulan (372), harga (264), dan tahun (141). Angka tersebut bukan pengaruh kausal dan tidak menunjukkan seberapa besar perubahan penjualan akibat peningkatan stok.



Gambar 4 Frekuensi penggunaan fitur pada model XGBoost

Berdasarkan Gambar 4, stok merupakan indikator yang dominan. Namun, bila angka stok pada catatan historis baru diketahui setelah transaksi berjalan, penggunaannya berpotensi menyebabkan kebocoran informasi pada evaluasi prediksi. Selain itu, fitur waktu pada data latih dibuat dari tanggal transaksi, sedangkan antarmuka prediksi menggunakan tanggal kedatangan produk; perbedaan makna tanggal ini memerlukan penyelarasan dan pengujian lanjutan.

Nilai pentingnya fitur berdasarkan frekuensi pemisahan pohon juga memiliki keterbatasan interpretasi. Fitur dengan banyak titik pemisahan bisa tampak lebih dominan dibanding fitur lain, walaupun hal tersebut tidak serta-merta menunjukkan kontribusi terhadap penurunan galat terbesar. Untuk menjelaskan prediksi per produk, penelitian lanjutan dapat melengkapi analisis dengan teknik interpretasi yang mempertimbangkan kontribusi tiap fitur pada keluaran model. Pada tahap sekarang, Gambar 4 hanya membuktikan fitur mana yang paling sering digunakan dalam struktur pohon yang telah dilatih. Kesimpulan tentang hubungan sebab-akibat antara perubahan harga, jumlah stok, dan penjualan tidak dapat diambil dari grafik tersebut.

Dominasi variabel stok juga perlu dibandingkan dengan cara pencatatan operasional. Chehrazi dkk. membahas friksi informasi inventori dalam penetapan harga barang mudah rusak; informasi yang tidak tepat waktu dapat membatasi keputusan yang diambil. Dalam penelitian ini, saat persisnya stok dicatat relatif terhadap penjualan belum diaudit secara menyeluruh. Dengan demikian, belum dapat dipastikan apakah seluruh informasi pada data historis tersedia sebelum kejadian yang hendak diprediksi. Pemeriksaan garis waktu setiap fitur seharusnya menjadi syarat validasi berikutnya sebelum hasil evaluasi dipakai untuk keputusan rutin.

Fungsi rekomendasi_diskon() memiliki kompleksitas siklomatis enam dan diuji menggunakan enam jalur independen. Semua skenario menghasilkan nilai sesuai aturan yang ditulis pengembang, sebagaimana dirangkum pada Tabel 4. Fungsi train_model() memiliki kompleksitas siklomatis tujuh. Status lolos uji menunjukkan konsistensi perilaku perangkat lunak terhadap spesifikasi, bukan validitas ekonomis atau keamanan kebijakan diskonnya.

Table 4. Ringkasan pengujian fungsi rekomendasi diskon

Kasus	P; R; E	Diskon	Hasil
P1	0; 10; 20	0%	Lulus
P2	7; 3; 0	30%	Lulus
P3	77; 123; 20	30%	Lulus
P4	77; 43; 45	20%	Lulus
P5	77; 43; 90	10%	Lulus
P6	7; 3; 4	0%	Lulus

Skenario P2 pada Tabel 4 menunjukkan kelemahan operasional paling penting: ketika E = 0, sistem tetap merekomendasikan diskon 30%. Hasil pengujian tersebut benar terhadap aturan awal, tetapi barang yang sudah melewati batas kelayakan jual semestinya diblokir,

bukan dipromosikan. Oleh sebab itu, validasi kelayakan produk harus ditempatkan sebelum fungsi pemberian diskon dijalankan.

Pengaruh stok dan masa simpan terlihat pada kasus P3 hingga P5. Pada P = 77 dan R = 123 dengan E = 20 hari, rasio stok sekitar 1,60 memicu diskon 30%. Apabila R = 43 dan E = 45 hari, rasio sekitar 0,56 menghasilkan diskon 20%; untuk E = 90 hari, hasilnya turun menjadi 10%. Ambang ini bersifat heuristik yang ditentukan pengembang dan belum dibuktikan optimal melalui estimasi elastisitas harga maupun perhitungan margin usaha.

Secara teknis, alur aplikasi dapat dibedakan menjadi tiga tahap. Pertama, catatan historis diunggah dan diperiksa sebelum model dipelajari; kesesuaian tipe tanggal serta keberadaan kolom diperlukan agar fitur yang dipakai stabil. Kedua, proses prediksi mengubah masukan pengguna ke susunan 22 fitur yang sesuai dengan pelatihan. Ketiga, nilai prediksi dipakai oleh fungsi rekomendasi dengan mempertimbangkan rasio stok dan masa simpan. Pemisahan tersebut penting untuk pengembangan berikutnya karena perubahan aturan diskon tidak mengharuskan seluruh model XGBoost dilatih kembali.

Selain memeriksa nilai diskon yang muncul, pengujian fungsi perlu mempertimbangkan batas masukan dan keadaan yang tidak biasa. Prediksi nol atau negatif diproses oleh prioritas pertama agar tidak dilakukan pembagian terhadap nilai P yang tidak positif. Ketika produk sudah kedaluwarsa, aturan yang ada masih menghasilkan diskon 30% jika lolos dari pemeriksaan pertama. Kondisi ini menunjukkan bahwa pemeriksaan validitas pangan harus berdiri sebagai penghalang sebelum kalkulasi diskon, bukan hanya salah satu tingkat potongan harga. Nilai prediksi dan hasil pengujian algoritma tidak memberikan izin untuk menjual produk yang tidak memenuhi syarat. Administrator tetap memerlukan data status mutu produk serta kebijakan operasional yang sesuai.

Secara praktis, penanganan barang kedaluwarsa harus dipisahkan dari algoritma pemilihan persentase diskon. Bila tanggal kedaluwarsa telah tercapai, penilaian kelayakan produk tidak boleh direduksi menjadi pemberian potongan harga tambahan. Aturan yang diuji mencerminkan perilaku perangkat lunak yang ada, bukan usulan bahwa produk kedaluwarsa patut dijual. Penguatan sistem dapat dilakukan dengan validasi kelayakan sebelum fungsi rekomendasi dipanggil dan dengan audit produk terhadap standar mutu yang berlaku. Perubahan tersebut merupakan usulan pengembangan; penelitian ini tidak mengklaim bahwa mekanisme pemblokiran telah diimplementasikan.

Penelitian menunjukkan kelayakan teknis alur dua tahap: XGBoost memperkirakan penjualan, kemudian aturan bisnis menghasilkan opsi diskon berdasarkan risiko persediaan. Dalam literatur produk mudah rusak, penetapan harga mempertimbangkan umur simpan, kualitas, potensi permintaan, dan biaya (Sanders, 2024; Lin & Januardi, 2023; Haijema et al., 2026; Hassini et al., 2026). Akan tetapi, penelitian ini tidak mengukur peningkatan penjualan setelah diskon, laba bersih, atau penurunan produk yang terbuang. Oleh karena itu, tingkat diskon yang ditawarkan masih berupa usulan awal, bukan harga optimum yang tervalidasi.

Keterbatasan lainnya ialah pemakaian dataset satu perusahaan dan 13 produk, satu pembagian waktu data latih-uji, serta ketiadaan model pembanding yang dilatih pada pembagian sama. Selain itu, terdapat risiko kebocoran informasi melalui fitur stok dan perbedaan atribut tanggal antara pelatihan dan penggunaan interaktif. Evaluasi white-box tidak menggantikan uji keamanan aplikasi, uji penerimaan pengguna, atau pemeriksaan

kelayakan produk. Pengembangan berikutnya perlu mendahulukan pemblokiran barang kedaluwarsa, audit ketersediaan fitur saat prediksi, pembandingan dengan model dasar, serta uji lapangan untuk menilai hasil ekonomi.

Penggunaan aplikasi sebagai alat bantu keputusan juga memerlukan kejelasan tanggung jawab pengguna. Rekomendasi sebaiknya disertai alasan yang dapat dibaca pengelola, termasuk prediksi jumlah terjual, stok tersisa, dan aturan masa simpan yang aktif. Riwayat prediksi dapat mendukung penelusuran keputusan ketika hasil di lapangan tidak sesuai perkiraan. Akan tetapi, tersimpannya riwayat tidak menggantikan pencatatan harga aktual, jumlah transaksi setelah promosi, dan produk yang akhirnya tidak terjual. Tanpa pencatatan hasil tersebut, evaluasi apakah diskon sungguh membantu bisnis tidak dapat dilakukan. Penelitian ini baru menyediakan dasar teknis untuk pengambilan keputusan dan belum melakukan percobaan kausal terkait efektivitas kebijakan.

Perbandingan dengan literatur internasional menunjukkan dua jalur pengembangan yang belum dicapai sistem. Long dkk. menekankan keperluan peramalan retail yang terukur pada data penjualan dengan tingkat intermitensi berbeda, sedangkan Wang dkk. meneliti penggunaan pola spasial-temporal dalam pemodelan permintaan. Penelitian ini belum melakukan pembandingan semacam itu pada pembagian data yang sama. Pada sisi keputusan, H. Li dkk. menghubungkan ramalan, harga, dan pengisian persediaan dengan keseimbangan kesegaran serta profitabilitas. Aplikasi yang dibahas di sini belum memasukkan biaya per produk, margin setelah diskon, atau pengukuran perubahan permintaan akibat harga.

SIMPULAN

Penelitian menghasilkan prototipe Django yang menggabungkan prediksi penjualan XGBoost dan rekomendasi diskon berbasis aturan. Model dilatih pada 624 catatan dan diuji pada 312 catatan, menghasilkan MAE 1,894050, RMSE 3,179004, serta MAPE 9,887142% untuk 211 catatan aktual positif. Sistem menampilkan opsi diskon 0%, 10%, 20%, dan 30%; enam skenario pengujian fungsi sesuai aturan yang diimplementasikan. Namun, aturan bagi produk yang telah kedaluwarsa harus diperbaiki sebelum penerapan. Penelitian selanjutnya perlu mengaudit fitur stok, membandingkan model, dan menguji dampak diskon terhadap laba serta limbah produk.

DAFTAR PUSTAKA

Chehrazi, N., Sanders, R. E., & Stamatopoulos, I. (2025). Inventory Information Frictions Explain Price Rigidity In Perishable Groceries. *Marketing Science*, *44*(2), 411–436. Https://Doi.Org/10.1287/Mksc.2023.0473

Dafitri, R., & Daryani, D. (2025). Sistem Hambatan Berlapis Digitalisasi UMKM Kuliner: Studi Kualitatif Berbasis Kerangka TOE. *Jurnal Bisnis & Kewirausahaan*, *21*(3), 153–164. Https://Doi.Org/Https://Doi.Org/10.31940/Jbk.V21i3.153-164

Dankorpho, P. (2024). Sales Forecasting For Retail Business Using Xgboost Algorithm. *Journal Of Computer Science And Technology Studies*, *6*, 136–141. Https://Doi.Org/10.32996/Jcsts.2024.6.2.15

Duan, L. (2024). Machine Learning Modeling For Forecasting Repeat Purchases In Online Shopping. *MALCOM: Indonesian Journal Of Machine Learning And Computer Science*, *4*(3), 863–874. Https://Doi.Org/10.57152/Malcom.V4i3.1388

Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive Analytics For Demand Forecasting - A Comparison Of SARIMA And LSTM In Retail SCM. *Procedia Computer Science*, *200*(1), 993–1003. Https://Doi.Org/10.1016/J.Procs.2022.01.298

Haijema, R., Duijvestijn, L., Akkerman, R., & Cruijssen, F. (2026). Dynamic Expiration Date-Based Discounting Of Fresh Food Products. *International Journal Of Production Economics*, *292*. Https://Doi.Org/10.1016/J.Ijpe.2025.109824

Hardilawati, W. L., Farhanidhya, N., & Hinggo, H. T. (2023). The Effect Of Market Orientation, E-Commerce, And Product Innovation On Marketing Performance In Msmes Culinary Sector. *Jurnal Manajemen Teknologi*, *22*(2), 168–181. Https://Doi.Org/10.12695/Jmt.2023.22.2.4

Hassini, E., Ben-Daya, M., & Bahroun, Z. (2026). Iot-Enabled Quality-Triggered Markdown Pricing For Perishable Food: Equity And Waste Implications. *Foods*, *15*(4). Https://Doi.Org/10.3390/Foods15040742

Jin, Y. (2024). Sales Prediction Based On Machine Learning Approach: Support Vector Machine, Xgboost And Random Forest. *Highlights In Science, Engineering And Technology IFMPT*, *88*, 104–109. Https://Doi.Org/10.54097/6dxq4p06

Li, H., Liu, J., Qiu, J., Zhou, Y., Zhang, X., Wang, Y., & Guo, W. (2024). ARIMA-Driven Vegetable Pricing And Restocking Strategy For Dual Optimization Of Freshness And Profitability In Supermarket Perishables. *Sustainability*, *16*(10). Https://Doi.Org/10.3390/Su16104071

Li, J., Lin, B., Wang, P., Chen, Y., Zeng, X., Liu, X., & Chen, R. (2024). A Hierarchical RF-Xgboost Model For Short-Cycle Agricultural Product Sales Forecasting. *Foods*, *13*(18). Https://Doi.Org/10.3390/Foods13182936

Lin, S.-W., & Januardi. (2023). Two-Stage Pricing Of Perishable Food Supply Chain With Quality-Keeping And Waste Reduction Efforts. *Managerial And Decision Economics*, *44*(3), 1749–1766. Https://Doi.Org/10.1002/Mde.3780

Long, X., Bui, Q., Oktavian, G., Schmidt, D. F., Bergmeir, C., Godahewa, R., Lee, S. P., Zhao, K., & Condylis, P. (2025). Scalable Probabilistic Forecasting In Retail With Gradient Boosted Trees: A Practitioner's Approach. *International Journal Of Production Economics*, *279*. Https://Doi.Org/10.1016/J.Ijpe.2024.109449

Meilariza, R., Mirah Delima, Z., & . Z. (2024). Digitalization Of Micro, Small, And Medium Enterprises (Msmes) In Indonesia. *Kne Social Sciences*, *9*, 245–257. Https://Doi.Org/10.18502/Kss.V9i17.16334

Nurjanah, Y. S., Wibisono, T., & Sari, A. N. I. S. (2026). Sales Data Management Effectiveness Through A Web-Based Information System In Culinary Msmes. *Brilliance: Research Of Artificial Intelligence*, *6*(1), 68–77. Https://Doi.Org/10.47709/Brilliance.V6i1.7971

Pratiwi, Y., & Widianti, L. W. (2025). Implementasi Whitebox Testing Dengan Teknik Basis Path Pada Pengujian Halaman Pencarian Program Promo. *Jurnal Kecerdasan Buatan Dan Teknologi Informasi*, *4*(2), 173–180. Https://Doi.Org/10.69916/Jkbti.V4i2.280

Ruscikasani, S. N., Oktalivia, R. R. N., Putra, F. R., Wahidin, A. J., Rahmatullah, B., & Kurniawati, I. (2025). Prediksi Pembelian E-Commerce Menggunakan Xgboost Berbasis Perilaku Sesi Pengguna. *RIGGS: Journal Of Artificial Intelligence And Digital Business*, *4*(4), 5666–5672. Https://Doi.Org/10.31004/Riggs.V4i4.4287

Sanders, R. E. (2024). Dynamic Pricing And Organic Waste Bans: A Study Of Grocery Retailers' Incentives To Reduce Food Waste. *Marketing Science*, *43*(2), 289–316. Https://Doi.Org/10.1287/Mksc.2020.0214

Shahar, N. A. M., Mutalib, S., Halim, S. A., & Ramdhan, W. (2024). Sales Forecasting Using Machine Learning Methods For Online Store. *Proceedings Of The International Conference On Innovation & Entrepreneurship In Computing, Engineering & Science Education*, *117*, 41–52. Https://Doi.Org/10.2991/978-94-6463-589-8_6

Tang, S. (2024). The Box Office Prediction Model Based On The Optimized Xgboost Algorithm In The Context Of Film Marketing And Distribution. *Plos ONE*, *19*(10 October), 1–21. Https://Doi.Org/10.1371/Journal.Pone.0309227

Wang, J., Chong, W. K., Lin, J., & Hedenstierna, C. P. T. (2024). Retail Demand Forecasting Using Spatial-Temporal Gradient Boosting Methods. *Journal Of Computer Information Systems*, *64*(5), 652–664. Https://Doi.Org/10.1080/08874417.2023.2240753