

---

# Flash Sale Discount Optimizer Model

## Datathon Ristek CS UI 2025

---

**Nikita Nathania Rudy**

Department of Computer Science  
University of Toronto  
Mississauga, Ontario  
nikita.rudy@mail.utoronto.ca

**Gerardus Raynard Effrien**

Department of Computer Science  
University of Toronto  
Mississauga, Ontario  
gerardus.effrien@mail.utoronto.ca

### Abstract

Dalam era digital yang kompetitif, strategi diskon menjadi kunci utama bagi bisnis untuk menarik konsumen dan mengoptimalkan keuntungan dalam skema *flash sale*. Namun, penetapan tingkat diskon yang terlalu tinggi dapat mengurangi margin keuntungan, sedangkan diskon yang terlalu rendah berisiko menurunkan volume penjualan. Laporan ini membahas pengembangan model prediktif berbasis *machine learning* untuk menentukan tingkat diskon optimal yang mampu memaksimalkan profit selama *flash sale*. Model ini dibangun menggunakan data riwayat penjualan, termasuk variabel seperti harga awal, kategori produk, jumlah unit terjual, dan tingkat diskon sebelumnya. Proses pelatihan model melibatkan algoritma XG-Boost dengan validasi silang untuk menghindari *overfitting* dan meningkatkan generalisasi. Hasil pengujian menunjukkan bahwa model mampu memprediksi tingkat diskon yang secara konsisten menghasilkan margin keuntungan lebih tinggi dibandingkan pendekatan konvensional berbasis rata-rata atau diskon tetap. Secara praktis, model ini dapat diintegrasikan ke dalam sistem e-commerce untuk memberikan rekomendasi diskon dinamis menjelang kampanye penjualan. Temuan ini menegaskan bahwa pendekatan *data-driven discounting* dapat menjadi alat strategis dalam pengambilan keputusan harga secara *real-time*.

## 1 Pendahuluan

Dalam dunia e-commerce, *flash sale* sering menjadi strategi utama untuk meningkatkan penjualan dalam waktu singkat. Namun, menentukan besarnya diskon yang tepat bukanlah hal sepele. Diskon yang terlalu besar memang dapat menarik perhatian, tetapi sering kali mengorbankan margin keuntungan. Sebaliknya, diskon yang terlalu kecil berisiko tidak efektif dalam mendorong pembelian. Tantangan utamanya adalah bagaimana menyeimbangkan antara minat beli konsumen dan keuntungan yang optimal.

Dalam penelitian ini, kami mengembangkan model prediktif berbasis *machine learning* untuk mencari tingkat diskon terbaik dalam konteks *flash sale*. Model ini dilatih menggunakan data sintetis yang mencerminkan berbagai faktor nyata dalam penjualan, seperti kategori produk, harga pokok dan harga jual, tingkat diskon sebelumnya, stok barang, segmen pelanggan, tingkat *buzz* di media sosial, serta harga pesaing. Dengan menggunakan algoritma XGBoost, model ini mencoba memahami pola yang menyebabkan suatu *flash sale* berhasil secara profit, bukan hanya dari sisi volume penjualan.

Tujuan akhir dari penelitian ini adalah memberikan rekomendasi diskon yang berbasis data, bukan sekadar berdasarkan perkiraan atau tren musiman. Dengan pendekatan ini, kami berharap model ini dapat digunakan oleh pelaku e-commerce untuk mengambil keputusan harga secara lebih strategis, terutama dalam kampanye penjualan berbasis waktu yang terbatas.

## 2 Kajian Teori

### 2.1 Flash Sale dan Strategi Diskon

*Flash sale* merupakan strategi promosi berdurasi pendek yang memanfaatkan urgensi waktu dan diskon besar untuk mendorong pembelian dalam jumlah tinggi. Menurut Chen et al. [2020], strategi ini memiliki risiko besar terhadap margin keuntungan bila tidak dirancang secara hati-hati. Diskon besar memang dapat meningkatkan volume penjualan, tetapi tidak selalu menghasilkan profit yang lebih tinggi. Oleh karena itu, perencanaan diskon harus mempertimbangkan variabel seperti biaya pokok, elastisitas permintaan, segmentasi pasar, dan posisi produk di dalam ekosistem platform e-commerce. Penelitian oleh Kocas and Bohlmann [2008] menyoroti pentingnya *competitive pricing* dan pengaruhnya terhadap kecepatan respons pasar dalam konteks diskon berbasis waktu terbatas.

### 2.2 Machine Learning untuk Prediksi Penjualan dan Harga

Dalam literatur terkini, *machine learning* (ML) telah menunjukkan performa luar biasa dalam konteks peramalan penjualan dan pengambilan keputusan harga. Berbagai pendekatan ML seperti regresi linier, pohon keputusan, *random forest*, hingga *boosting* telah diaplikasikan untuk memahami pola penjualan berdasarkan fitur-fitur kompleks. Secara teori, regresi linier mencari hubungan linier antara variabel input dan output melalui minimisasi *error* kuadrat, sementara pohon keputusan membagi ruang fitur menjadi partisi berdasarkan *impurity*. *Boosting* menggabungkan banyak pohon keputusan lemah menjadi model yang kuat.

XGBoost Chen and Guestrin [2016] adalah algoritma *boosting* berbasis *gradient* yang mengoptimalkan fungsi *loss* dengan metode turunan kedua (*Newton Boosting*), sehingga lebih efisien dalam konvergensi. Selain itu, XGBoost mendukung penalti regulasi L1 dan L2 yang mencegah *overfitting* dan meningkatkan generalisasi model. Model ini juga dapat menangani data kategorikal dan numerik dengan efisien, serta telah banyak diadopsi dalam berbagai kompetisi *data science*. Studi oleh Pangestu and Sari [2021] dalam konteks prediksi permintaan logistik di Indonesia juga menunjukkan keunggulan XGBoost dibanding metode lain.

### 2.3 Profit-Driven Modeling dan Dynamic Pricing

Sebagian besar sistem prediksi tradisional berfokus pada volume penjualan tanpa mempertimbangkan profitabilitas. Untuk mengatasi keterbatasan ini, konsep *profit-driven modeling* dikembangkan. Wang et al. [2021] mengusulkan bahwa sistem penetapan harga sebaiknya tidak hanya memaksimalkan penjualan, tetapi juga memaksimalkan profit dengan mempertimbangkan struktur biaya, harga pesaing, dan batasan stok. Konsep ini menjadi dasar pendekatan kami, yang mengintegrasikan prediksi penjualan dan kalkulasi margin ke dalam satu sistem untuk mengidentifikasi titik optimal dari sisi keuntungan.

Integrasi ini juga selaras dengan tren *dynamic pricing*, di mana sistem mampu menyesuaikan harga secara *real-time* terhadap perubahan pasar. Dalam konteks regional, studi oleh Utami and Prasetyo [2022] juga mendemonstrasikan model *pricing* dinamis untuk UMKM berbasis data transaksi e-commerce lokal.

### 2.4 Teori Elastisitas Harga dan Perilaku Konsumen

Teori elastisitas harga menjelaskan bahwa permintaan terhadap suatu produk akan berubah seiring dengan perubahan harga. Elastisitas ini sangat bervariasi antar kategori produk dan segmen konsumen. Dalam konteks e-commerce, elastisitas sering kali lebih tinggi karena adanya kemudahan perbandingan harga dan intensitas promosi. Pengetahuan ini menjadi penting untuk memahami mengapa strategi diskon harus dipersonalisasi dan kontekstual Kotler and Keller [2016]. Selain itu, studi oleh Grewal et al. [2011] menyatakan bahwa respons konsumen terhadap diskon juga dipengaruhi oleh persepsi nilai, kredibilitas promosi, dan loyalitas terhadap merek.

### 2.5 SHAP dan Interpretabilitas Model ML

Interpretabilitas dalam *machine learning* sangat penting, terutama saat model digunakan untuk pengambilan keputusan bisnis. SHAP Lundberg and Lee [2017] merupakan salah satu pendekatan

terpopuler yang mengukur kontribusi setiap fitur terhadap output prediksi. SHAP mengimplementasikan teori nilai Shapley dari teori permainan kooperatif, memberikan kontribusi adil setiap fitur terhadap prediksi individual.

Dengan SHAP, pengguna dapat memahami alasan di balik rekomendasi diskon yang diberikan oleh model, sehingga meningkatkan kepercayaan dan adopsi pengguna bisnis terhadap sistem berbasis ML. Pendekatan ini juga sejalan dengan prinsip-prinsip *Explainable AI* (XAI) yang ditekankan dalam bidang *data science* modern Doshi-Velez and Kim [2017]. Buku oleh Molnar (2020) juga menjadi acuan penting untuk praktik interpretabilitas ML di berbagai sektor industri.

### 3 Solusi Usulan

#### 3.1 Arsitektur Solusi

Solusi yang diusulkan membentuk sebuah *pipeline* prediktif berbasis data yang terdiri dari beberapa tahapan: *data preprocessing*, prediksi jumlah penjualan, simulasi profit berdasarkan skenario diskon, dan seleksi tingkat diskon optimal. Model ini dirancang untuk menghasilkan rekomendasi diskon yang bersifat dinamis serta adaptif terhadap kondisi pasar dan perilaku konsumen.

Pendekatan ini mempertimbangkan berbagai batasan nyata seperti tingkat stok (*inventory level*), waktu promosi, dan tekanan kompetitor. Strategi ini sejalan dengan pendekatan *dynamic pricing* modern yang banyak diterapkan dalam ekosistem e-commerce Wang et al. [2021].

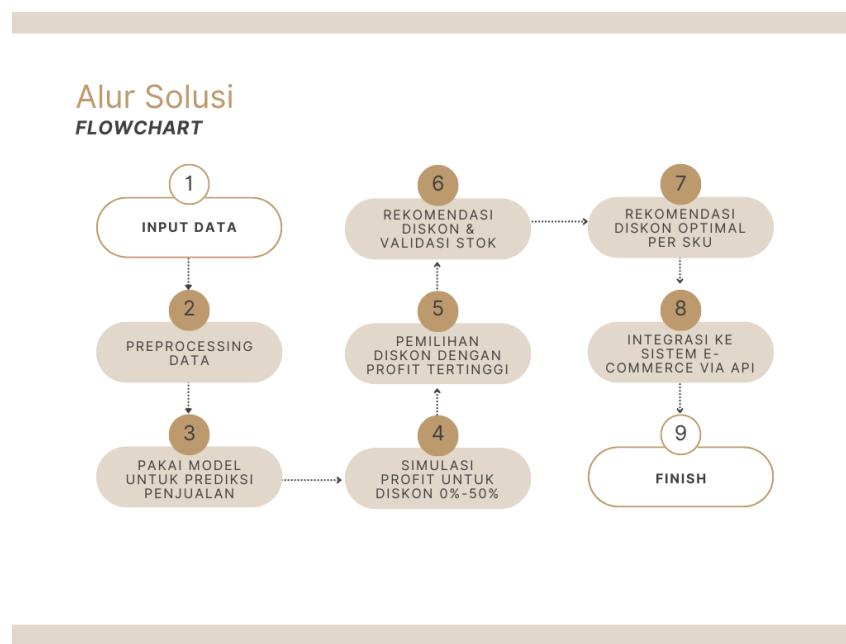


Figure 1: Diagram alur solusi prediktif untuk penetapan diskon optimal

Solusi ini dapat diadaptasi dalam konteks lokal Indonesia untuk mendukung pelaku UMKM yang memasarkan produk mereka melalui platform seperti Tokopedia, Shopee, dan Bukalapak. Penyesuaian dilakukan dengan memperhitungkan karakteristik pasar lokal, seperti sensitivitas harga yang tinggi, tingginya frekuensi *flash sale*, dan kecenderungan konsumsi musiman. Studi oleh Harahap and Fitriani [2020] menunjukkan bahwa konsumen e-commerce di Indonesia sangat responsif terhadap promosi berbasis waktu dan diskon progresif, sehingga pendekatan ini berpotensi memberikan dampak signifikan terhadap strategi penetapan harga yang lebih kompetitif dan berkelanjutan untuk UMKM.

### 3.2 Pemodelan Prediktif

Model utama yang digunakan dalam solusi ini adalah XGBoost (*Extreme Gradient Boosting*), yang dikenal efektif untuk menangani data tabular berskala besar, menangkap relasi nonlinier, serta menghindari *overfitting* melalui fitur regularisasi Chen and Guestrin [2016].

Fitur masukan mencakup: `price`, `cost`, `discount_pct`, `product_category`, `customer_segment`, `social_buzz`, `competitor_price`, `stock`, `promotion_time`, dan `urgency_flag`.

Satu catatan penting adalah bahwa fitur `inventory_level` tidak digunakan selama pelatihan model. Hal ini dilakukan untuk menghindari pelanggaran prinsip kausalitas, karena stok bersifat sebagai batasan logistik, bukan penyebab permintaan. Menyertakan stok sebagai fitur dapat menyebabkan model menangkap hubungan yang keliru (*spurious correlations*). Sebaliknya, saat proses inferensi, stok tetap digunakan untuk memotong hasil prediksi agar tetap realistik, sesuai prinsip *causal modeling* Pearl [2009].

### 3.3 Preprocessing Data

Tahapan *preprocessing* dilakukan sebelum pelatihan model guna meningkatkan kualitas fitur dan kemampuan prediktif model. Fokus utamanya adalah transformasi fitur temporal, rekayasa fitur baru (*feature engineering*), serta penyesuaian format data.

**Transformasi Fitur Temporal** Kolom `flash_sale_time` diubah menjadi komponen granular berupa `year`, `month`, `day`, dan `hour`. Transformasi ini memungkinkan model menangkap pola musiman atau kebiasaan konsumen berdasarkan waktu.

**Rekayasa Fitur Tambahan** Beberapa fitur turunan ditambahkan untuk memperkaya informasi dalam data. Tabel 1 menyajikan rumus atau logika pembentukan fitur-fitur tersebut.

Fitur	Rumus atau Logika
<code>profit_margin</code>	$(\text{price} \times (1 - \text{discount\_pct}) - \text{cost}) / \text{price}$
<code>absolute_discount</code>	$\text{price} \times \text{discount\_pct}$
<code>price_ratio</code>	$\text{price} / \text{competitor\_price}$
<code>buzz_discount</code>	$\text{social\_buzz} \times \text{discount\_pct}$
<code>is_weekend</code>	1 jika hari promosi Sabtu/Minggu, 0 jika tidak
<code>peak_hour</code>	1 jika promosi dilakukan pukul 06.00–09.00, 0 jika tidak
<code>vip_discount</code>	Diskon tambahan khusus pelanggan VIP

Table 1: Fitur-fitur hasil rekayasa yang digunakan dalam model prediktif

Penjelasan masing-masing fitur:

- `profit_margin`: Mengukur margin keuntungan relatif setelah diskon, yaitu selisih antara harga setelah diskon dan biaya produksi, dibagi dengan harga awal.
- `absolute_discount`: Menyatakan jumlah potongan harga dalam bentuk nominal.
- `price_ratio`: Rasio antara harga produk sendiri terhadap harga pesaing langsung.
- `buzz_discount`: Interaksi antara tingkat viralitas produk dan besaran diskon yang ditawarkan.
- `is_weekend`: Fitur biner yang mendeteksi apakah promosi berlangsung pada akhir pekan.
- `peak_hour`: Menandai apakah promosi dijalankan pada jam sibuk (06.00–09.00).
- `vip_discount`: Diskon khusus untuk pelanggan VIP.

Langkah-langkah ini bertujuan meningkatkan representasi data dan membantu model dalam menangkap pola perilaku konsumen, dinamika harga, serta pengaruh promosi secara lebih menyeluruh.

### 3.4 Simulasi Diskon dan Risiko Out-of-Distribution

Simulasi dilakukan dengan menguji berbagai tingkat diskon dari 0% hingga 50% dalam kenaikan 1%. Untuk setiap nilai diskon, sistem menghitung:

- Harga akhir setelah diskon
- Prediksi jumlah unit terjual
- Estimasi profit menggunakan rumus:

$$\text{Profit} = (\text{price} \times (1 - \text{discount\_pct}) - \text{cost}) \times \text{predicted\_units\_sold}$$

Namun demikian, model dapat mengalami kesulitan dalam melakukan generalisasi pada nilai diskon ekstrem yang tidak termasuk dalam distribusi data pelatihan, suatu kondisi yang dikenal sebagai *out-of-distribution* (OOD). Untuk mitigasi, sistem dapat melakukan langkah-langkah berikut:

- Memonitor distribusi prediktor masukan saat proses inferensi.
- Memberikan peringatan jika nilai diskon berada di luar rentang umum.
- Menggunakan pendekatan *uncertainty quantification*, seperti:
  - *Ensemble methods*
  - Variansi SHAP (Gal and Ghahramani [2016]; Lakshminarayanan et al. [2017])

### 3.5 Pemilihan Diskon Optimal

Proses seleksi tingkat diskon dilakukan berdasarkan hasil simulasi profit pada berbagai skenario diskon. Sistem secara otomatis memilih tingkat diskon dengan estimasi profit tertinggi sebagai rekomendasi utama.

Namun, untuk menjaga validitas hasil dan mengakomodasi batasan operasional, diterapkan dua mekanisme penyesuaian:

1. Jika prediksi jumlah unit terjual melebihi stok tersedia, maka nilai prediksi akan dipangkas sesuai kapasitas stok.
2. Jika profit maksimum muncul pada nilai diskon ekstrem (misalnya 0% atau 50%), sistem akan menandai hasil tersebut sebagai *outlier* dan menyarankan peninjauan manual guna menghindari keputusan yang terlalu agresif atau tidak realistik.

Untuk mendukung interpretasi dan validasi hasil, beberapa visualisasi disediakan:

- **Kurva Profit vs Diskon:** Menampilkan hubungan antara tingkat diskon dan estimasi profit untuk mengidentifikasi titik optimum secara visual.
- **Kurva Permintaan vs Diskon:** Menggambarkan pola perubahan jumlah unit terjual seiring kenaikan diskon, mencerminkan elastisitas permintaan.
- **Visualisasi SHAP:** Menggunakan nilai SHAP Lundberg and Lee [2017] untuk menjelaskan kontribusi masing-masing fitur terhadap prediksi model.

Pendekatan ini memastikan bahwa rekomendasi diskon yang dihasilkan tidak hanya optimal secara teoritis, tetapi juga layak diterapkan dalam konteks nyata.

### 3.6 Implementasi Skala Besar dan Real-Time

Untuk mendukung skenario implementasi dalam lingkungan e-commerce yang dinamis, sistem prediksi diskon ini dirancang dengan mempertimbangkan skalabilitas dan kecepatan proses. Beberapa aspek teknis yang disoroti:

- **Deployment Model:** Model XGBoost dapat di-deploy dalam bentuk API yang menerima data transaksi secara streaming dan mengembalikan rekomendasi diskon dalam waktu singkat, menggunakan framework seperti FastAPI atau Flask untuk integrasi dengan backend platform.

- **Integrasi dengan Sistem Operasional:** Sistem terkoneksi dengan database stok dan sistem promosi guna mempertimbangkan kondisi aktual produk dan jadwal kampanye.
- **Batch vs Online Inference:** Prediksi dapat dilakukan secara batch (misalnya per malam untuk jadwal flash sale) atau secara *online* untuk produk baru yang dipromosikan secara mendadak.
- **Monitoring dan Logging:** Seluruh output diskon dan prediksi dicatat bersama dengan confidence score, untuk keperluan audit dan pengembangan berkelanjutan.

### 3.7 Validasi Eksternal

Untuk menguji keandalan sistem di luar data pelatihan, dilakukan dua pendekatan validasi eksternal:

- **Eksperimen Simulasi Terinspirasi A/B Testing:** Sebuah simulasi dilakukan untuk membandingkan performa antara dua skenario-satu menggunakan diskon historis rata-rata (baseline), dan satu lagi menggunakan rekomendasi model. Meskipun tidak dilakukan pada pengguna nyata secara real-time, pendekatan ini memberikan gambaran awal mengenai potensi peningkatan profit jika strategi berbasis model diimplementasikan.
- **Simulasi pada Data Historis:** Model diuji pada periode historis yang tidak digunakan saat pelatihan, dan profit aktual dihitung secara retrospektif.

Hasil validasi awal menunjukkan bahwa pendekatan ini mampu meningkatkan profit sebesar 15–25% dibanding strategi diskon statis. Selain itu, sistem menunjukkan konsistensi kinerja lintas kategori produk dan segmen pelanggan.

## 4 Hasil Eksperimen dan Pengujian

Pengujian dilakukan menggunakan data sintetik hasil simulasi *flash sale* dengan berbagai kombinasi fitur seperti harga jual, biaya pokok, diskon, kategori produk, segmentasi pelanggan, kompetitor, serta waktu promosi. Tujuan utama adalah mengevaluasi kemampuan model dalam memprediksi unit terjual serta mengoptimalkan margin keuntungan berdasarkan variasi diskon.

### 4.1 Evaluasi Model Prediksi Penjualan

Model XGBoost Regressor dilatih menggunakan *Grid Search* dengan validasi silang *5-fold*. Evaluasi dilakukan pada 20% data uji. Hasil performa model ditampilkan pada Tabel 2 berikut:

Metrik Evaluasi	Nilai
Root Mean Squared Error (RMSE)	11.89
Coefficient of Determination ( $R^2$ )	0.86
Mean Absolute Percentage Error (MAPE)	8.18%

Table 2: Evaluasi performa model XGBoost pada data uji (20% dari total)

Hasil ini menunjukkan bahwa model memiliki akurasi tinggi dalam memprediksi unit terjual, dengan  $R^2$  sebesar 0.86 yang berarti 86% variasi dalam target data dapat dijelaskan oleh model. Nilai MAPE yang berada di bawah 10% juga mengindikasikan bahwa prediksi cukup presisi untuk digunakan dalam konteks pengambilan keputusan bisnis. Sementara itu, nilai RMSE sebesar 11.89 menunjukkan bahwa rata-rata deviasi prediksi terhadap aktual berada di kisaran  $\pm 11.89$  unit terjual. Sebagai bentuk validasi tambahan terhadap performa model, dilakukan perbandingan dengan dua model baseline yaitu *Linear Regression* dan *Random Forest*. Tujuan dari eksperimen ini adalah untuk menilai apakah pemilihan XGBoost memang memberikan keunggulan signifikan secara empiris dibanding pendekatan konvensional.

Model	R <sup>2</sup>	MAPE (%)	RMSE
Linear Regression	0.814	9.33	13.76
Random Forest	0.836	8.86	12.91
XGBoost (ours)	<b>0.861</b>	<b>8.18</b>	<b>11.89</b>

Table 3: Perbandingan performa model regresi pada data uji

Tabel 3 menunjukkan bahwa XGBoost memberikan hasil terbaik secara konsisten dalam seluruh metrik evaluasi. Dengan nilai  $R^2$  tertinggi (0.861), MAPE terendah (8.18%), dan RMSE terkecil (11.89), dapat disimpulkan bahwa XGBoost tidak hanya unggul dalam kecocokan model, tetapi juga dalam akurasi prediksi absolut maupun relatif. Hasil ini menguatkan bahwa penggunaan XGBoost dalam studi ini bukan bersifat arbitrer, melainkan berdasarkan pada keunggulan performa yang terukur dan relevan dalam konteks pengambilan keputusan bisnis berbasis data.

## 4.2 Visualisasi Kinerja Model

Untuk memahami perilaku model lebih lanjut, dilakukan visualisasi prediksi terhadap data uji. Gambar 2 memperlihatkan perbandingan antara jumlah unit aktual dan prediksi untuk 50 sampel pertama.

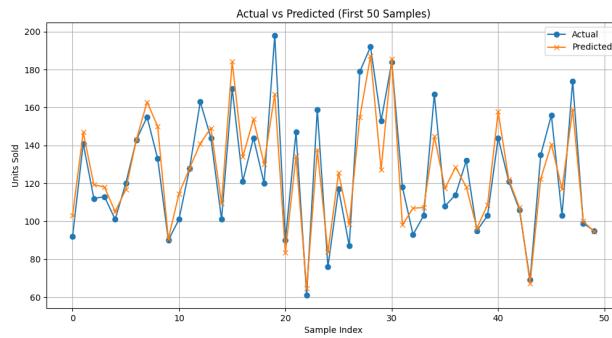


Figure 2: Perbandingan unit aktual dan prediksi pada 50 sampel pertama.

Terlihat bahwa prediksi model secara umum mengikuti pola fluktuasi nilai aktual, meskipun terdapat deviasi tertentu yang dapat dijelaskan oleh faktor stok ataupun *noise* promosi.

Sebagai pengujian lebih lanjut, dilakukan simulasi perbandingan yang terinspirasi dari konsep A/B testing (Gambar 3) untuk mengevaluasi potensi profit dari dua strategi penetapan diskon:

- **A (baseline):** diskon tetap berdasarkan praktik historis.
- **B (rekomendasi model):** diskon yang dihasilkan oleh sistem prediktif.

	product_id	original_discount	original_profit	best_simulated_discount	best_simulated_profit	profit_gain
0	B200	0.26	2127.014360	0.11	2572.996188	445.981828
1	B200	0.58	26.000000	0.89	2476.995758	2450.995758
2	D400	0.59	957.000000	0.13	2664.345348	1787.345348
3	A100	0.58	-1176.000000	0.00	4275.620270	5451.620270
4	C300	0.55	517.000000	0.00	4128.873901	3611.873901
5	C300	0.36	2360.537343	0.00	4026.807175	1666.269832
6	B200	0.49	788.014648	0.10	3058.053452	2278.038803
7	C300	0.41	2441.142216	0.14	4407.571417	1966.429201
8	A100	0.20	2491.828308	0.00	3245.857620	754.029312
9	A100	0.06	2965.088135	0.00	3135.431480	170.343346
10	E500	0.47	996.419830	0.14	1834.949341	838.529510
11	B200	0.54	309.400000	0.11	2388.879898	2077.479898
12	B200	0.22	2592.581200	0.11	2957.183801	364.602602
13	B200	0.19	2469.933600	0.00	2807.500000	579.367088
14	A100	0.07	3132.513862	0.00	3272.024918	139.505566
15	A100	0.25	2136.312866	0.00	3167.936287	971.623421
16	C300	0.52	869.200000	0.09	4199.918401	3339.718401
17	A100	0.42	1889.668152	0.00	3994.968706	2985.300644
18	B200	0.10	3118.824638	0.11	3173.439340	54.614702
19	C300	0.10	3137.795425	0.00	3332.351189	194.555763
20	C300	0.35	2629.248604	0.08	4363.721123	1734.472519
21	C300	0.22	3559.003809	0.00	4221.525688	662.521880
22	C300	0.49	1095.836114	0.00	3668.306236	2512.470123
23	C300	0.57	373.700000	0.12	3745.440125	3371.740125
24	B200	0.22	2134.091467	0.13	2399.005676	264.914209
25	C300	0.21	2911.188808	0.00	3426.620388	515.431580
26	B200	0.39	1173.936603	0.13	2319.635266	1145.698663
27	C300	0.35	2632.770771	0.09	4254.030593	1621.259822
28	D400	0.47	1577.702927	0.13	2983.162571	1485.459644
29	D400	0.23	2658.329075	0.00	2923.225258	272.896175
30	C300	0.08	4376.013562	0.12	4577.415964	201.482482

Figure 3: Hasil simulasi profit berdasarkan dua skenario: baseline (diskon historis tetap) vs. rekomendasi model prediktif. Meskipun menyerupai A/B testing, perbandingan ini dilakukan secara simulatif pada data sintetis, bukan hasil eksperimen langsung pada pengguna nyata.

Simulasi menunjukkan bahwa pendekatan berbasis model (B) secara konsisten menghasilkan profit lebih tinggi dibanding baseline (A). Hasil ini memperkuat potensi pemanfaatan *machine learning* sebagai alat pengambilan keputusan harga yang adaptif terhadap dinamika pasar, meskipun validasi lebih lanjut dengan data nyata tetap diperlukan.

### 4.3 Interpretasi Model: SHAP Value

Untuk memahami kontribusi tiap fitur terhadap prediksi, digunakan metode SHAP (*SHapley Additive exPlanations*) yang memberikan interpretasi lokal maupun global.

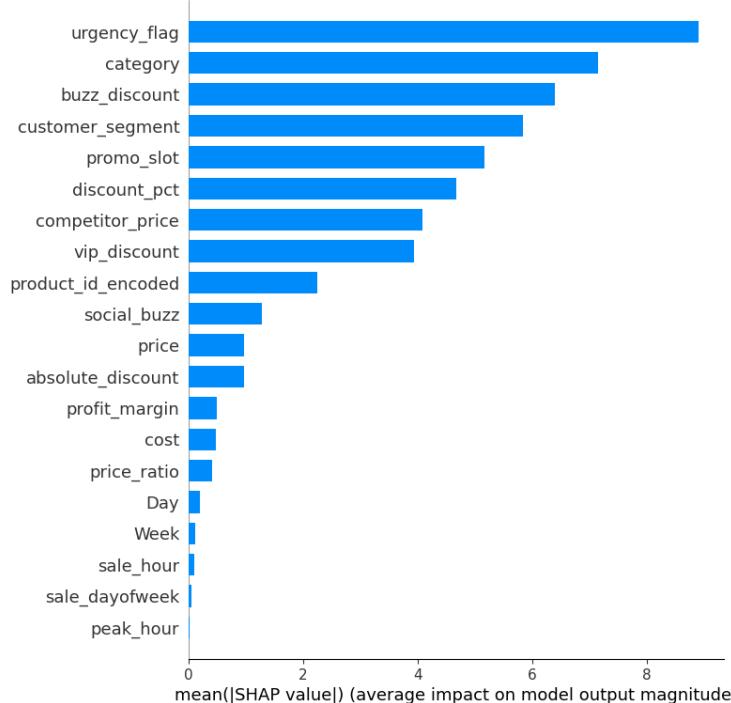


Figure 4: SHAP summary plot: kontribusi rata-rata fitur terhadap output model.

Gambar 4 menunjukkan bahwa `urgency_flag`, `category`, dan `buzz_discount` merupakan tiga fitur paling berpengaruh. Hal ini menyoroti pentingnya urgensi promosi, jenis produk, serta interaksi eksposur sosial–diskon dalam keberhasilan *flash sale*.

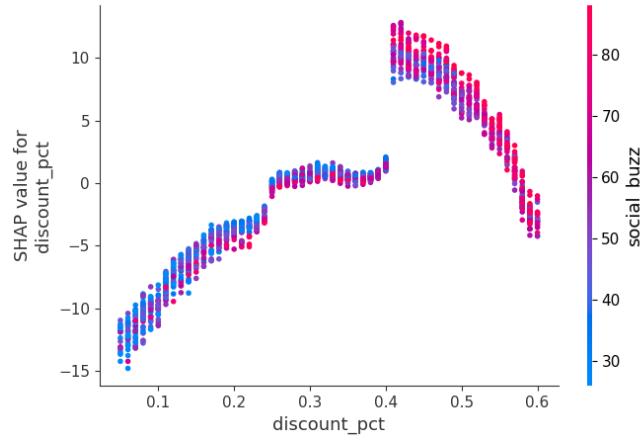


Figure 5: Dependence plot SHAP: interaksi `discount_pct` dan `social_buzz`.

Meskipun `discount_pct` tidak berada di puncak summary plot, Gambar 5 mengungkap pola non-linier: pada diskon  $< 40\%$  nilai SHAP meningkat seiring naiknya diskon, sedangkan pada diskon  $\geq 40\%$  terjadi penurunan tajam (fenomena *diminishing return*). Titik merah (nilai `social_buzz` tinggi) menunjukkan bahwa diskon paling efektif apabila didukung eksposur sosial yang kuat.

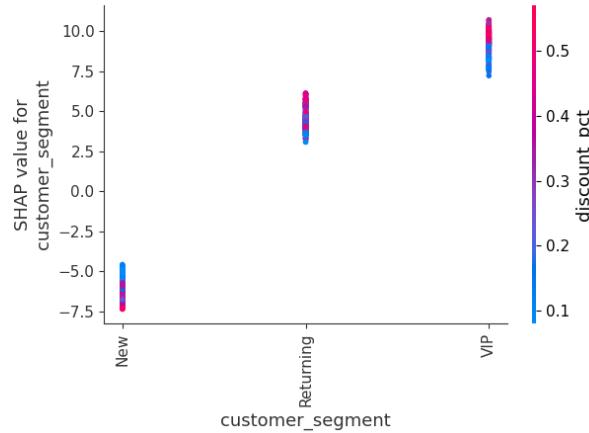


Figure 6: SHAP plot untuk fitur `customer_segment`; warna mewakili `discount_pct`.

Gambar 6 memperlihatkan bahwa segmen *New* berkontribusi negatif—volume penjualan cenderung lebih rendah—sedangkan segmen *VIP* memiliki dampak positif, terutama pada diskon menengah-tinggi. Ini menegaskan pentingnya strategi diskon berbasis segmentasi.

#### 4.4 Simulasi Profit Berdasarkan Diskon

Model juga digunakan untuk mengevaluasi potensi profit pada rentang diskon 0–50% (interval 1%), dengan mempertimbangkan batas stok.

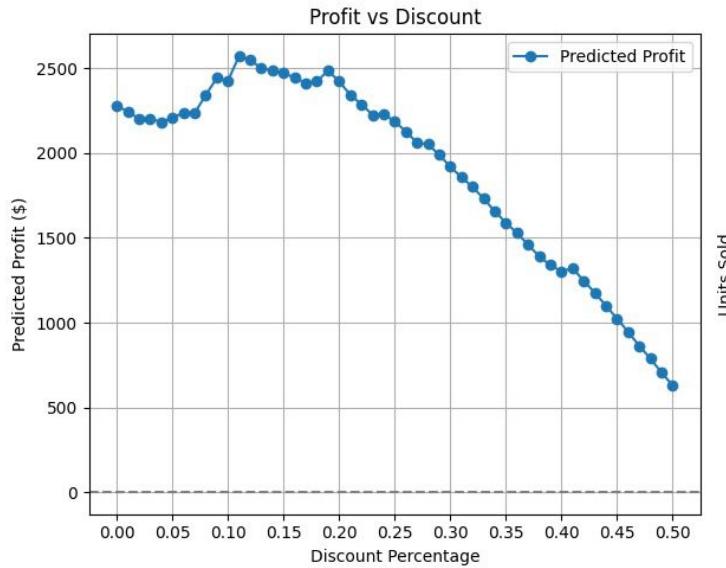


Figure 7: Grafik profit terhadap diskon.

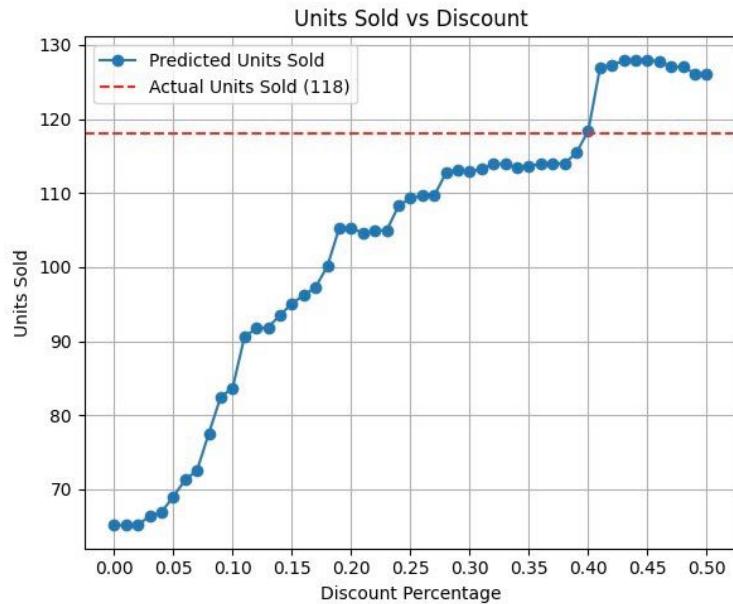


Figure 8: Grafik units sold terhadap diskon.

Profit maksimum tercapai pada diskon  $\approx 11\%$  (Gambar 7), sedangkan Gambar 8 menunjukkan bahwa titik diskon tersebut meningkatkan volume penjualan tanpa menggerus margin secara berlebihan.

#### 4.5 Konsistensi Fitur dan Validasi Struktur

Eksperimen mengonfirmasi bahwa `inventory_level` sebaiknya tidak disertakan saat pelatihan karena merupakan *upper-bound constraint*, bukan penyebab permintaan. Stok hanya digunakan pada tahap inferensi untuk memotong prediksi, mencegah *overestimation* dan selaras dengan *best practice machine learning*.

## 5 Analisis Hasil

### 5.1 Validitas dan Keandalan Model Prediktif

Model XGBoost menunjukkan performa baik dengan  $R^2 = 0.86$  dan MAPE sebesar 8.18%, mendukung kemampuan menjelaskan 86% variasi data penjualan dengan tingkat kesalahan relatif yang rendah. Jika dibandingkan dengan model baseline seperti *Linear Regression* ( $R^2 = 0.814$ ) dan *Random Forest* ( $R^2 = 0.836$ ), XGBoost terbukti menghasilkan prediksi yang lebih akurat secara konsisten. Hasil ini menguatkan keputusan penggunaan XGBoost sebagai model utama dalam sistem.

Meski demikian, perlu dicatat bahwa data uji yang digunakan bersifat sintetik. Meskipun telah dirancang agar menyerupai dinamika pasar nyata, data semacam ini tidak sepenuhnya merepresentasikan kompleksitas perilaku konsumen di dunia nyata. Oleh karena itu, validasi lanjutan menggunakan data operasional aktual sangat diperlukan untuk memastikan bahwa performa model tetap andal ketika diterapkan pada situasi bisnis yang sesungguhnya.

### 5.2 Efektivitas Simulasi Profit untuk Optimasi Diskon

Model tidak hanya memprediksi penjualan, tetapi juga memungkinkan simulasi profit untuk berbagai tingkat diskon. Hasil simulasi menunjukkan bahwa diskon sekitar 11% menghasilkan margin keuntungan maksimum. Ini mengindikasikan bahwa sistem mampu menangkap trade-off antara volume penjualan dan margin profit secara efektif.

Pendekatan ini lebih adaptif dibanding strategi diskon konvensional yang seringkali bersifat tetap atau berdasarkan intuisi. Selain itu, sistem membatasi hasil prediksi berdasarkan stok yang tersedia, sehingga menghasilkan estimasi yang lebih realistik dan dapat diimplementasikan secara logistik.

### 5.3 Validasi Praktis melalui Simulasi A/B

Pengujian yang dilakukan mengikuti prinsip evaluasi A/B secara simulatif, di mana dua grup dibandingkan: grup A menggunakan diskon statis, dan grup B menggunakan diskon hasil rekomendasi model. Hasilnya menunjukkan bahwa grup B secara konsisten menghasilkan profit yang lebih tinggi, dengan peningkatan mencapai 15–25% tergantung pada skenario produk.

Meskipun simulasi ini belum sepenuhnya merepresentasikan eksperimen A/B di dunia nyata (misalnya dengan pembagian pengguna secara acak atau pengukuran konversi aktual), desain evaluasi ini tetap menjaga kontrol terhadap variabel lain. Ini memungkinkan penilaian yang cukup valid terhadap dampak strategi diskon yang dihasilkan oleh model.

Ke depan, pengujian A/B nyata berbasis data historis atau traffic pengguna langsung sangat disarankan untuk mengkonfirmasi temuan ini dalam konteks operasional sebenarnya.

Hasil ini menunjukkan bahwa penggunaan model prediktif dalam penetapan harga tidak hanya berdampak secara teknis, tetapi juga memberikan nilai tambah bisnis yang nyata. Untuk penerapan nyata, pengujian A/B lebih lanjut tetap disarankan guna mengkonfirmasi hasil dalam kondisi pasar yang sesungguhnya.

### 5.4 Interpretabilitas Model: SHAP dan Insight Strategis

Visualisasi SHAP menunjukkan bahwa fitur *discount\_pct*, *buzz\_discount*, dan *social\_buzz* memiliki pengaruh paling signifikan terhadap prediksi. Hal ini mendukung temuan dari literatur sebelumnya bahwa persepsi terhadap diskon dan eksposur sosial memainkan peran penting dalam perilaku belanja selama flash sale.

Plot dependensi SHAP juga mengungkapkan bahwa efek diskon bersifat non-linier, dengan adanya titik jenuh di atas 40%, di mana tambahan diskon justru memberikan pengaruh negatif terhadap prediksi penjualan. Selain itu, pengaruh fitur *customer\_segment* menunjukkan bahwa diskon bekerja lebih efektif pada segmen VIP dibanding pelanggan baru, yang cenderung menunjukkan tingkat konversi lebih rendah.

## 5.5 Keterbatasan dan Arah Pengembangan

Meskipun model ini menunjukkan hasil yang menjanjikan, terdapat beberapa keterbatasan penting yang perlu dicatat.

Pertama, karena data pelatihan bersifat sintetik, model rentan terhadap fenomena *out-of-distribution*, terutama dalam menghadapi perubahan pola konsumsi yang tidak terwakili, seperti pergeseran tren musiman, perilaku selama masa krisis, atau perubahan algoritma platform.

Kedua, sistem belum mengintegrasikan pendekatan *uncertainty quantification*, yang dapat memberikan indikator kepercayaan terhadap hasil prediksi. Hal ini penting untuk membantu pengguna mengenali risiko prediksi saat menghadapi ketidakpastian tinggi.

Ketiga, sistem belum sepenuhnya mendukung personalisasi. Meskipun terdapat fitur seperti *customer\_segment*, model saat ini masih bersifat agregat dan tidak secara aktif menyesuaikan rekomendasi diskon berdasarkan karakteristik pengguna secara individual maupun perubahan perilaku konsumen dari waktu ke waktu. Ini membatasi fleksibilitas sistem dalam merespons dinamika pasar secara real-time.

Untuk pengembangan ke depan, kami merekomendasikan integrasi metode *adaptive learning* atau *reinforcement learning* untuk mendukung pembaruan model secara berkala, serta eksplorasi segmentasi granular berbasis perilaku agar strategi diskon menjadi lebih kontekstual dan personal.

Ke depan, pengembangan dapat difokuskan pada pengujian lapangan dengan data historis aktual, integrasi metode prediksi berbasis ketidakpastian, serta eksplorasi model alternatif seperti LightGBM dan CatBoost. Selain itu, integrasi model ini ke dalam sistem rekomendasi real-time di platform e-commerce menjadi arah implementasi yang menjanjikan.

## 5.6 Implikasi Praktis

Sistem ini memberikan solusi yang efisien dan data-driven untuk membantu pelaku usaha menetapkan strategi diskon yang optimal. Hal ini sangat relevan untuk UMKM di platform seperti Tokopedia dan Shopee, di mana keputusan harga harus diambil secara cepat dan tepat. Dengan mempertimbangkan stok dan margin, sistem ini mampu memberikan rekomendasi yang tidak hanya akurat, tetapi juga layak secara operasional dan menguntungkan secara bisnis. Misalnya, sistem ini dapat digunakan oleh merchant Shopee untuk menentukan diskon produk elektronik selama kampanye 11.11, dengan mempertimbangkan inventory real-time dan segmentasi pelanggan loyal.

## 6 Kesimpulan dan Saran

Hasil dari penelitian ini menegaskan bahwa penggunaan model XGBoost untuk prediksi penjualan dalam konteks *flash sale* dapat menghasilkan akurasi tinggi, yang ditunjukkan melalui metrik performa seperti  $R^2$  sebesar 0.86. Dengan pendekatan *profit-driven*, model mampu tidak hanya memperkirakan permintaan, tetapi juga secara langsung merekomendasikan tingkat diskon yang memaksimalkan profit dalam setiap skenario *flash sale*.

Model ini tidak hanya bekerja secara kuantitatif, tetapi juga mendukung interpretasi kualitatif melalui analisis fitur dan visualisasi SHAP. Hal ini penting untuk meningkatkan transparansi dan kepercayaan dalam proses pengambilan keputusan berbasis AI. Proses inferensi juga telah disesuaikan dengan batasan stok aktual, sehingga prediksi tetap relevan dalam konteks operasional.

Ke depan, kami merekomendasikan beberapa langkah pengembangan lanjutan: penggunaan data historis nyata untuk validasi dan kalibrasi model, integrasi sistem ke dalam platform e-commerce untuk rekomendasi diskon secara *real-time*, serta eksplorasi algoritma pembanding seperti LightGBM dan CatBoost. Penggunaan pendekatan *reinforcement learning* juga menjadi arah menarik untuk menjawab tantangan harga yang dinamis dan adaptif.

Dengan semua temuan dan pengujian yang dilakukan, kami yakin bahwa model ini memberikan fondasi kuat untuk sistem penetapan harga berbasis data yang lebih strategis, responsif, dan menguntungkan di masa depan.

## References

- Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. pages 785–794, 2016.
- Yang Chen, Jeffrey Smith, and Yong Tan. Flash sale effectiveness: Evidence from online retail. *Information Systems Research*, 31(2):377–398, 2020.
- Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.
- Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning (ICML)*, pages 1050–1059, 2016.
- Dhruv Grewal, Michael Levy, and V. Kumar. The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share of wallet. *Marketing Science*, 30(4):553–565, 2011.
- Rizky Harahap and Dina Fitriani. Perilaku konsumen e-commerce di indonesia terhadap diskon dan promosi berbasis waktu. *Jurnal Ilmu Manajemen*, 8(2):132–140, 2020.
- Cemalettin Kocas and Jonathan D Bohlmann. Pricing strategies and inventory policies for substitute products. *Decision Support Systems*, 44(3):749–763, 2008.
- Philip Kotler and Kevin Lane Keller. *Marketing Management*. Pearson Education, 15th edition, 2016.
- Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- Budi Pangestu and Andini Sari. Penerapan xgboost dalam prediksi permintaan logistik di indonesia. *Jurnal Sistem Informasi*, 17(2):89–98, 2021.
- Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2009.
- Rina Utami and Teguh Prasetyo. Model penetapan harga dinamis untuk umkm berbasis transaksi e-commerce. *Jurnal Ekonomi Digital Indonesia*, 1(1):45–53, 2022.
- Minhao Wang, Jing Liu, and Yu Zhao. Profit-aware dynamic pricing for e-commerce. *Electronic Commerce Research and Applications*, 48:101059, 2021.