

README: Analisis Churn dalam E-commerce

Pendahuluan

Di era digital yang terus berkembang, e-commerce telah menjadi bagian integral dari cara bisnis beroperasi. Namun, dengan pertumbuhan pesat ini, tantangan dalam memahami dan mempertahankan konsumen semakin kompleks. Churn, atau kehilangan konsumen, menjadi isu krusial yang harus diatasi oleh setiap perusahaan e-commerce. Dokumentasi ini bertujuan untuk menjelaskan tujuan, pendekatan, dan metrik yang digunakan dalam analisis churn konsumen.

Apa itu E-commerce?

E-commerce adalah proses jual beli barang dan layanan melalui internet, yang memanfaatkan platform digital untuk memudahkan transaksi antara individu dan perusahaan. Pertumbuhan pesat dalam sektor ini menjadikannya peluang usaha yang menguntungkan.

Latar Belakang

Menurut IBM, e-commerce telah mengalami transformasi signifikan sejak tahun 1990, menjadi ekosistem kompleks yang mendorong pertumbuhan ekonomi global. Dengan pemanfaatan teknologi, khususnya machine learning, perusahaan berkomitmen untuk menganalisis perilaku konsumen dan mengidentifikasi pola-pola churn.

Gambaran Proyek

Proyek ini berfokus pada analisis churn konsumen menggunakan teknik machine learning. Dengan mengidentifikasi faktor-faktor yang berkontribusi pada churn, kami bertujuan untuk membantu bisnis e-commerce meningkatkan strategi retensi.

Tujuan Analisis

1. **Identifikasi konsumen berisiko churn:** m data untuk mengetahui faktor-faktor yang mempengaruhi konsumen yang memiliki risiko tinggi untuk berhenti menggunakan layanan.
2. **Optimalkan strategi retensi:** merancang model machine learning yang efektif untuk mempertahankan konsumen.

Pendekatan Analisis

1. **Pengumpulan data:** mengumpulkan data relevan, termasuk demografi, preferensi, dan aktivitas transaksi konsumen.
2. **Model klasifikasi:** menggunakan algoritma machine learning untuk membangun model yang memprediksi kemungkinan churn berdasarkan pola perilaku konsumen.
3. **Evaluasi metrik:**
 - **Churn Rate:** menghitung tingkat churn untuk memahami proporsi konsumen yang berhenti menggunakan layanan.
 - **F2-Score:** mengoptimalkan pengelolaan churn dengan fokus pada recall, memastikan identifikasi konsumen berisiko tinggi yang lebih akurat.

Metrik Evaluasi

1. **Biaya churn:**
 - **Cost of Acquisition (CAC):** biaya untuk mendapatkan konsumen baru.

- **Cost of Retention (CRC):** biaya untuk mempertahankan konsumen yang ada. Penelitian menunjukkan bahwa mempertahankan konsumen lebih efisien dibandingkan akuisisi.

2. Jenis kesalahan dalam analisis:

- **Type 1 Error (False Positive):** mengklasifikasikan konsumen loyal sebagai churn.
- **Type 2 Error (False Negative):** gagal mengidentifikasi konsumen yang churn, yang lebih berdampak pada profitabilitas.

Analisis Data

Kami mengumpulkan dan menganalisis data, termasuk demografi konsumen, riwayat transaksi, dan umpan balik. Faktor-faktor kunci yang mempengaruhi churn diidentifikasi melalui analisis statistik, dan model prediktif dikembangkan untuk meramalkan kemungkinan churn.

Hasil

Analisis mengungkapkan beberapa faktor signifikan yang berkontribusi terhadap churn konsumen, beserta strategi yang dapat diterapkan untuk mengatasinya:

Faktor	Analisis	Aksi untuk Menghindari/ Mengatasi Churn	Penanggung Jawab
Complain	Tingginya churn di kalangan konsumen yang mengajukan keluhan.	Menerapkan sistem respons cepat untuk menangani keluhan dan melakukan follow-up untuk memastikan kepuasan.	Manajemen Eksekutif yang dapat berkolaborasi dengan Tim Customer Service
MaritalStatus	Konsumen single memiliki tingkat churn yang lebih tinggi.	Menyusun penawaran dan promosi khusus yang menarik untuk konsumen single.	Tim Pemasaran
PreferredOrderCat	Tingkat churn yang tinggi pada kategori produk mobile phone.	Meningkatkan kualitas produk dan memberikan jaminan atau garansi untuk meningkatkan kepercayaan konsumen.	Tim R&D, Tim Pemasaran
PreferredPaymentMode	Churn tinggi pada metode pembayaran cash on delivery dan e-wallet.	Meningkatkan keamanan transaksi dan memberikan insentif tambahan untuk penggunaan metode pembayaran tersebut.	Tim Pemasaran yang dapat berkolaborasi dengan Tim IT
PreferredLoginDevice	Masalah teknis pada pengguna komputer dapat berkontribusi pada churn.	Melakukan perbaikan teknis secara rutin terutama untuk komputer dan menyediakan tutorial penggunaan untuk meningkatkan pengalaman pengguna.	Tim R&D yang dapat berkolaborasi dengan Tim IT
CityTier	Tier 3 menunjukkan tingkat churn yang lebih tinggi.	Membangun infrastruktur yang lebih baik di tier 3 untuk memastikan layanan yang memadai, termasuk pengiriman yang lebih cepat.	Manajemen Eksekutif yang dapat berkolaborasi dengan Tim Operational
Tenure	Konsumen baru memiliki tingkat churn yang lebih tinggi.	Menerapkan program onboarding yang menarik dan memberikan penawaran khusus untuk konsumen baru selama periode awal.	Manajemen Eksekutif yang dapat berkolaborasi dengan Tim Customer Service
CashbackAmount	Cashback yang tidak menarik dapat mempengaruhi loyalitas.	Mengkaji dan meningkatkan program cashback agar lebih kompetitif dan menarik bagi konsumen.	Tim Pemasaran dan Manajemen Eksekutif
WarehouseToHome	Jarak yang lebih jauh dapat menyebabkan waktu pengiriman yang lebih lama.	Mengoptimalkan rute pengiriman dan menambah jumlah gudang di lokasi strategis untuk mempercepat proses pengiriman.	Manajemen Eksekutif yang dapat berkolaborasi dengan Tim Logistik

Kesimpulan dan Rekomendasi

Kesimpulan definisi churn:

Analisis menunjukkan bahwa churn tidak hanya dipengaruhi oleh demografi, preferensi, atau keterlibatan, tetapi dapat dilacak melalui indikator perilaku tertentu seperti:

- Perubahan status akun
- Logout dari aplikasi
- Uninstall aplikasi
- Berhenti dari keanggotaan

Rekomendasi untuk bisnis:

Untuk menangani faktor-faktor churn yang teridentifikasi dan meningkatkan retensi, kami merekomendasikan hal-hal berikut:

1. **Program pelatihan:** pelatihan rutin untuk semua tim tentang analisis data dan machine learning.
2. **Pelaporan terintegrasi:** mengembangkan sistem pelaporan yang lebih kohesif dengan Business Intelligence (BI) untuk visualisasi data secara real-time.
3. **Model machine learning yang lebih lanjut:** memperluas cakupan aplikasi machine learning ke area bisnis lain seperti perkiraan penjualan dan manajemen inventaris.
4. **Kemitraan:** bekerjasama dengan perusahaan analisis data untuk mendapatkan informasi yang lebih mendalam tentang perilaku konsumen.

Performa Model

Langkah-langkah yang diambil:

- **Fitur engineering:** mengatasi missing value, menggunakan OneHotEncoder untuk fitur kategorikal, dan menerapkan Random Over Sampling (ROS) untuk imbalanced data.
- **Evaluasi model:** melakukan tuning hyperparameter, memastikan kinerja yang kuat melalui teknik cross-validation.

Model terbaik:

Model LightGBM menunjukkan kemampuan prediksi yang sangat baik, dengan F2 Score meningkat dari 0.93 menjadi 0.94 setelah proses tuning. Biaya kesalahan klasifikasi model adalah yang terendah di antara model lainnya, menunjukkan efisiensi tinggi dalam mendeteksi konsumen yang berisiko churn.

Rekomendasi Modeling Selanjutnya

Beberapa langkah yang dapat dilakukan untuk pengembangan lebih lanjut meliputi:

- **Eksplorasi hyperparameter:** menggunakan teknik seperti Bayesian Optimization untuk menemukan parameter terbaik.
- **Cross validation yang kompleks:** menerapkan stratified k-fold untuk mengurangi kemungkinan overfitting.
- **Fitur baru:** mempertimbangkan fitur tambahan yang dapat memberikan informasi lebih tentang perilaku konsumen.
- **Pemantauan kinerja model:** mengimplementasikan sistem untuk memantau kinerja model secara berkala dan mendeteksi penurunan kinerja dengan cepat.

Streamlit:

Dalam konteks e-commerce, memahami perilaku konsumen dan memprediksi kemungkinan churn merupakan langkah penting untuk meningkatkan retensi. Dengan menggunakan Streamlit, penyusun membuat aplikasi sederhana namun cukup efektif, yang memungkinkan user untuk melakukan prediksi churn untuk satu konsumen dengan memasukkan data secara manual. Selain itu, aplikasi ini juga mendukung pengunggahan file CSV, sehingga user dapat menganalisis data churn dari banyak konsumen sekaligus. Dengan demikian, streamlit ini dapat menjadi alat yang berguna untuk membantu perusahaan dalam merumuskan strategi pencegahan churn dengan lebih baik. Streamlit dapat diakses pada **file "streamlit_finpro"** pada [GitHub](#).

Visualisasi Interaktif:

Visualisasi menggunakan Tableau dapat diakses melalui tautan berikut: [Lihat Visualisasi di Tableau](#).

About Us:

Notebook ini disusun oleh tim mindfuel: [FaUziah](#) dan [ELisa](#), sebagai proyek akhir dari program Job Connector Data Science Purwahika Bandung. Nama 'mind' kami pilih untuk merepresentasikan kreativitas dan inovasi yang kami miliki, sementara 'fuel' (yang juga merupakan singkatan nama kami) diharapkan dapat mencerminkan energi dan semangat yang diperlukan untuk mewujudkan ide-ide tersebut. Dengan nama ini kami ingin menunjukkan bahwa setiap gagasan yang muncul dari pikiran kami dapat menjadi bahan bakar untuk menciptakan perubahan kearah yang lebih positif.
