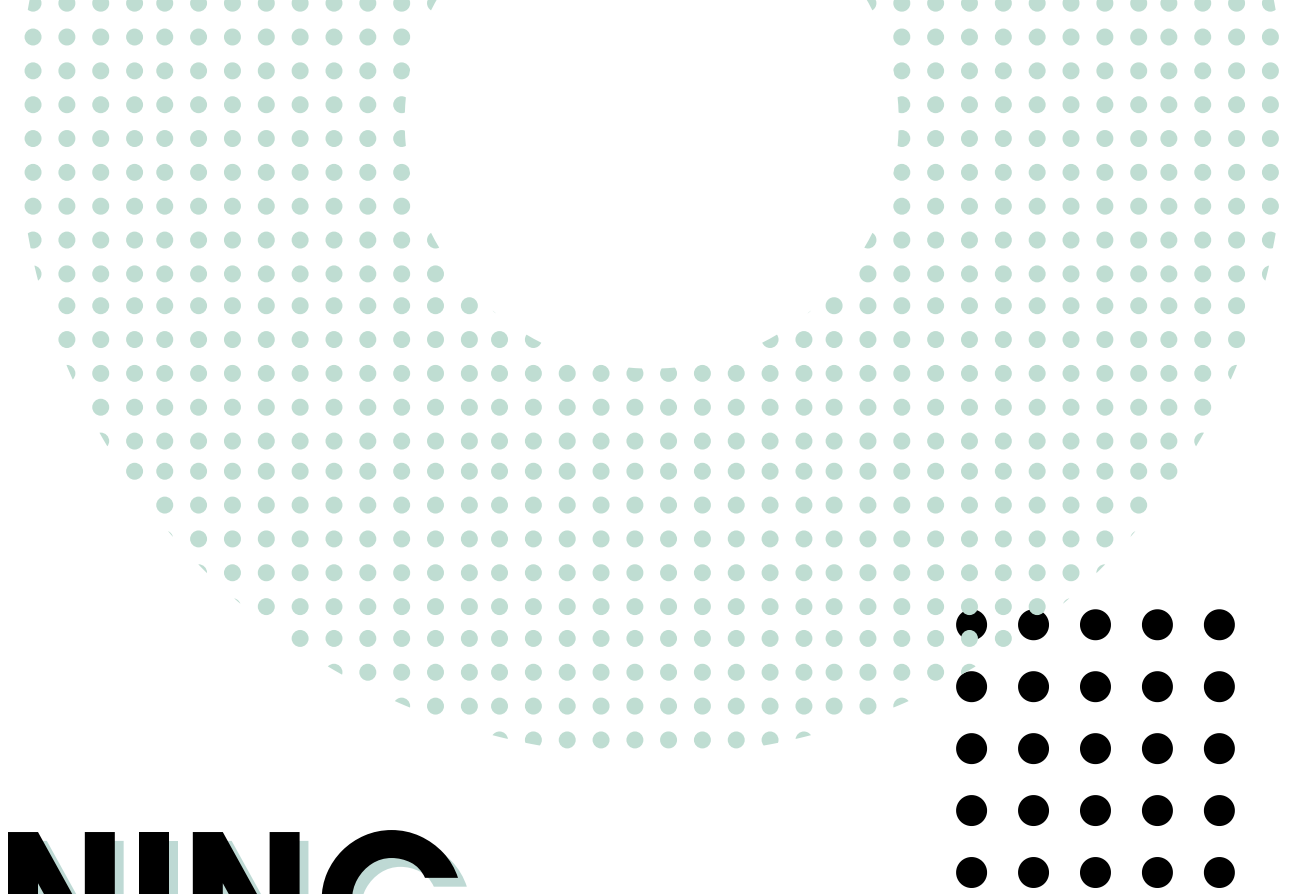


PREDIKSI PEMBATALAN BOOKING HOTEL DENGAN MACHINE LEARNING



Latar Belakang

Pembatalan pemesanan yang tidak terprediksi merugikan sebuah hotel. Pembatalan yang tidak terprediksi ini menyebabkan masalah yang berdampak kepada beberapa divisi. Proyek ini bertujuan menyediakan model machine learning sebagai dasar pengambilan keputusan agar hotel lebih siap dalam menghadapi pembatalan.

Dampak Permasalahan terhadap Divisi

Divisi	Masalah
Management	<ul style="list-style-type: none">• Perumusan dan evaluasi strategi• Kamar penuh tapi tak pasti dipakai
Front Office	<ul style="list-style-type: none">• Perencanaan staffing dan alokasi kamar yang terkendala• Slot kamar terisi dengan booking yang tak pasti
Marketing	<ul style="list-style-type: none">• Menjaga tingkat konversi sambil mendorong peningkatannya

Tujuan Pengembangan Machine Learning

Mengidentifikasi potensi pembatalan booking

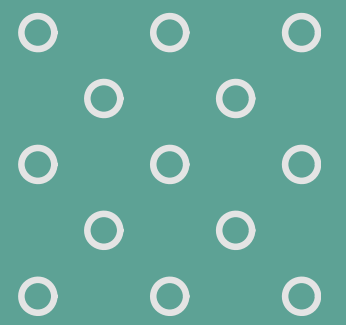
Mengidentifikasi kemungkinan pembatalan agar bisa dijadikan dasar perumusan strategi maupun pengambilan langkah preventif.

Mengklasifikasi kemungkinan pembatalan booking

Klasifikasi tingkat risiko pembatalan berdasarkan tinggi rendahnya kemungkinan batal

Mendukung pengambilan keputusan lintas divisi berbasis data

Output dari identifikasi dan klasifikasi kemungkinan batal membantu sebagai dasar tambahan dalam mengambil keputusan



Dataset dan Pendekatan Machine Learning

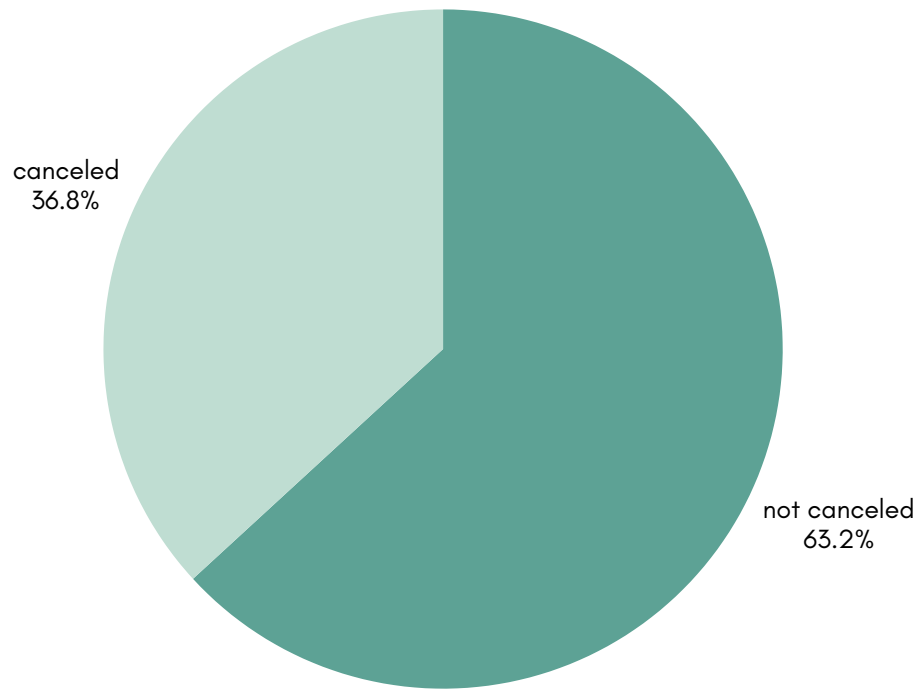


Dataset : Hotel Booking Demand

*Purwadhika for capstone version

- Jumlah Baris dan Kolom**
83573 baris
11 kolom
- Target dan Fitur**
Target : is_canceled
Fitur : kolom lainnya selain is_canceled

- Distribusi Pembatalan Booking**



Kolom dan Deskripsi

Kolom	Deskripsi
country	Negara asal pelanggan.
market_segment	Penunjukan segmen pasar.
previous_cancellations	Jumlah pemesanan sebelumnya yang dibatalkan.
booking_changes	Jumlah perubahan pada pemesanan.
deposit_type	Indikasi apakah pelanggan membayar deposit.
days_in_waiting_list	Jumlah hari dalam daftar tunggu.
customer_type	Jenis pemesanan.
reserved_room_type	Kode tipe kamar yang dipesan.
required_car_parking_space	Jumlah ruang parkir yang diminta.
total_of_special_request	Jumlah permintaan khusus pelanggan.
is_canceled	Indikasi pembatalan (1 = ya, 0 = tidak).

Pendekatan Machine Learning

Model : Randomforest

- Robust terhadap overfitting
- Menggabungkan banyak perspektif untuk hasil yang lebih stabil
- Cocok untuk data kompleks yang mencerminkan kondisi di lapangan.

Metric : Recall

Menangkap sebanyak mungkin calon tamu yang berpotensi membatalkan pesanan (mengurangi false negative rate).

Pros :

Membantu tim mengenali lebih banyak tamu berisiko

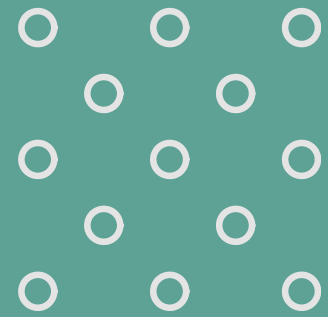
Cons :

Follow up dari hasil prediksi bisa jadi salah sasaran





Hasil Train



Recall Score:

CV score

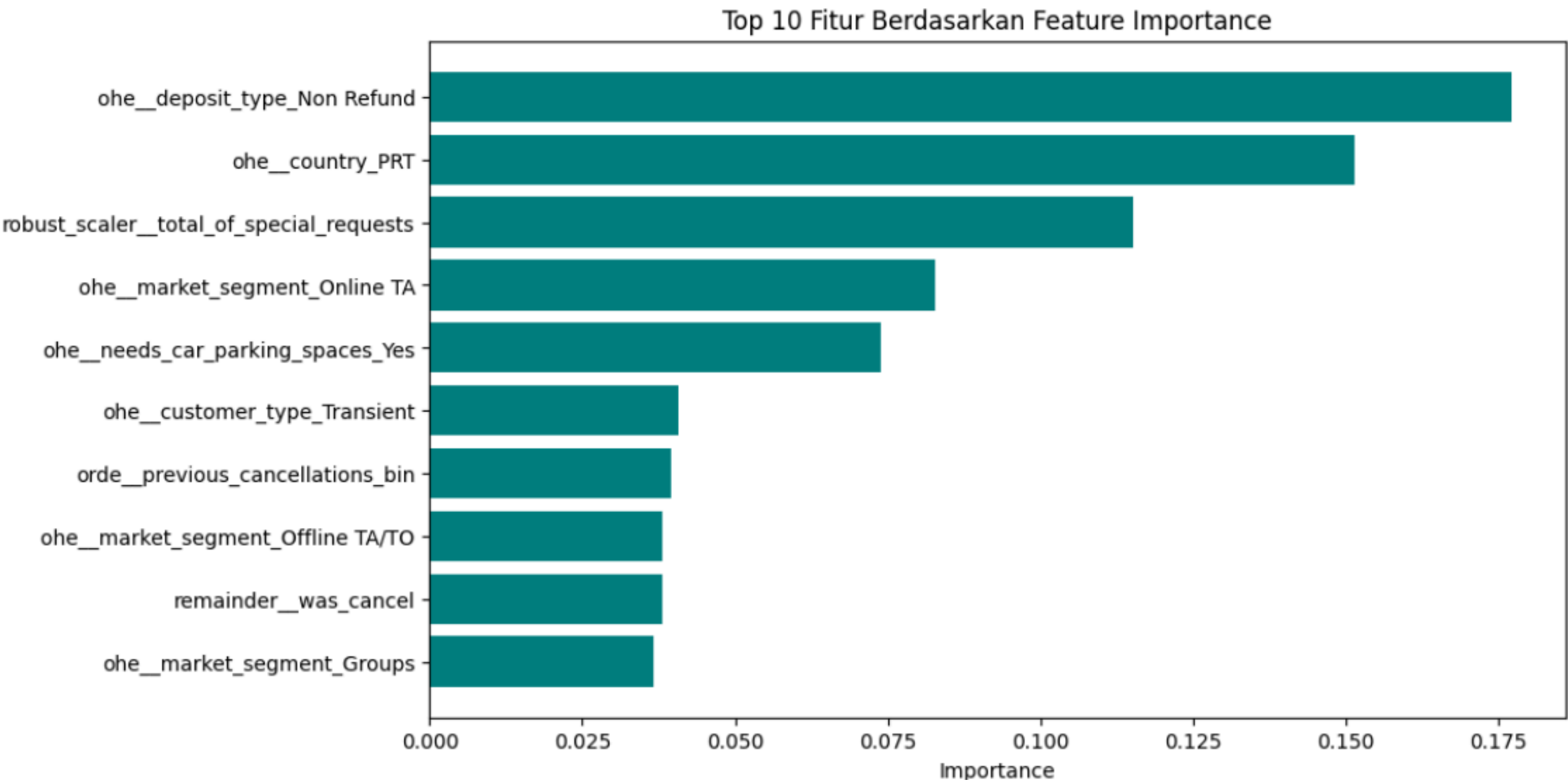
Best CV Score: 0.783

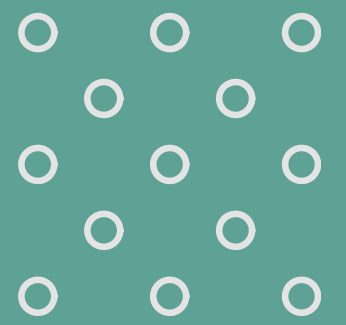
Rata-rata skor CV: 0.769

Standar Deviasi : 0.01

Test score : 0.769

feature importance





Confusion Matrix dan Implikasinya dalam Bisnis



1. Kerugian dari False Positive (FP)

False Positive terjadi ketika model memprediksi bahwa tamu akan membatalkan, padahal sebenarnya tidak. Dalam kondisi ini hotel berpotensi mengalami dua jenis kerugian:

- Intervensi yang tidak diperlukan
- Konsekuensi operasional pada kondisi overbook

2. Kerugian dari False Negative (FN)

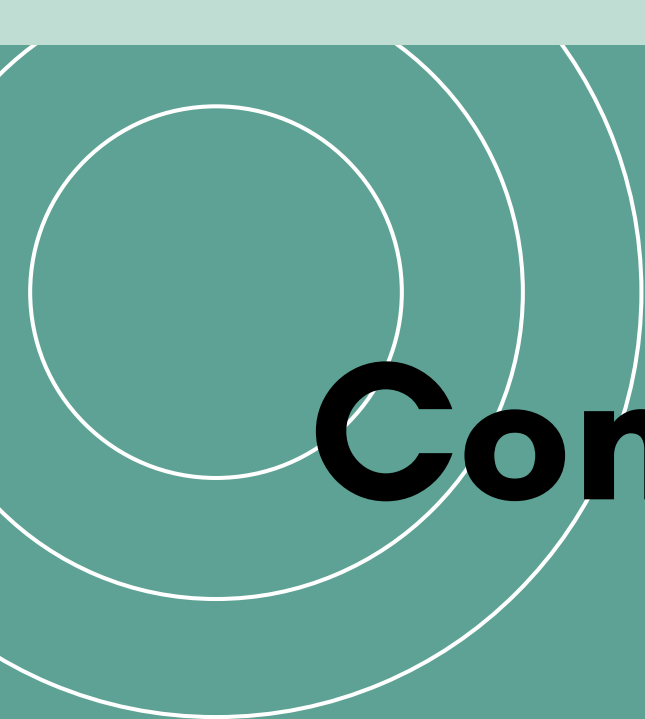
False Negative terjadi ketika model memprediksi bahwa tamu **tidak akan membatalkan**, padahal sebenarnya **tamu batal**. Dalam kondisi ini hotel berpotensi mengalami kerugian berupa:

- Potential Revenue Loss

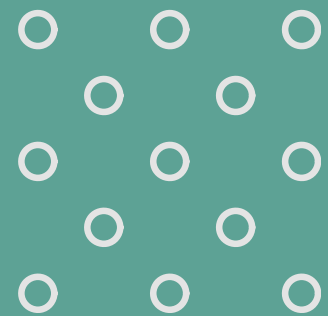
3. Keuntungan dari True Positive (TP)

True Positive terjadi ketika model memprediksi bahwa **tamu akan membatalkan**, dan **prediksi ini benar**. Hotel dapat melakukan intervensi preventif, misalnya memberikan promosi atau insentif, sehingga pembatalan dapat dicegah. Dampak bisnisnya meliputi:

- Pendapatan yang dipertahankan
- Realokasi kamar (khusus high season)



Contoh Penggunaan



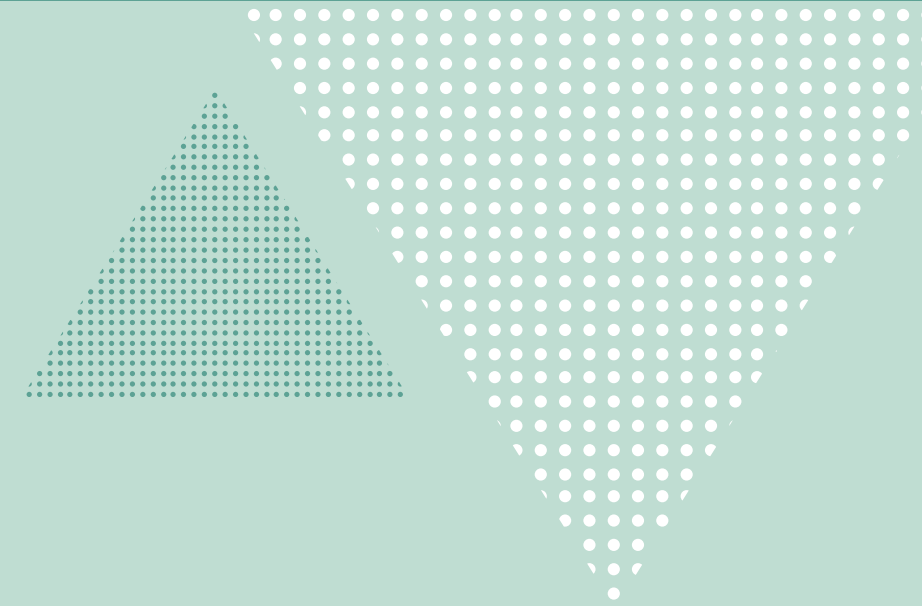
Tabel Probabilitas

- Dibuat berdasarkan tingkat kemungkinan batal
- Klasifikasi bisa disesuaikan dengan kebutuhan perusahaan
- Memabantu sebagai decision support

	Prob_0 (%)	Prob_1 (%)	Risk_Level	Action
14970	14.31	85.69	Tinggi	Tawarkan insentif
15366	23.75	76.25	Tinggi	Tawarkan insentif
1538	63.12	36.88	Sangat Rendah	Tidak perlu follow-up
6835	28.40	71.60	Tinggi	Tawarkan insentif
357	97.78	2.22	Sangat Rendah	Tidak perlu follow-up
11461	37.59	62.41	Sedang	Tawarkan engagement
6801	0.24	99.76	Sangat Tinggi	Kontak manual
16260	77.84	22.16	Sangat Rendah	Tidak perlu follow-up
4365	99.69	0.31	Sangat Rendah	Tidak perlu follow-up
14565	34.37	65.63	Sedang	Tawarkan engagement
12608	77.30	22.70	Sangat Rendah	Tidak perlu follow-up
6295	50.32	49.68	Rendah	Pantau lebih lanjut
13504	96.27	3.73	Sangat Rendah	Tidak perlu follow-up
1079	99.22	0.78	Sangat Rendah	Tidak perlu follow-up
8280	32.95	67.05	Sedang	Tawarkan engagement



Rekomendasi dan Kesimpulan



Rekomendasi Untuk Divisi

Manajemen

- Segmentasi risiko pembatalan tamu digunakan sebagai dasar pendukung dalam membuat dan mengevaluasi strategi promosi serta kebijakan operasional, sehingga keputusan dapat disesuaikan dengan tingkat risiko pembatalan pada masing-masing segmen tamu.

Front Office/Operasional

- Informasi risiko pembatalan membantu front office dan operasional dalam perencanaan staf dan alokasi kamar, khususnya untuk mengantisipasi potensi pembatalan tanpa mengubah kebijakan layanan secara langsung.

Marketing

- Prediksi risiko pembatalan dimanfaatkan sebagai informasi tambahan bagi bagian marketing untuk menjaga dan meningkatkan konversi reservasi menjadi kedatangan aktual, melalui pendekatan yang lebih selektif terhadap segmen berisiko.

Kesimpulan

Pengembangan model machine learning dalam studi ini ditujukan sebagai alat pendukung pengambilan keputusan melalui segmentasi risiko pembatalan tamu. Dengan menyediakan informasi tambahan yang relevan, model ini dapat membantu berbagai divisi dalam merespons potensi pembatalan secara lebih terkoordinasi, tanpa menggantikan peran pengambilan keputusan manajerial. Adapun pengaruh model terhadap bisnis sepenuhnya bergantung pada kemampuan manajemen dalam menginterpretasikan output model dan menggunakannya sebagai dasar pengambilan keputusan yang tepat.



Terima Kasih

O