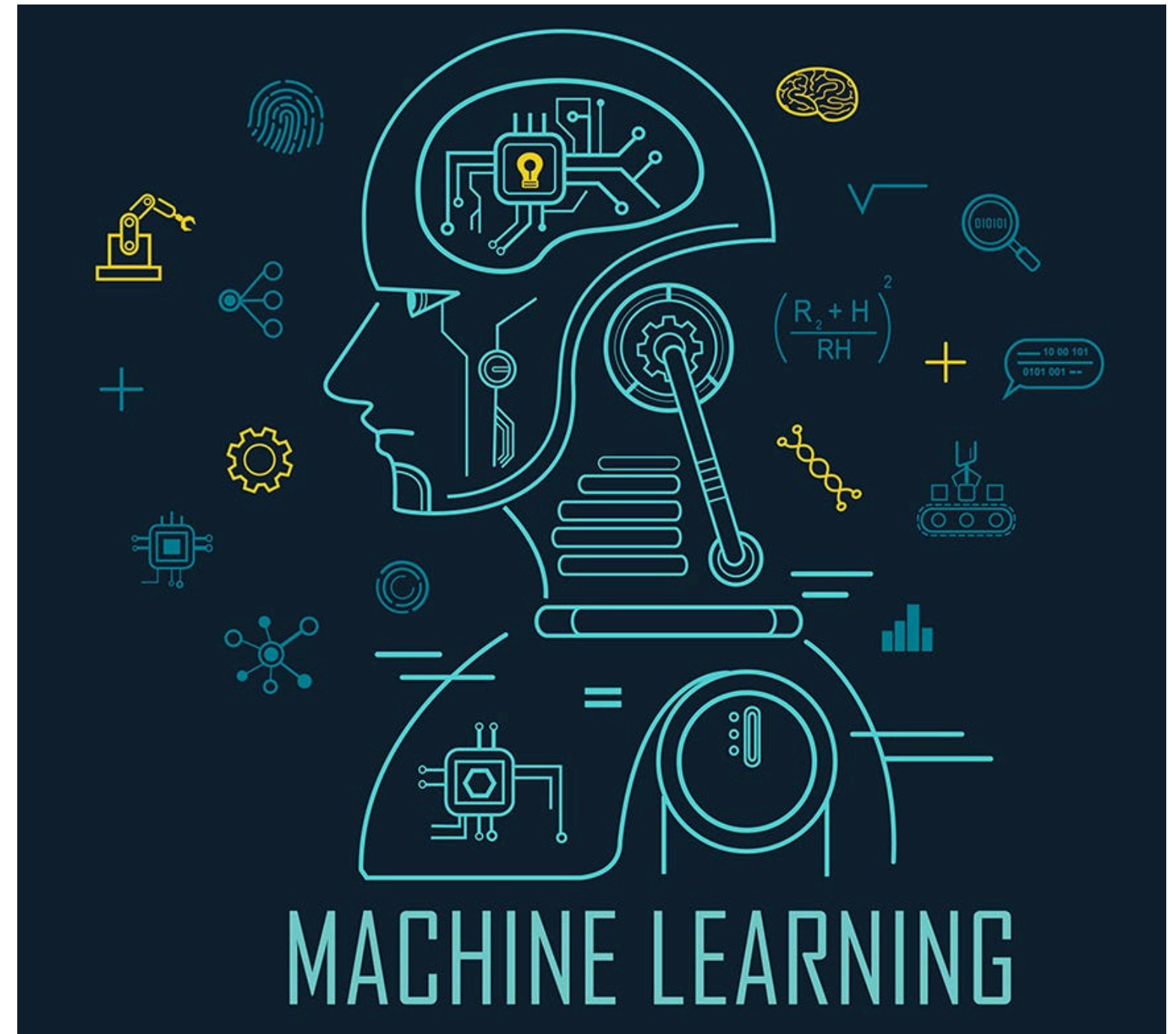


MACHINE LEARNING

Machine Learning untuk Memprediksi Cancel
Pada Pengunjung Hotel Portugal

Yogyakarta, 27 Mei 2024



Habib Ja'far Nuur
Capstone 3

LIST OF CONTENTS

03	BUISNIS PROBLEM UNDERSTANDING
04	MATRIX EVALUATION
05	DATA PREPARATION
07	FEATURE
08	EXPERIMENT
09	CARA KERJA MODEL
10	EVALUASI MODEL
11	KINERJA MODEL
12	BATASAN-BATASAN
13	SIMULASI
14	CONCLUTION DAN RECOMENDATION

BUSNIS PROBLEM UNDERSTANDING

Sebuah hotel yang berada di Portugal dalam kegiatan pelayanannya seringkali menerima pembatalan mendadak yang dilakukan oleh pemesan. Hal ini menyebabkan kerugian finansial yang dialami pihak hotel dan mempengaruhi reputasi bisnis.

GOAL

Pihak Hotel ingin memiliki kemampuan untuk mendeteksi pelanggan mana saja yang berpotensi untuk melakukan pembatalan pemesanan

Analytic Approach

Pihak hotel ingin membangun system yang dapat membantu hotel untuk memprediksi pengunjung yang berpotensi membatalkan pesanan

MATRIX EVALUATION

F2 Score

Type 1 error: False Positif

Konsekuensi= pihak hotel harus memberikan upgrade kamar jika kamar yang sama telah penuh. potensi mengalami kerugian akibat upgrade kamar sebesar \$18,69

Type 2 Error: False Negative

konsekuensi: Pihak hotel mengalami kerugian dikarenakan kamar yang seharusnya terisi menjadi kosong atau tidak tersewakan. kerugian yang dialami pihak hotel akibat tidak tersewanya kamar sebesar \$99

DATA PREPARATION

MASALAH

- TERDAPAT DUPLICATE
- TERDAPAT OUTLIER
- TERDAPAT CORELASI ANTAR FEATURE
- IMBALACE

PENANGANAN

- HAPUS DUPLICATE
- MENGHAPUS OUTLIER DALAM EXPERIMENT
- MEMILIH SALAH SATU FEATURE YANG TERDAPAT KORELASI
- MENANGANI IMBALANCE DENGAN EXPERIMENT

KORELASI ANTAR FEATURE

Terdapat korelasi antar feature yang moderate yang berpotensi collinearity maka kolom tersebut tidak di masukkan kedalam feature. kolom tersebut yaitu country,market segment,deposit_type tidak dimasukkan dalam fitur

	country	market_segment	deposit_type	customer_type	reserved_room_type	is_canceled
country	1.000000	0.452452	0.550297	0.238182	0.238119	0.444530
market_segment	0.452452	1.000000	0.508427	0.556138	0.280519	0.359348
deposit_type	0.550297	0.508427	1.000000	0.103519	0.245738	0.297877
customer_type	0.238182	0.556138	0.103519	1.000000	0.182412	0.200644
reserved_room_type	0.238119	0.280519	0.245738	0.182412	1.000000	0.097515
is_canceled	0.444530	0.359348	0.297877	0.200644	0.097515	1.000000

FEATURE

Machine learning secara otomatis belajar melalui pola atau struktur dari feature untuk memprediksi secara baik.

- 1 PREVIOUS CANCELLATIONS**
- 2 BOOKING CHANGES**
- 3 DAYS IN WAITING LIST**
- 4 CUSTOMER TYPE**
- 5 RESERVED ROOM TYPE**
- 6 REQUIRED CAR PARKING SPACES**
- 7 TOTAL OF SPECIAL REQUESTS**

EXPERIMENT MODEL MACHINE LEARNING

**Experiment 5 Model Gradient
Boosting(Final Model)**

Encoding, Nearmis (0.7726)

01

Experiment 1

Melakukan encoding
(0.6882)

02

Experiment 2

encoding, Binning (0.6945)

03

Experiment 3

encoding, Binning,
Polynomial (0.7147)

04

Experiment 4

Encoding, Polynomial
(0.6881)

05

Experiment 5

Encoding, Nearmis (0.7726)

06

Experiment 6

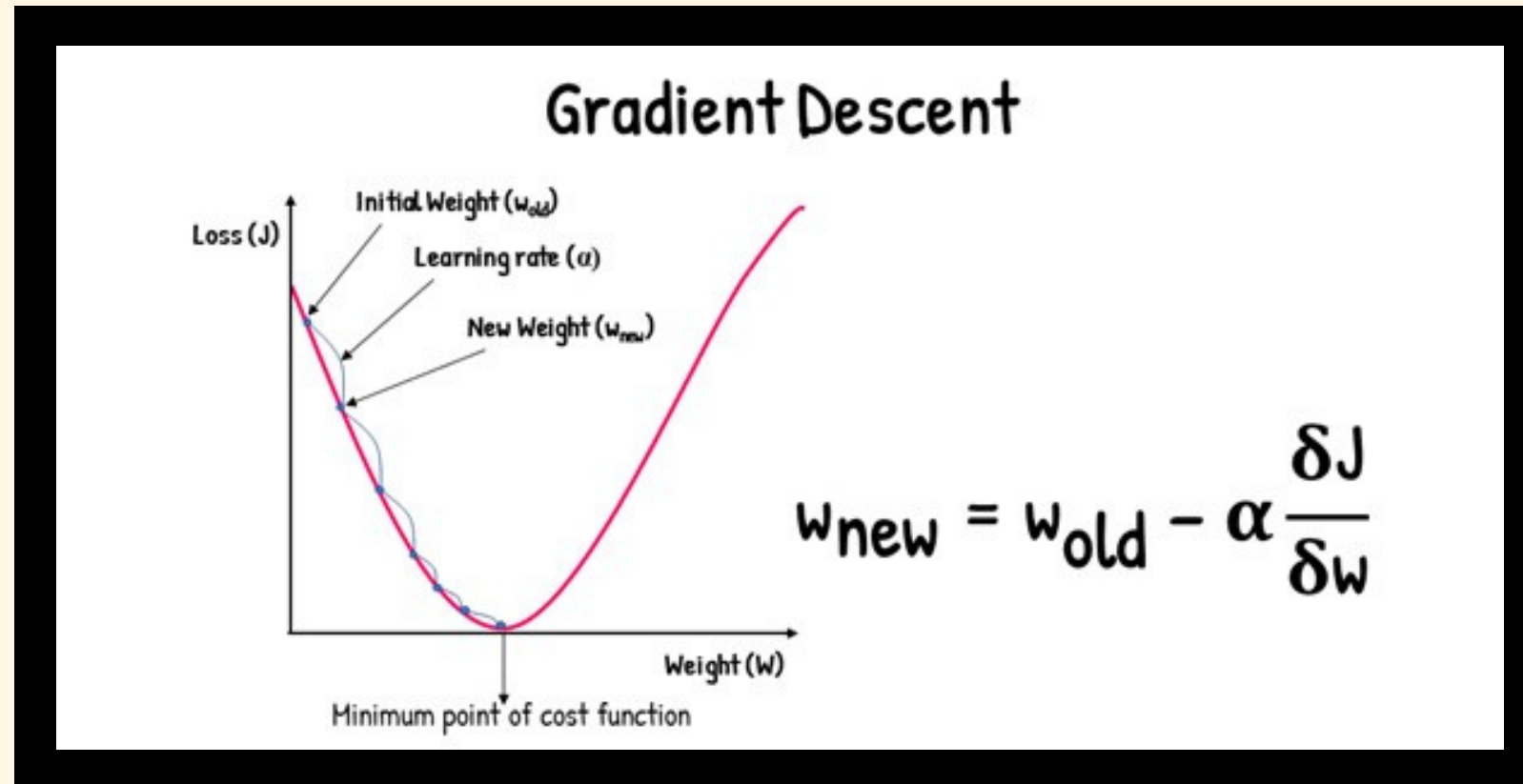
Encoding, Nearmis, Votting
Classifier (0.7746)

07

Experiment 7

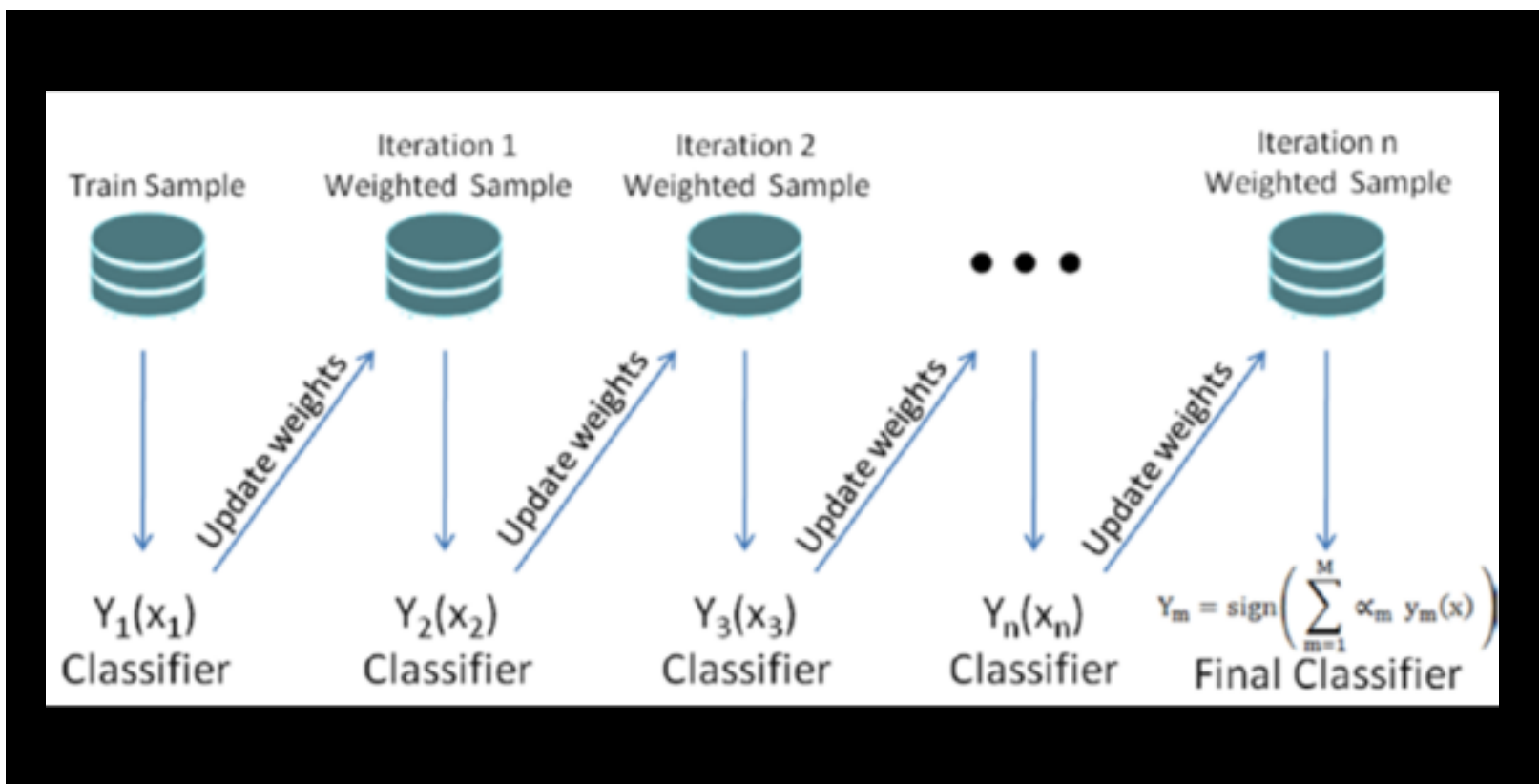
Encoding, Feature Selection
(0.7169)

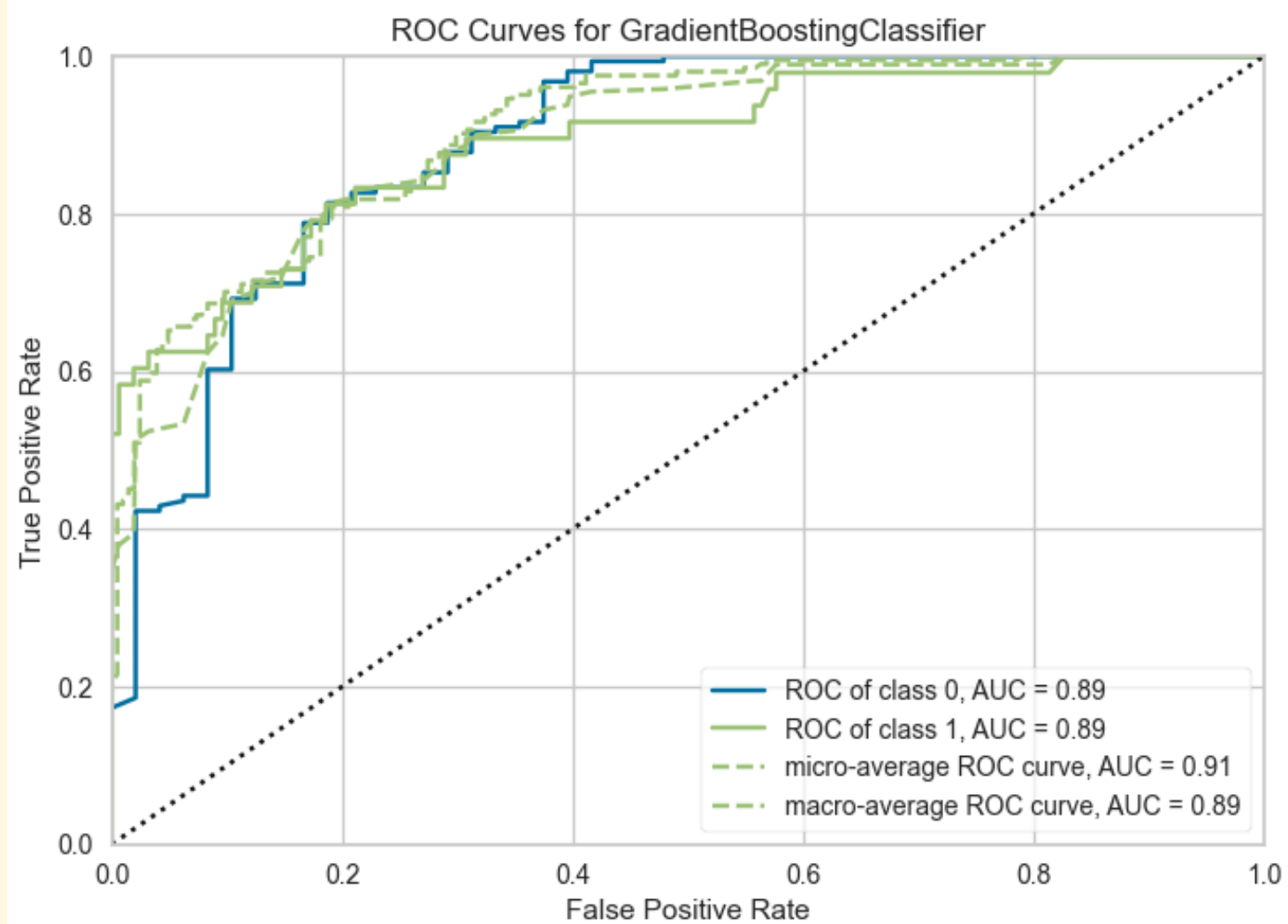
CARA KERJA MODEL



Final Model yang digunakan yaitu gradient Boosting. Pada setting default yang kita gunakan gradient boosting yang kita gunakan berbasis tree. Model ini dilakukan boosting dengan menggunakan penurunan gradient untuk melakukan pengupdatean weight dengan menggunakan loss log sebagai pengurang pada weight old.

Boosting merupakan suatu proses machine learning yang bersifat iteratif. Boosting merubah model weak learner menjadi strong learner. Weak learner merupakan model machine learning yang performanya sedikit lebih dari pada kita melakukan klasifikasi tetapi secara acak. Boosting memperbaiki model dengan memberikan perhatian lebih terhadap data yang diprediksi salah dengan memberikan bobot/weight yang lebih terhadap data yang diprediksi salah. Proses ini terjadi secara terus menerus sampai pada titik tertentu.

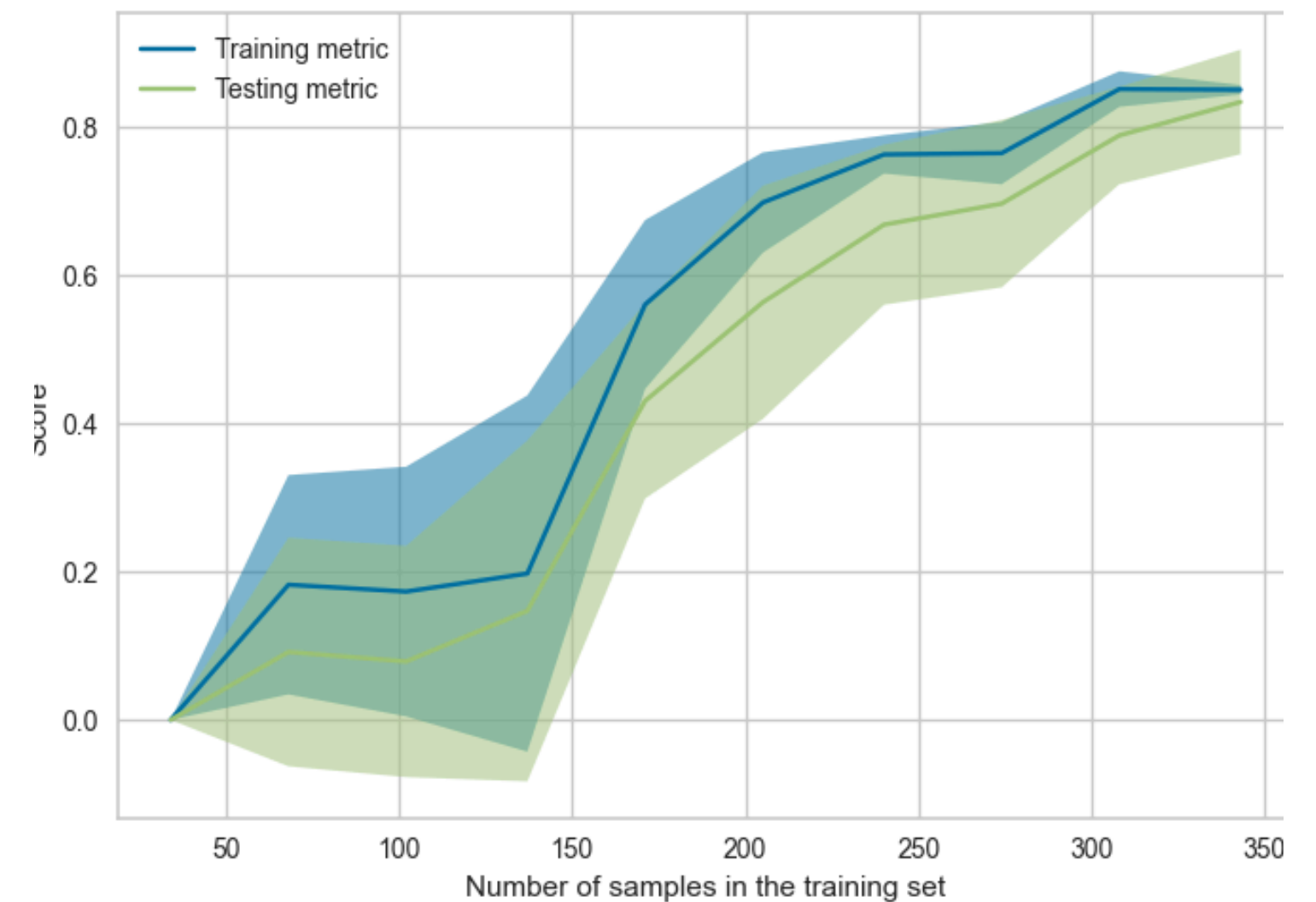




EVALUASI MODEL

Dari grafik disamping kita dapat melihat bahwa model gradien boositing dapat membedakan 2 kelas dengan baik karena memiliki AUC yang mendekati 0.9. Nilai AUC yang semakin mendekati 1 maka semakin baik model membedakan 2 kelas dengan baik.

Dari hasil leraning curve disamping dapat lihat bahwa train size yang kita tetapkan tidak terjadi overfitting ataupun underfitting



KINERJA OPTIMAL MACHINE LERANING

Machine Learning dapat bekerja dengan baik jika berada pada rentang probabilitas tertentu serta kinerja model kurang baik jika berada pada rentang probabilitas tertentu juga

Rentang kinerja terbaik

Rentang probabilitas machine learning ini dapat bekerja dengan baik jika berada di rentang 0.00-0.50 dan 0.80-1.00. Dengan F2 score diatas 90%

Rentang kinerja yang kurang baik

Rentang probabilitas Mechine Learning yang mengalami penurunan berada pada rentang 0.50-0.80. dengan F2 score sebesar 69-75%

BATASAN MACHINE LEARNING

Machine Learning ini di buat dengan batasan-batasan saat melakukan training data. Jika ada data yang di input diluar data training maka hasil prediksi tidak dapat dipercaya

1. previous_cancellations memiliki rentang antara 0-26
2. booking_changes memiliki rentang anantara 0-21
3. days_in_waiting_list diantara rentang 0-391
4. total_of_special_requests diantara rentang 0-5
5. tipe dari customer type antara lain: Transient-Party, Transient, Contract, Group
6. kategori dari reserved room type antara lain: A, C, E, D, B, H, F, G, P, L

SIMULASI MACHINE LEARNING

Simulasi untuk melihat efek penggunaan machine learning dalam meminimalkan kerugian dari cancelation hotel. Dengan Machine Learning dapat memperkecil kerugian sebesar \$ 4132.29



255

PEMESAN HOTEL

\$ 19061.13

Kerugian Tanpa Machine Learning. 100% kesalahan Prediksi

\$6488.13

Kerugian Tanpa Machine Learning. 50% kesalahan Prediksi

\$ 2355.84

Kerugian Dengan Menggunakan Machine Learning.

CONCLUTION & REKOMENDATION

Conclution

machine learning untuk mendeteksi cancelation hotel di portugal menggunakan model gradient boosting dengan matrix evaluation F2 score sebesar 77.26%. Dengan menggunakan Machine learning hotel dapat menekan kergian sebesar \$ 4132.29 dibandingkan tanpa menggunakan machine learning.

Recomendation

Dalam menggunakan machine learning ini harus memperhatikan batasan-batasan yang telah ditetapkan agar hasil prediksi akurat.