



INSTITUT TEKNOLOGI
SUMATERA

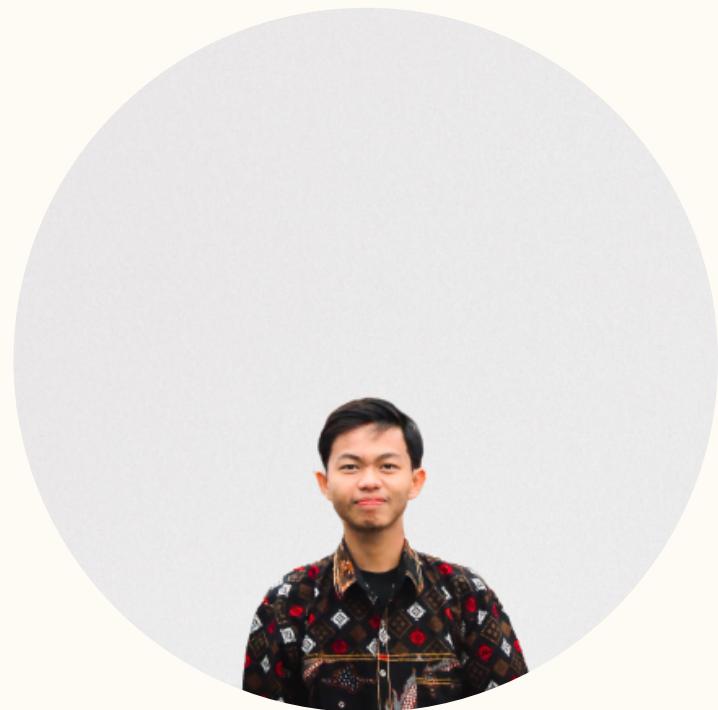
TERRARIUM

BPJS Healthkathon 2022

Machine Learning

Anggota TERRARIUM

[Back to Overview](#)



Lanang Almasyuri
Anggota Tim



**Muhammad Dhoni
Apriyadi**
Ketua Tim

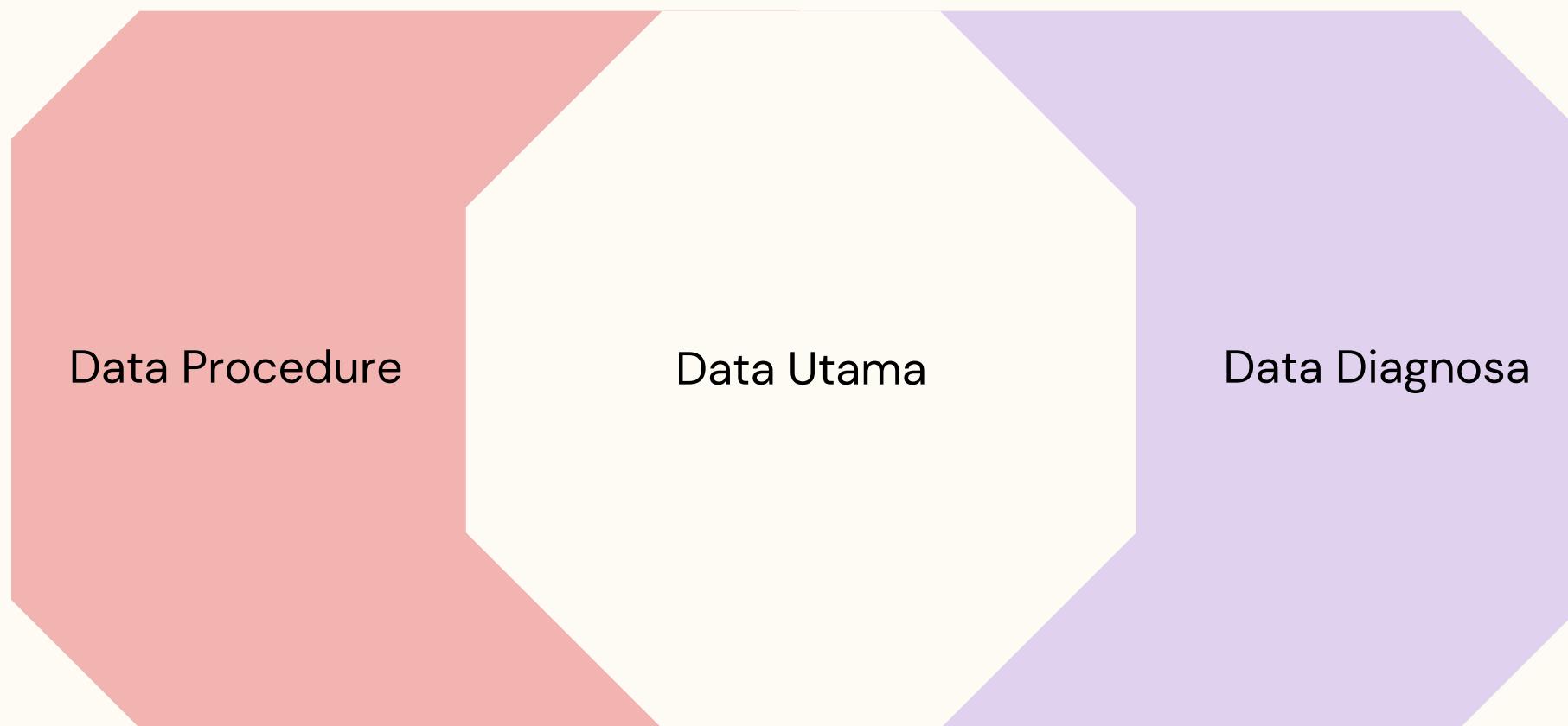


**Shahnaz Salsabila
Ishak**
Anggota Tim

Overview

- Introduction
- Data preparation
- Features selection
- Features engineering
- Pemilihan algoritma
- Save and load Model

Data Preparation



- Mengisi Data yang Nilainya Kosong
- Merubah Tipe Data
- Melakukan remove beberapa baris data untuk data yang tidak memiliki format yang sesuai

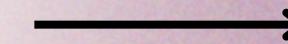
[Back to Overview](#)

Data Utama

[Back to Overview](#)

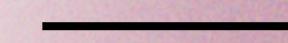
PENGISIAN PADA DATA YANG HILANG

Data Nominal



'N'

Data Ratio



median

Data Preparation

MENGUBAH TIPE DATA

Int64



Int8

Int16

Int32

tanggal (string/object)



datetime

Data Procedure

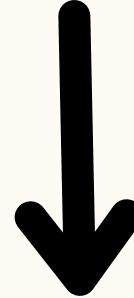
[Back to Overview](#)

Aturan penulisan procedure berdasarkan ICD-9-CM

4 digit angka

2 digit pertama : kode dasar.
karakter setelah desimal menunjukkan
subkategori procedure lebih spesifik

titik desimal antara digit
ke 2 dan 3



Menghapus baris yang formatnya tidak sesuai dengan aturan
penulisan kode procedure

Abjad

simbol seperti '-'

Data Procedure

[Back to Overview](#)

Mengambil 2 digit pertama kode procedure

90.59 → 90

69.01 → 69

88.78 → 88

Menggabungkan procedure berdasarkan id

id	procedure
6	90
6	69
6	88
27	90
27	88

→

id	procedure
6	90, 69, 88
27	90, 88

Data Diagnosa

[Back to Overview](#)

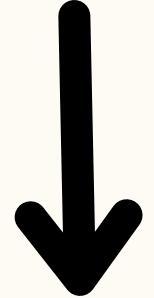
Aturan penulisan diagnosa berdasarkan ICD-10

3 - 7 karakter

titik desimal antara
karakter ke 3 dan 4

3 karakter pertama : kode dasar.
karakter setelah titik desimal menunjukkan
subkategori diagnosa lebih spesifik

diawali 1 huruf
abjad dengan
huruf besar



Filter kode diagnosa yang mengandung abjad

Mengubah abjad di semua kode diagnosa menjadi huruf besar

Data Preparation

Data Diagnosa

[Back to Overview](#)

Mengambil 3 karakter pertama kode diagnosa

006.9 → 006

M10.99 → M10

T14.70 → T14

Menggabungkan diagnosa dan levelid berdasarkan id

id	diagnosa	levelid
11426954	M79	2
11426954	R51	1
11426954	JO6	2
11426957	M79	1
11426957	K30	2
11426957	R10	2

id	diagnosa	levelid
11426954	M79, R51, JO6	1,2
11426957	M79, K30, R10	1,2

Features Selection

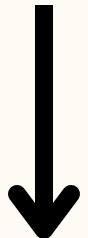
[Back to Overview](#)

Data Utama

[Back to Overview](#)

22 features

id	id_peserta	dati2	typefaskes	usia	jenkel	pisat	tgldatang	tglpulang	label
----	------------	-------	------------	------	--------	-------	-----------	-----------	-------	-------



7 features

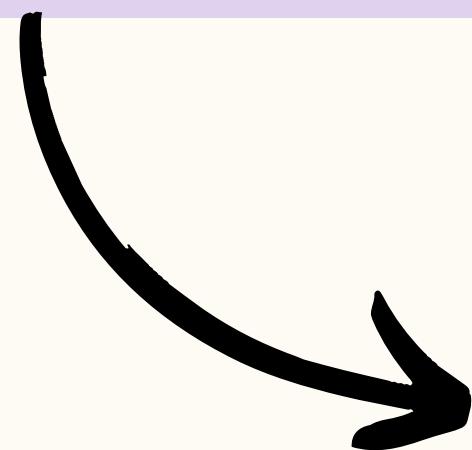
id	dati2	tgldatang	tglpulang	politujuan	biaya	label
----	-------	-----------	-----------	------------	-------	-------

Features Selection

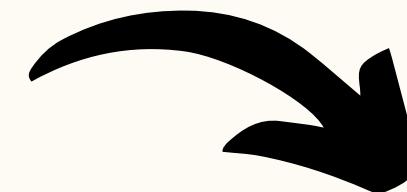
Data Utama

[Back to Overview](#)

Memilih politujuan yang sering muncul pada data yang diberi label 1 atau memiliki potensi inefisiensi pada klaim



Memilih politujuan yang
muncul lebih dari 500 kali



politujuan	sum
IGD	20800
MAT	19288
INT	18755
BED	15040

23 politujuan

Features Selection

*preview

Procedure dan Diagnosa

[Back to Overview](#)

Memilih kode procedure dan kode diagnosa yang sering muncul pada data yang diberi label 1 atau memiliki potensi inefisiensi pada klaim

Memilih kode procedure yang muncul lebih dari 500 kali

26
procedure

procedure	sum
99	26634
89	19288
90	18755
93	15040

*preview

Memilih kode diagnosa yang muncul lebih dari 1500 kali

30
diagnosa

diagnosa	sum
PO3	13219
H52	8112
E11	5671
K30	4716

*preview

Features Selection

Features Engineering

[Back to Overview](#)

Data Utama

[Back to Overview](#)

Menambah fitur untuk menentukan apakah pasien menjalani rawat inap

Menghitung selisih tanggal datang dan
tanggal pulang



Data Utama

[Back to Overview](#)

**Memberikan variabel dummy pada setiap setiap data berdasarkan
politujuan yang telah didapatkan pada features selection**

id	label	IGD	MAT	INT	HDL
165666	0	0	0	0	0
166042	0	0	0	1	0
168937	0	0	1	0	0

Procedure

[Back to Overview](#)

Memberikan variabel dummy pada setiap setiap data berdasarkan kode procedure yang telah didapatkan pada features selection

id	99	89	90	56
1	0	0	1	0
3	0	0	1	0
4	0	0	1	0

Diagnosa

[Back to Overview](#)

Memberikan variabel dummy pada setiap setiap data berdasarkan kode diagnosa yang telah didapatkan pada features selection dan levelid

id	H52	E11	K30	J06	1	2
2	0	0	0	0	1	0
3	1	0	0	0	1	0
4	0	0	0	0	1	0
5	0	0	1	0	1	0

Pemilihan Algoritma

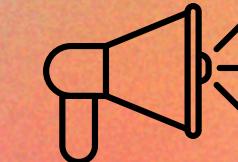
[Back to Overview](#)

**eXtreme Gradient Boosting
(XGBoost)**



Mengapa Menggunakan Metode XGBoost ?

Ada beberapa alasan yang mempengaruhi kami menggunakan metode ini.



Kelebihan Menggunakan Metode XGBoost

Banyak keuntungan yang dapat dipertimbangkan dan menjadi alasan kami menggunakan metode ini.

[Back to Overview](#)

	Precision	recall	F1- Score
0	0.94	0.99	0.96
1	0.92	0.57	0.70

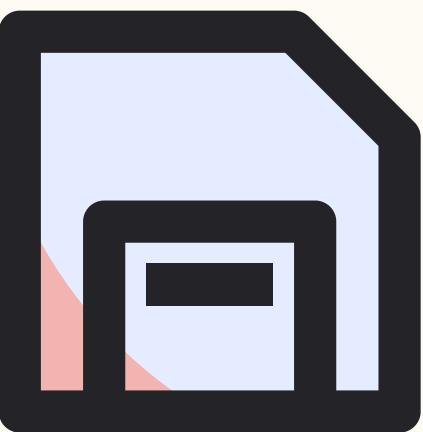
Recall : 0.57

Precision : 0.92

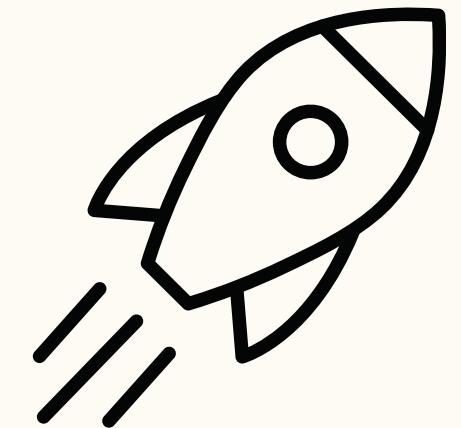
Accuracy : 0.9162

Specifity : 0.9973

Save and Load Model



SAVE



LOAD

[Back to Overview](#)

Q&A Session

Thank you for listening!