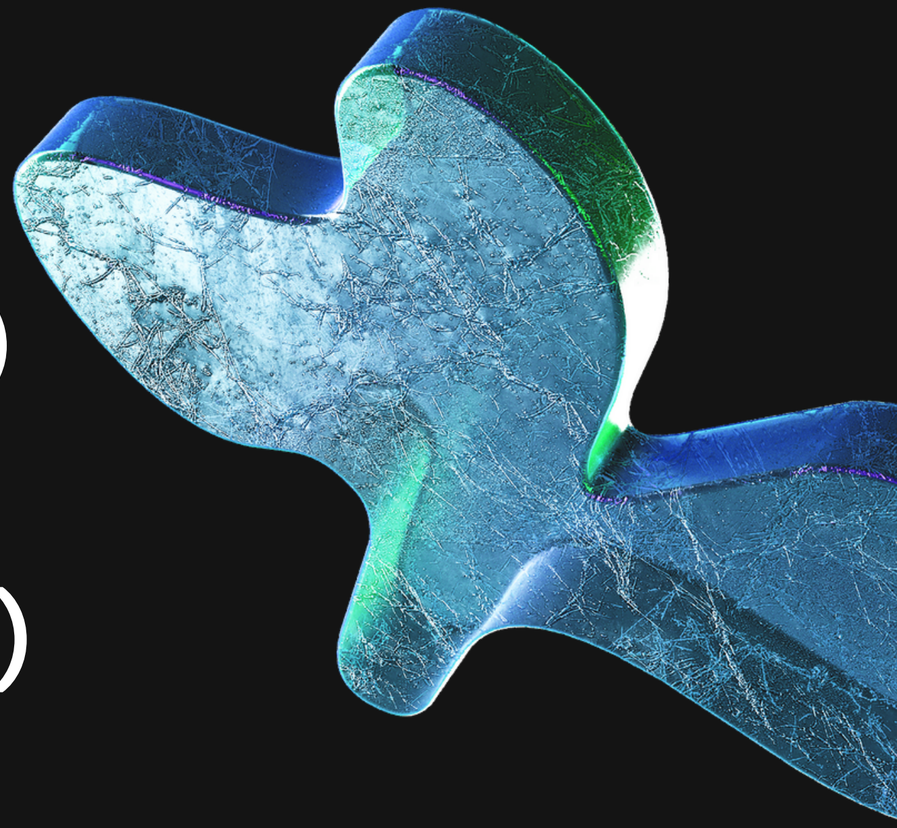
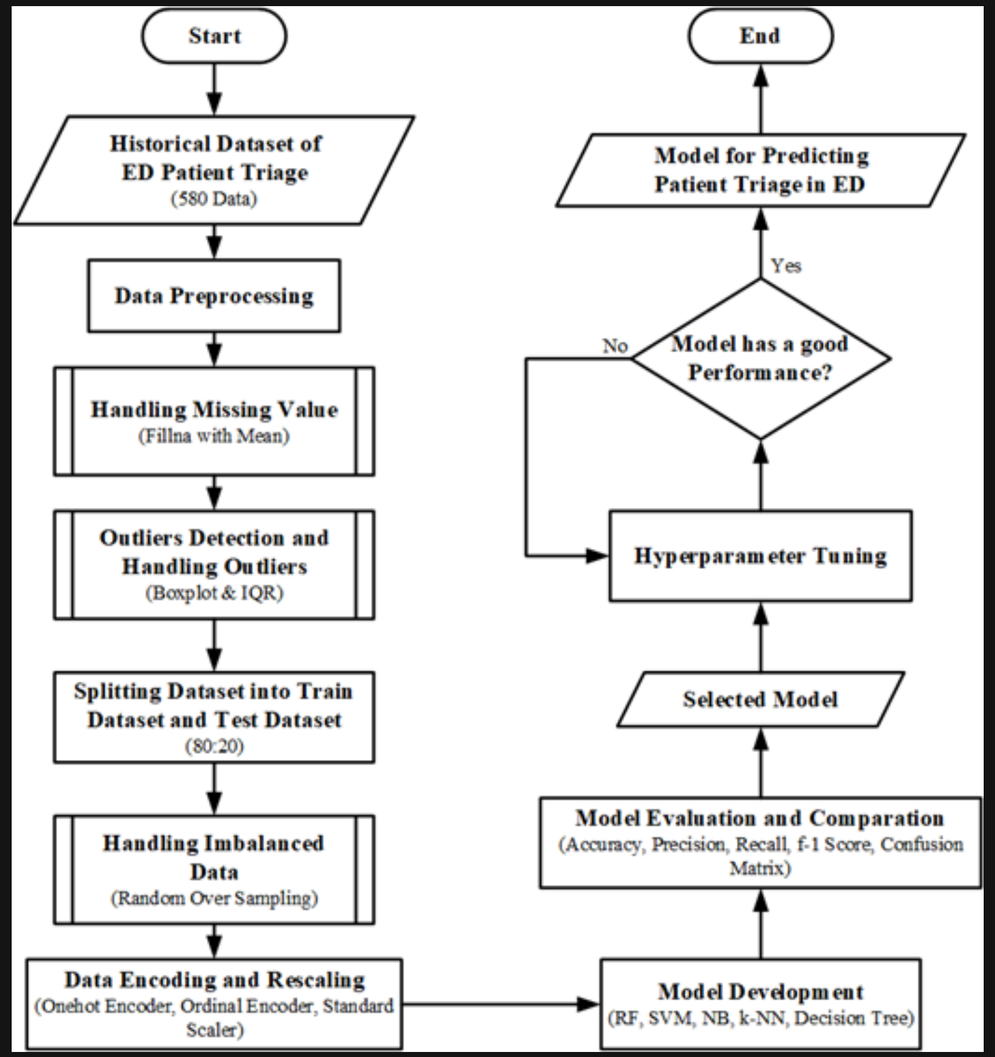
PERAMALAN TRIASE PASIEN INSTALASI GAWAT DARURAT (IGD) MENGGUNAKAN PENDEKATAN KLASIFIKASI MACHINE LEARNING (STUDI KASUS: RUMAH SAKIT UNS)

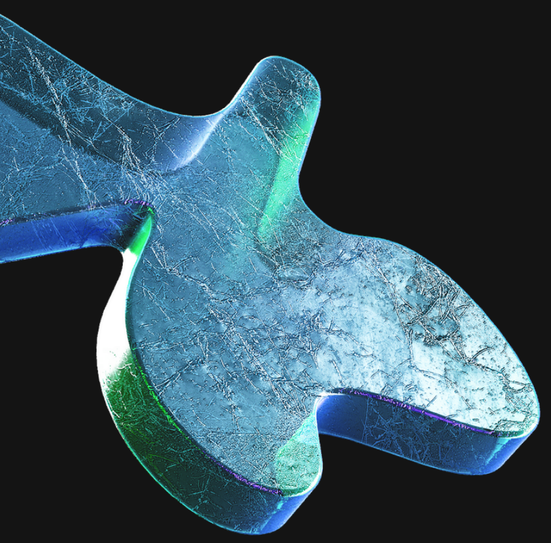


Rizki Ananda Putra Nur Rohmat

Flowchart Pengolahan Data



Data



About the Data...

SUMBER DATA

data historis rekam medis pasien IGD yang disediakan oleh PhysioNet, penyedia data riset medis yang diatur oleh Massachusetts Institute of Technology (MIT), Laboratory of Computational Physiology dan diambil dari penelitian Salman et al. (2021)

UKURAN DATA

Data yang dikumpulkan sebanyak 580 data dengan 13 fitur

FITUR DATA

ECG Records, Spo2, H. Blood(mHg), L. Blood(mHg), Chest Pain, Shortness of Breath, Palpitation., rest?, Peaks, QRS Width, Peak to Peak, ST El., dan OutputTriage level.

Dataset

	p. no.	ECG Records	Spo2	H. Blood(mHg)	L. Blood(mHg)	Chest Pain	Shortness of Breath	Palpitation.	rest?	Peaks	QRS width	Peak to Peak	ST El.	OutputTriage level
0	1	Sleep Apnea Records	97	23	12	False	False	False	False	67	0.06	Regular	YES	Sick
1	2	Sleep Apnea Records	97	23	12	False	False	False	True	67	0.06	Regular	YES	Sick
2	3	Sleep Apnea Records	97	23	12	False	False	True	False	67	0.06	Regular	YES	Sick
3	4	Sleep Apnea Records	97	23	12	False	False	True	True	67	0.06	Regular	YES	Sick
4	5	Sleep Apnea Records	97	23	12	False	True	False	False	67	0.06	Regular	YES	Sick
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
575	576	Normal ECG with HR 110	92	12	8	False	False	False	True	110	NaN	Regular	UNKNOWN	Sick
576	577	Normal ECG with HR 110	92	12	8	False	False	False	True	110	NaN	Regular	UNKNOWN	Sick
577	578	Normal ECG with HR 110	92	12	8	False	False	False	True	110	NaN	Regular	UNKNOWN	Sick
578	579	Normal ECG with HR 110	92	12	8	False	False	False	True	110	NaN	Regular	UNKNOWN	Sick
579	580	Normal ECG with HR 110	92	12	8	False	False	False	True	110	NaN	Regular	UNKNOWN	Sick

580 rows × 14 columns

Descriptive Statistics: Numerical Data

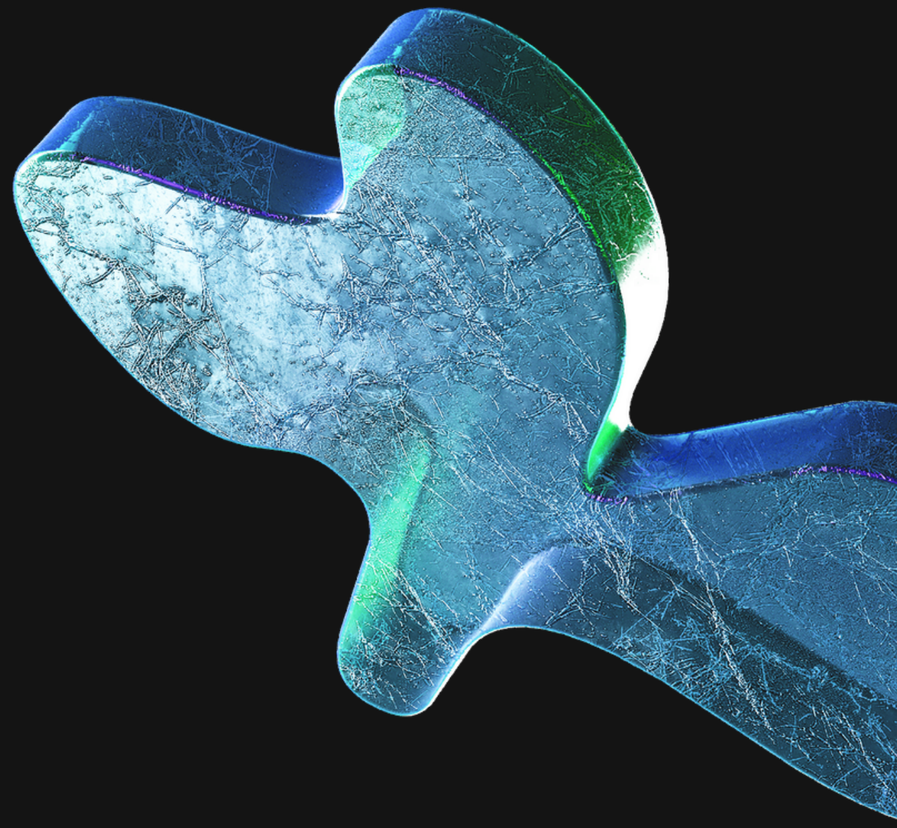
	count	mean	std	min	25%	50%	75%	max
Spo2	580.0	91.193103	6.785432	80.000	92.00	92.000	97.0000	97.0
H. Blood(mHg)	580.0	16.386207	4.611596	12.000	12.00	15.000	23.0000	23.0
L. Blood(mHg)	580.0	9.903448	1.634364	8.000	8.00	10.000	12.0000	12.0
QRS width	580.0	0.197800	0.182178	0.047	0.06	0.169	0.1978	0.5

Descriptive Statistics: Categorical Data

	count	unique	top	freq
ECG Records	580	6	Arrythmia Records	288
Chest Pain	580	2	False	296
Shortness of Breath	580	2	False	300
Palpitation.	580	2	False	304
rest?	580	2	True	304
Peaks	580	6	67	144
Peak to Peak	580	1	Regular	580
ST El.	580	3	YES	288
OutputTriage level	580	5	Sick	291

Model Development

Data Preprocessing



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 580 entries, 0 to 579
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   p. no.                580 non-null    int64
 1   ECG Records           580 non-null    object
 2   Spo2                  580 non-null    int64
 3   H. Blood(mHg)         580 non-null    int64
 4   L. Blood(mHg)         580 non-null    int64
 5   Chest Pain            580 non-null    object
 6   Shortness of Breath   580 non-null    object
 7   Palpitation.          580 non-null    object
 8   rest?                 580 non-null    object
 9   Peaks                 580 non-null    object
 10  QRS width             500 non-null    float64
 11  Peak to Peak          580 non-null    object
 12  ST El.                580 non-null    object
 13  OutputTriage level    580 non-null    object
dtypes: float64(1), int64(4), object(9)
memory usage: 63.6+ KB
```

DETECTING MISSING VALUES

- TERDAPAT 580 ENTRI DATA
- TERDAPAT 13 FITUR/PREDIKTOR
- TIPE DATA: NUMERICAL (INT64 & FLOAT64) DAN CATEGORICAL (OBJECT)
- TERDAPAT DATA NULL (NAN) PADA QRS WIDTH

how to handle?

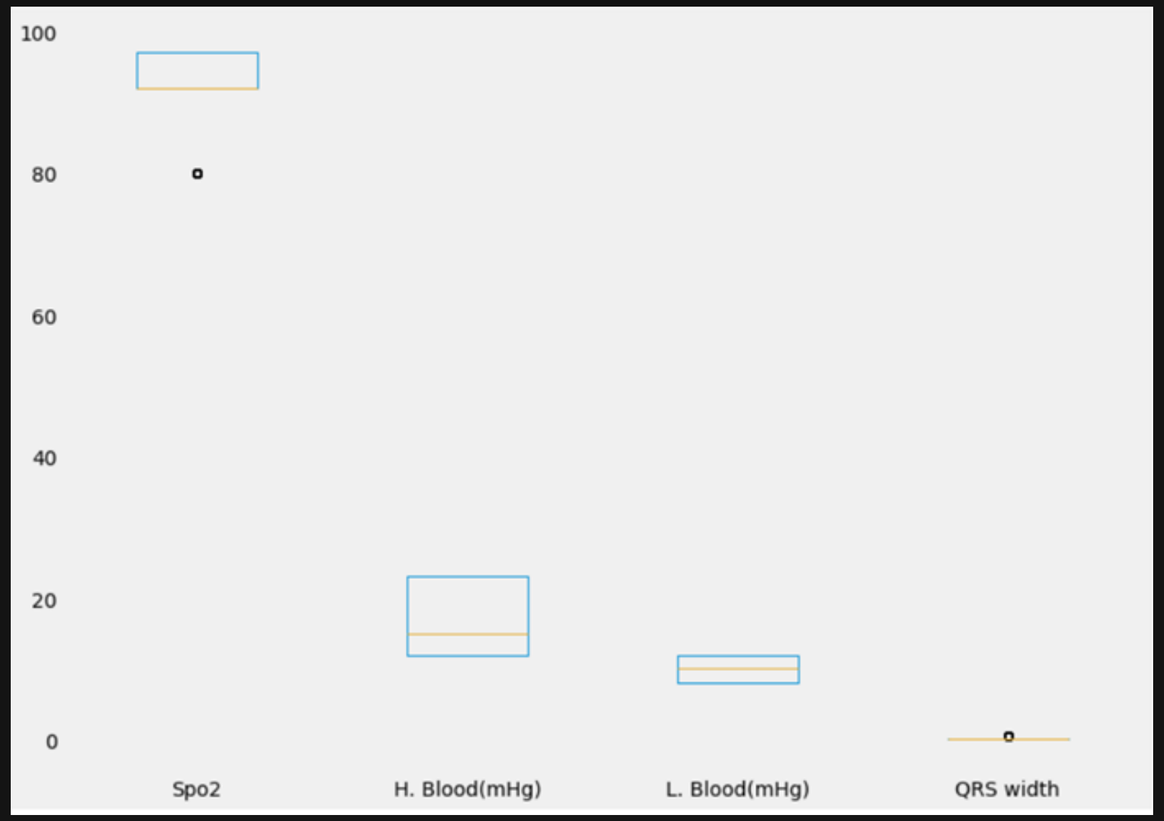
fillna with mean value of QRS Width

Before Handling Missing Value

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 580 entries, 0 to 579
Data columns (total 14 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   p. no.                580 non-null    int64
 1   ECG Records           580 non-null    object
 2   Spo2                  580 non-null    int64
 3   H. Blood(mHg)         580 non-null    int64
 4   L. Blood(mHg)         580 non-null    int64
 5   Chest Pain            580 non-null    object
 6   Shortness of Breath   580 non-null    object
 7   Palpitation.          580 non-null    object
 8   rest?                 580 non-null    object
 9   Peaks                 580 non-null    object
 10  QRS width             500 non-null    float64
 11  Peak to Peak          580 non-null    object
 12  ST El.                580 non-null    object
 13  OutputTriage level    580 non-null    object
dtypes: float64(1), int64(4), object(9)
memory usage: 63.6+ KB
```

After Handling Missing Value

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 580 entries, 0 to 579
Data columns (total 13 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ECG Records           580 non-null    object
 1   Spo2                  580 non-null    int64
 2   H. Blood(mHg)         580 non-null    int64
 3   L. Blood(mHg)         580 non-null    int64
 4   Chest Pain            580 non-null    object
 5   Shortness of Breath   580 non-null    object
 6   Palpitation.          580 non-null    object
 7   rest?                 580 non-null    object
 8   Peaks                 580 non-null    object
 9   QRS width             580 non-null    float64
 10  Peak to Peak          580 non-null    object
 11  ST El.                580 non-null    object
 12  OutputTriage level    580 non-null    object
dtypes: float64(1), int64(3), object(9)
memory usage: 59.0+ KB
```



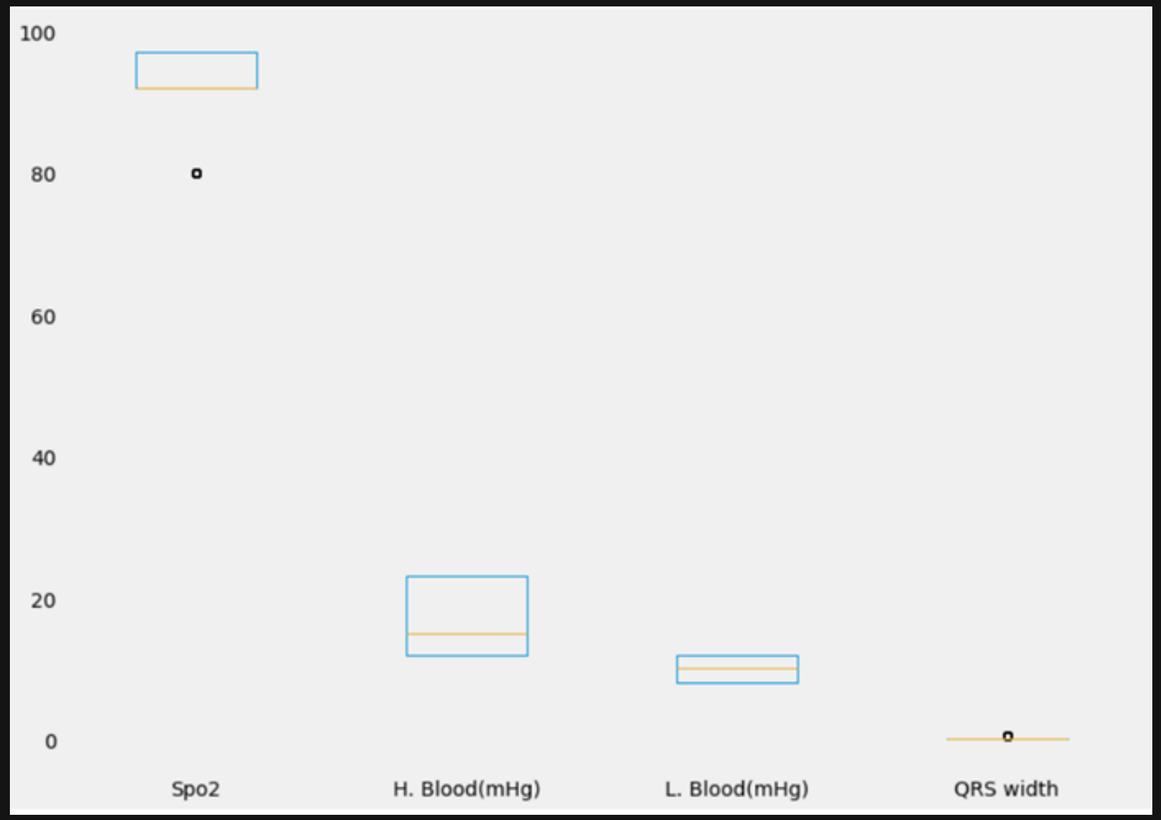
DETECTING OUTLIERS

- Numerical Data: Spo2, H. Blood(mHg), L. Blood(mHg), QRS Width
- Terdapat outlier pada Spo2 dan QRS Width (ditandai dengan titik hitam)

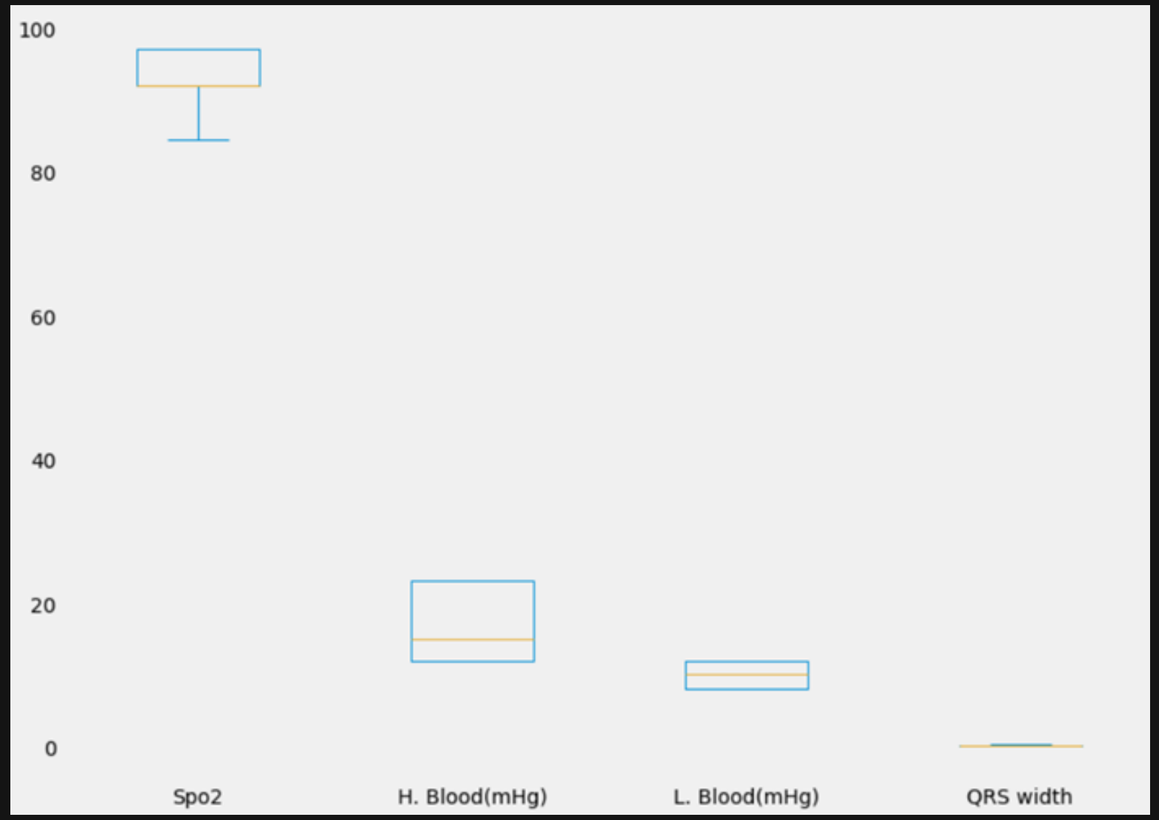
how to handle?

masking with IQR method

Before Handling Outliers

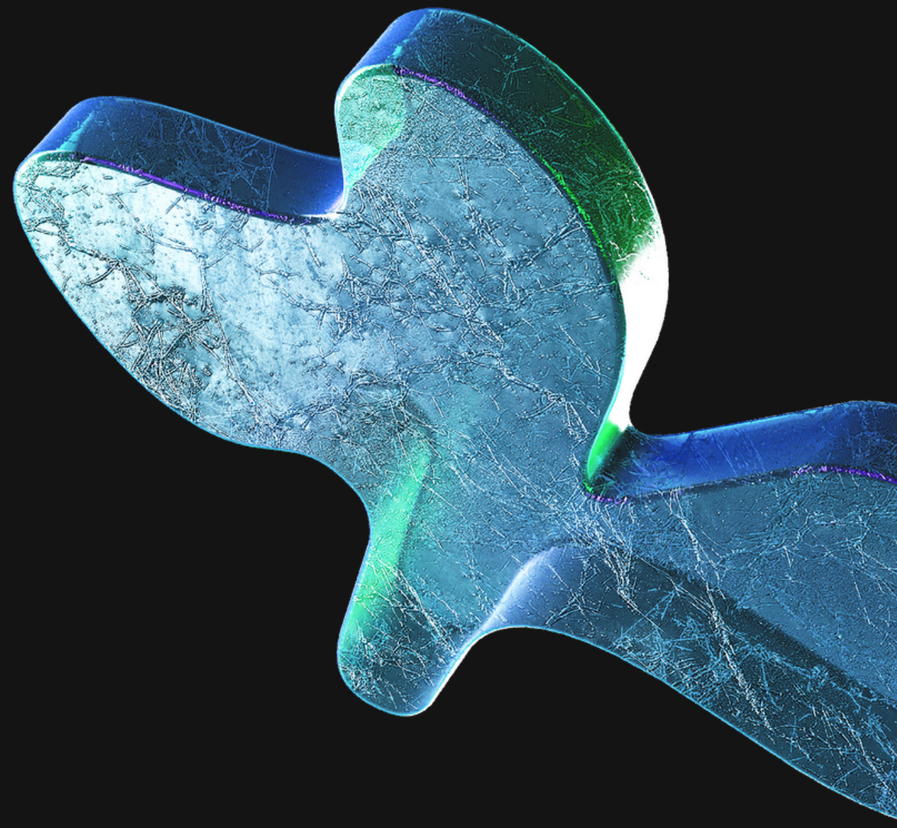


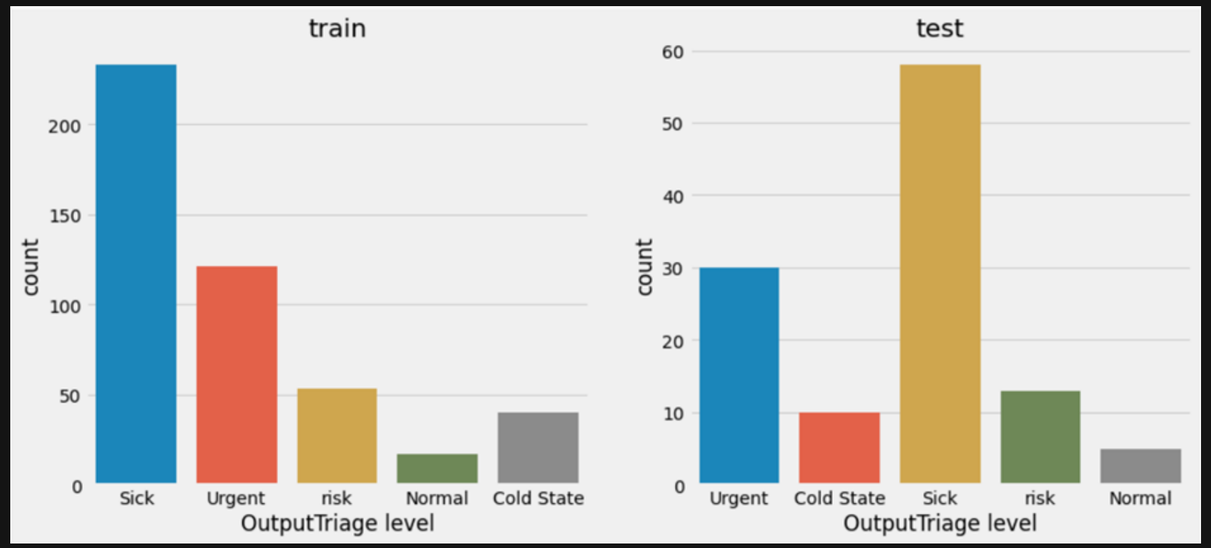
After Handling Outliers



Model Development

Splitting Dataset





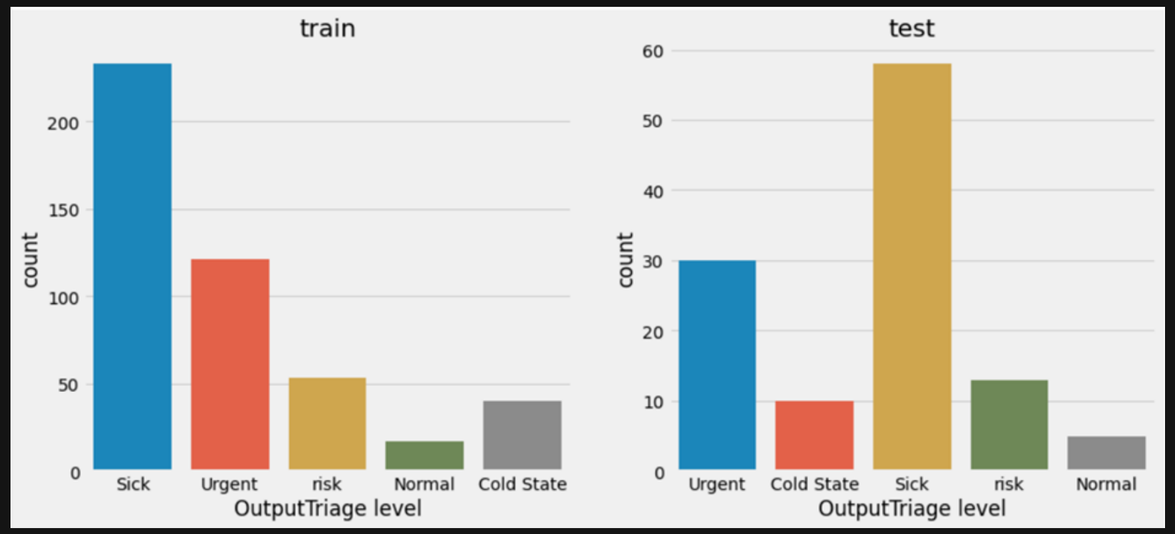
SPLITTING DATASET

- Dibagi menjadi train dataset dan test dataset (80:20)
- Terdapat imbalanced data pada train dataset
- Kelas terbanyak: Sick (233 data)
- Kelas paling sedikit: Normal (17 data)

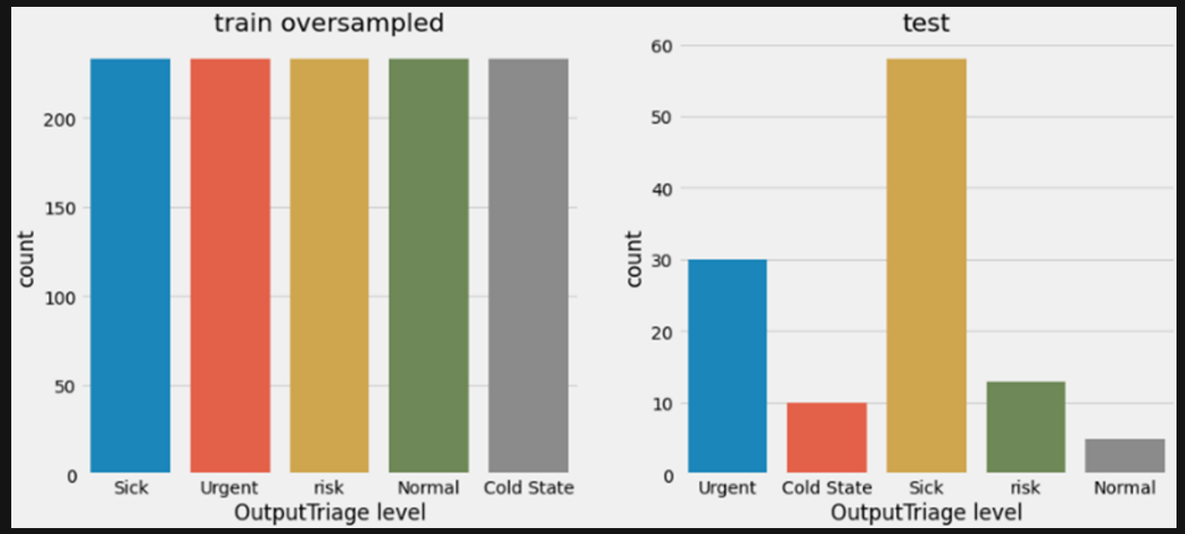
how to handle?

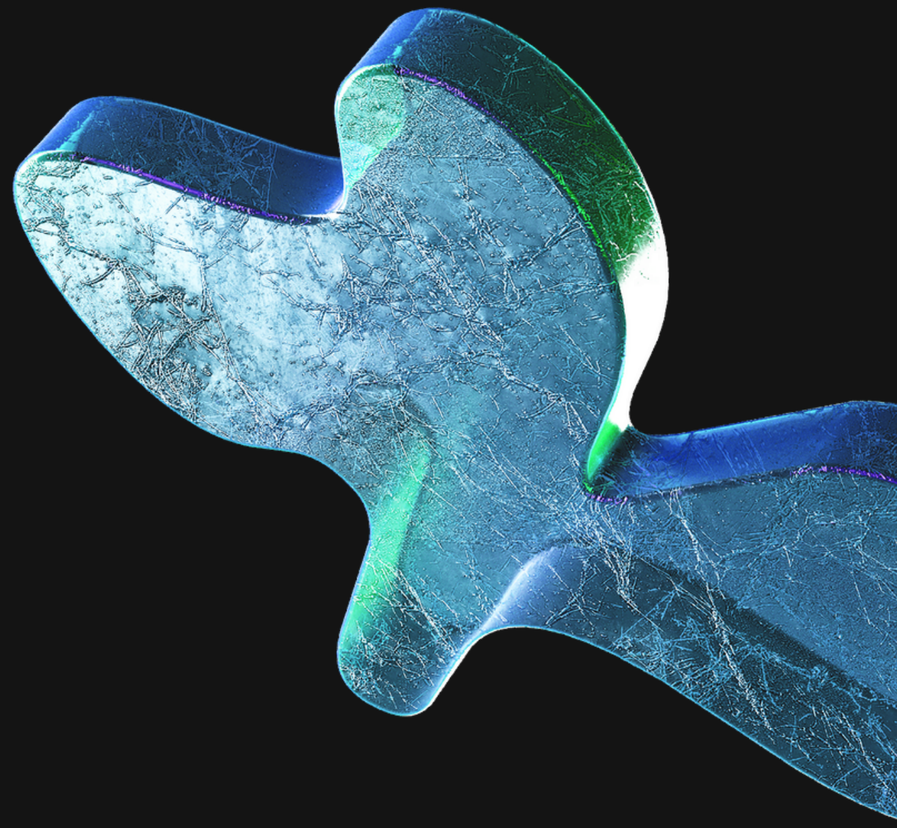
Handling Imbalanced Data with Random Over Sampling

Before Handling Imbalanced Data



After Handling Imbalanced Data





Model Development

Data Encoding and Rescaling

Data Encoding?

TUJUAN

Mengubah data kategorikal menjadi data numerikal yang dapat diproses oleh library machine learning

JENIS DATA

Data nominal: Data kategorikal berisi True/False
Data ordinal: Data kategorikal berisi beberapa kelas

TOOLS

Data Nominal: OneHotEncoder
Data Ordinal: OrdinalEncoder

After Data Encoding: Nominal Data

	Chest Pain_False	Chest Pain_True	Shortness of Breath_False	Shortness of Breath_True	Palpitation._False	Palpitation._True	rest?_False	rest?_True
0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0
1	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0
2	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0
3	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0
4	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0
...	...	...	...	...	...	...	...	...
1160	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0
1161	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0
1162	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0
1163	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0
1164	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0

1165 rows × 8 columns

After Data Encoding: Ordinal Data

	ECG Records	Peaks	Peak to Peak	ST El.
0	0.0	1.0	0.0	0.0
1	2.0	3.0	0.0	2.0
2	0.0	4.0	0.0	2.0
3	5.0	3.0	0.0	2.0
4	0.0	1.0	0.0	0.0
...	...	...	...	...
1160	0.0	1.0	0.0	0.0
1161	0.0	4.0	0.0	2.0
1162	0.0	4.0	0.0	2.0
1163	5.0	3.0	0.0	2.0
1164	0.0	1.0	0.0	0.0

1165 rows × 4 columns

Data Rescaling?

TUJUAN

Mengubah data agar memiliki nilai rata rata 0 dan nilai standar deviasi 1

TOOLS

Menggunakan StandardScaler

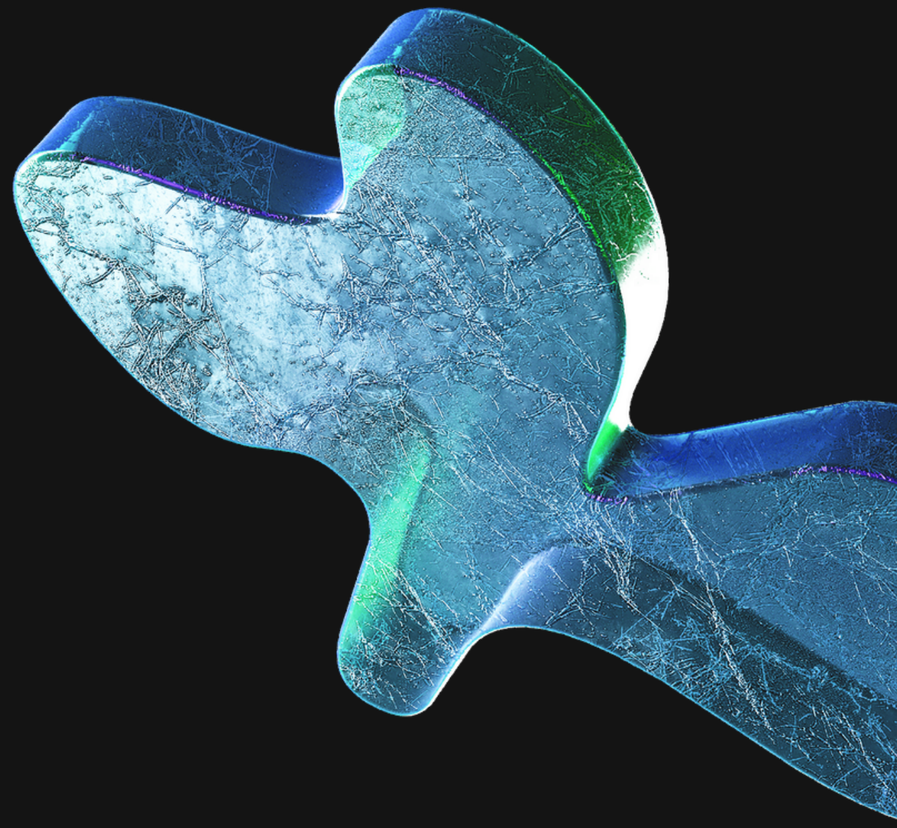
After Data Rescaling:

	Spo2	H. Blood(mHg)	QRS width	Chest Pain_False	Chest Pain_True	Shortness of Breath_False	Shortness of Breath_True	Palpitation._False	Palpitation._True	rest?_False	rest?_True	ECG Records	Peaks	Peak to Peak	ST El.
0	97.0	15	0.4045	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0
1	84.5	23	0.0600	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	2.0	3.0	0.0	2.0
2	92.0	15	0.0470	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	4.0	0.0	2.0
3	97.0	23	0.0600	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	5.0	3.0	0.0	2.0
4	84.5	23	0.4045	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1160	84.5	23	0.4045	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0
1161	92.0	15	0.0470	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	4.0	0.0	2.0
1162	92.0	23	0.0470	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	4.0	0.0	2.0
1163	92.0	23	0.0600	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	5.0	3.0	0.0	2.0
1164	92.0	23	0.4045	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0

1165 rows × 15 columns

Model Development

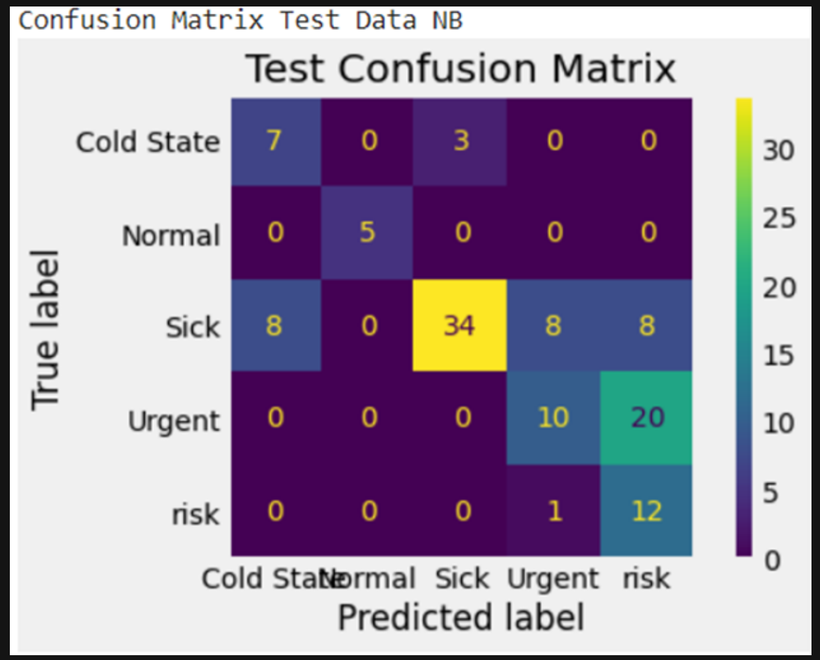
Model Evaluation and Comparison



Classification Report Test Data NB

	precision	recall	f1-score	support
Cold State	0.47	0.70	0.56	10
Normal	1.00	1.00	1.00	5
Sick	0.92	0.59	0.72	58
Urgent	0.53	0.33	0.41	30
risk	0.30	0.92	0.45	13
accuracy			0.59	116
macro avg	0.64	0.71	0.63	116
weighted avg	0.71	0.59	0.61	116

Confusion Matrix Test Data NB



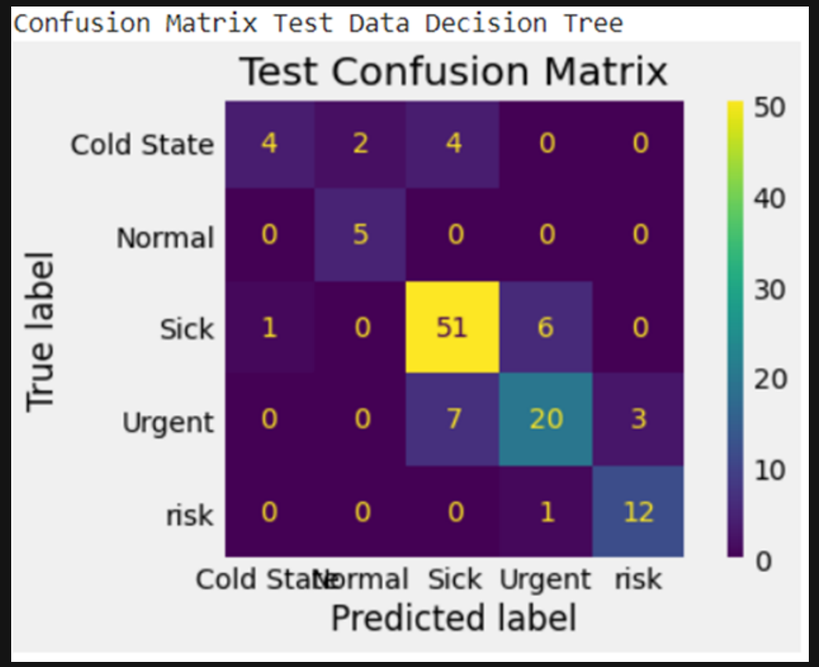
NB MODEL EVALUATION

- Accuracy: 0,59
- Precision: 0,39; 1,00; 0,92; 0,53; dan 0,30
- Recall: 0,70; 1,00; 0,62; 0,30; dan 0,85
- f-1 score: 0,50; 1,00; 0,74; 0,38; dan 0,44
- Confusion matrix: 48 data yang salah diprediksi dan masuk ke dalam kelas lain

Classification Report Test Data Decision Tree

	precision	recall	f1-score	support
Cold State	0.80	0.40	0.53	10
Normal	0.71	1.00	0.83	5
Sick	0.82	0.88	0.85	58
Urgent	0.74	0.67	0.70	30
risk	0.80	0.92	0.86	13
accuracy			0.79	116
macro avg	0.78	0.77	0.76	116
weighted avg	0.79	0.79	0.78	116

Confusion Matrix Test Data Decision Tree



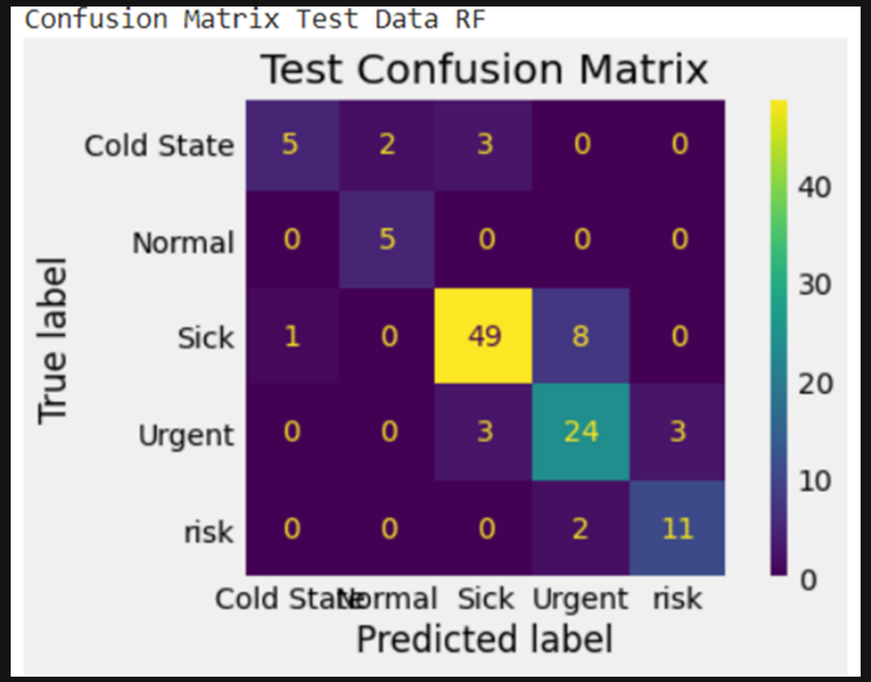
DT MODEL EVALUATION

- Accuracy: 0,79
- Precision: 0,80; 0,71; 0,82; 0,74; dan 0,80
- Recall: 0,40; 1,00; 0,88; 0,67; dan 0,92
- f-1 score: 0,53; 0,83; 0,85; 0,70; dan 0,86
- Confusion matrix: 24 data yang salah diprediksi dan masuk ke dalam kelas lain

Classification Report Test Data RF

	precision	recall	f1-score	support
Cold State	0.83	0.50	0.62	10
Normal	0.71	1.00	0.83	5
Sick	0.89	0.84	0.87	58
Urgent	0.71	0.80	0.75	30
risk	0.79	0.85	0.81	13
accuracy			0.81	116
macro avg	0.79	0.80	0.78	116
weighted avg	0.82	0.81	0.81	116

Confusion Matrix Test Data RF



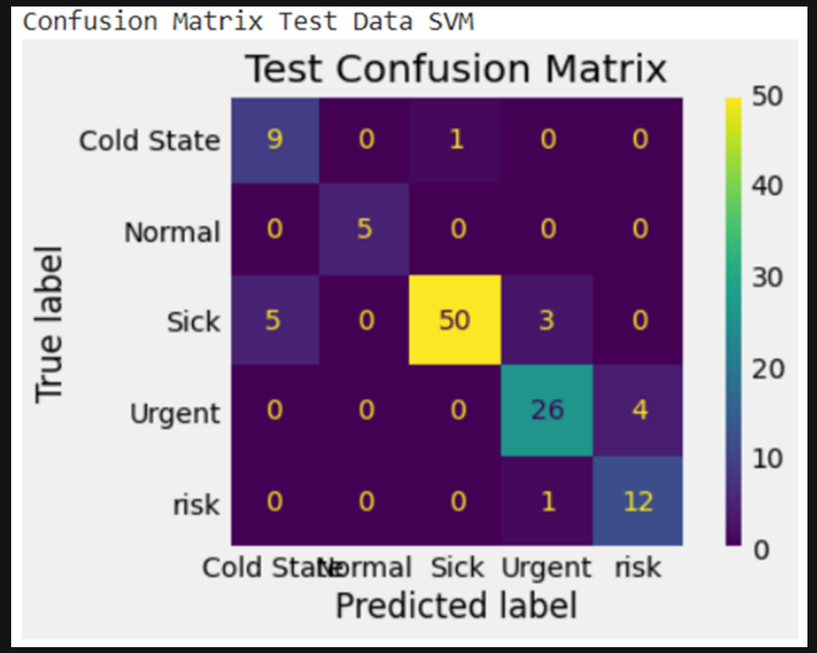
RF MODEL EVALUATION

- Accuracy: 0,81
- Precision: 0,83; 0,71; 0,89; 0,71; dan 0,79
- Recall: 0,50; 1,00; 0,84; 0,80; dan 0,85
- f-1 score: 0,62; 0,83; 0,87; 0,75; dan 0,81
- Confusion matrix: 25 data yang salah diprediksi dan masuk ke dalam kelas lain

Classification Report Test Data SVM

	precision	recall	f1-score	support
Cold State	0.64	0.90	0.75	10
Normal	1.00	1.00	1.00	5
Sick	0.98	0.86	0.92	58
Urgent	0.87	0.87	0.87	30
risk	0.75	0.92	0.83	13
accuracy			0.88	116
macro avg	0.85	0.91	0.87	116
weighted avg	0.90	0.88	0.88	116

Confusion Matrix Test Data SVM



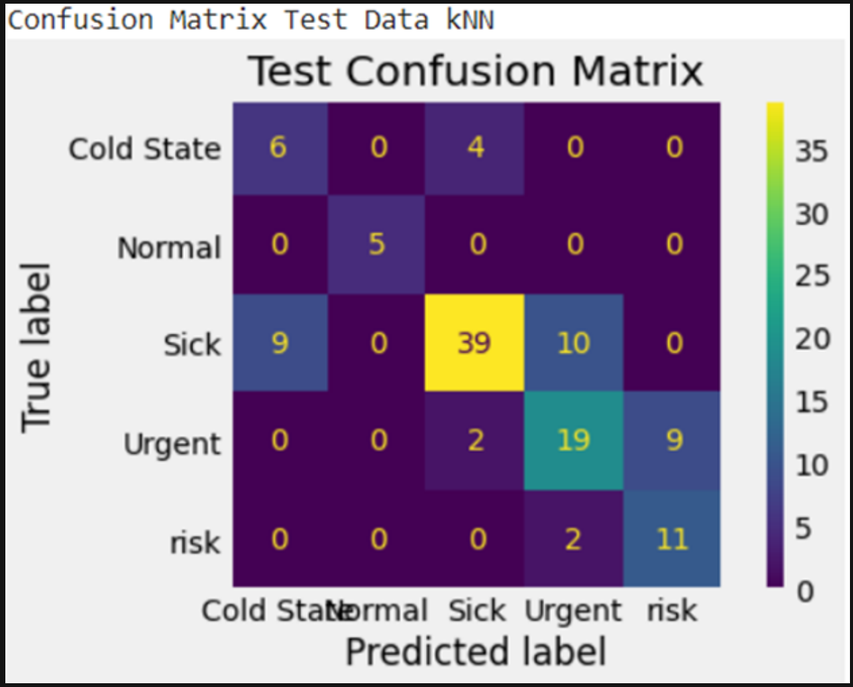
SVM MODEL EVALUATION

- Accuracy: 0,88
- Precision: 0,64; 1,00; 0,98; 0,87; dan 0,75
- Recall: 0,90; 1,00; 0,86; 0,87; dan 0,92
- f-1 score: 0,75; 1,00; 0,92; 0,87; dan 0,83
- Confusion matrix: 11 data yang salah diprediksi dan masuk ke dalam kelas lain

Classification Report Test Data kNN

	precision	recall	f1-score	support
Cold State	0.40	0.60	0.48	10
Normal	1.00	1.00	1.00	5
Sick	0.87	0.67	0.76	58
Urgent	0.61	0.63	0.62	30
risk	0.55	0.85	0.67	13
accuracy			0.69	116
macro avg	0.69	0.75	0.71	116
weighted avg	0.73	0.69	0.70	116

Confusion Matrix Test Data kNN



K-NN MODEL EVALUATION

- Accuracy: 0,69
- Precision: 0,40; 1,00; 0,87; 0,61; dan 0,55
- Recall: 0,60; 1,00; 0,67; 0,63; dan 0,85
- f-1 score: 0,48; 1,00; 0,76; 0,62; dan 0,67
- Confusion matrix: 36 data yang salah diprediksi dan masuk ke dalam kelas lain

Accuracy Score Recap

	Decision Tree	NB	SVM	K - NN	RF
Accuracy Score	0.79	0.59	0.88	0.69	0.81

Model SVM memiliki nilai accuracy tertinggi

Precision Score Recap

	Decision Tree	NB	SVM	K – NN	RF
Cold State	0.8	0.39	0.64	0.4	0.83
Normal	0.71	1	1	1	0.71
Sick	0.82	0.92	0.98	0.87	0.89
Urgent	0.74	0.53	0.87	0.61	0.71
Risk	0.8	0.3	0.75	0.55	0.79
Rata - Rata	0.774	0.628	0.848	0.686	0.786

Model SVM memiliki rata - rata nilai precision tertinggi

Recall Score Recap

	Decision Tree	NB	SVM	K – NN	RF
Cold State	0.4	0.7	0.9	0.6	0.5
Normal	1	1	1	1	1
Sick	0.88	0.62	0.86	0.67	0.84
Urgent	0.67	0.3	0.87	0.63	0.8
Risk	0.92	0.85	0.92	0.85	0.85
Rata - Rata	0.774	0.694	0.91	0.75	0.798

Model SVM memiliki rata - rata nilai recall tertinggi

f-1 Score Recap

	Decision Tree	NB	SVM	K - NN	RF
Cold State	0.53	0.5	0.75	0.48	0.62
Normal	0.83	1	1	1	0.83
Sick	0.85	0.74	0.92	0.76	0.87
Urgent	0.7	0.38	0.87	0.62	0.75
Risk	0.86	0.44	0.92	0.67	0.81
Rata - Rata	0.754	0.612	0.892	0.706	0.776

Model SVM memiliki rata - rata nilai f-1 score tertinggi

Misspreadiction Recap

	Decision Tree	NB	SVM	K - NN	RF
Kesalahan Prediksi	24	48	11	36	25

Model SVM memiliki jumlah misprediksi terendah

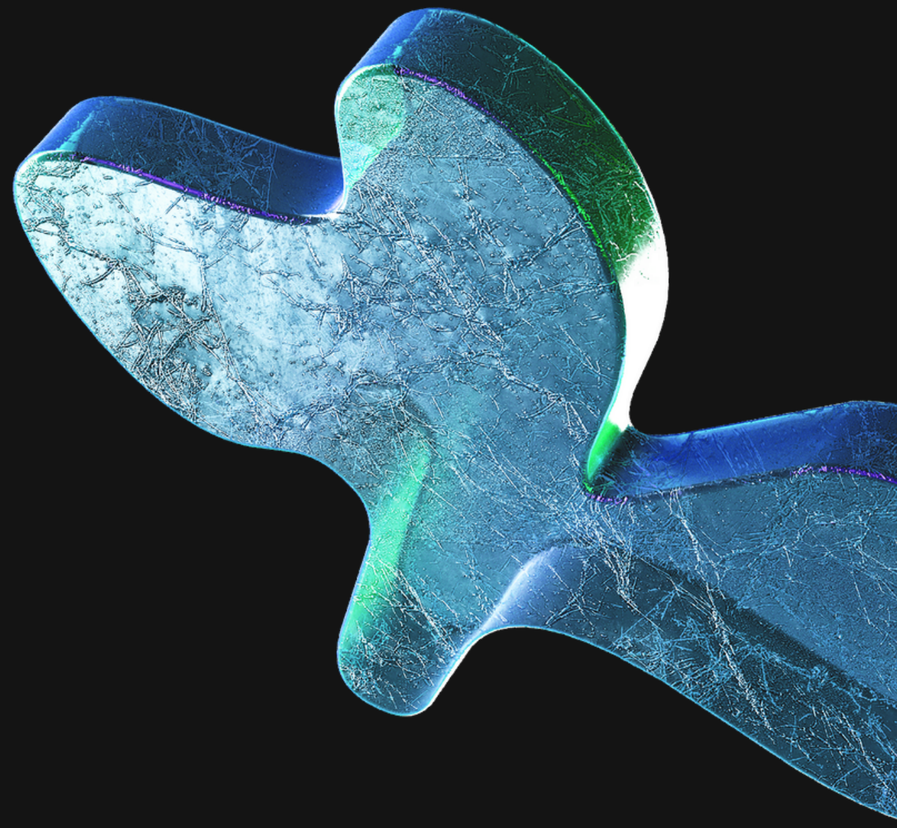
Kesimpulan

Model SVM adalah model terbaik guna memprediksi triase pasien IGD

Hal ini dikarenakan karena model SVM memiliki nilai accuracy, precision, recall, dan f-1 score tertinggi serta jumlah misprediksi terendah dibandingkan dengan model lainnya

Model Development

Hyperparameter Tuning



Parameter yang Akan Dilakukan Hyperparameter Tuning

Parameter	**Definisi**	**Nilai Parameter yang Akan Dilakukan Proses *Tuning***
Kernel	Jenis *kernel* yang digunakan dalam model (linear, *poly*, rbf, sigmoid, *precomputed*, dll)	*Poly*
Degree	Derajat fungsi *kernel* polinomial	3,4,5,6,7,8,9,10
Gamma	Koefisien *kernel* untuk rbf, *poly*, dan *sigmoid*	*Scale, auto*
Coef0	Independensi fungsi *kernel*	0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9
Decision_function_shape	Bentuk pemilihan keputusan (ovo atau ovr)	*ovo* dan *ovr*
Probability	Apakah estimasi probabilitas diperlukan? (*True* atau *False*)	*True* dan *false*

HYPERPARAMETER TUNING

- Menggunakan library GridSearchCV
- Jumlah kombinasi: 640 kombinasi
- Pemilihan set hyperparameter terbaik berdasarkan nilai accuracy tertinggi

	coef0	decision_function_shape	degree	gamma	kernel	probability	Accuracy
0	0.0	ovo	3	scale	poly	True	0.931330
1	0.0	ovo	3	scale	poly	False	0.931330
2	0.0	ovo	3	auto	poly	True	0.929614
3	0.0	ovo	3	auto	poly	False	0.929614
4	0.0	ovo	4	scale	poly	True	0.901288
...	...	...	...	...	...	...	...
635	0.9	ovr	9	auto	poly	False	0.954506
636	0.9	ovr	10	scale	poly	True	0.945923
637	0.9	ovr	10	scale	poly	False	0.945923
638	0.9	ovr	10	auto	poly	True	0.945923
639	0.9	ovr	10	auto	poly	False	0.945923

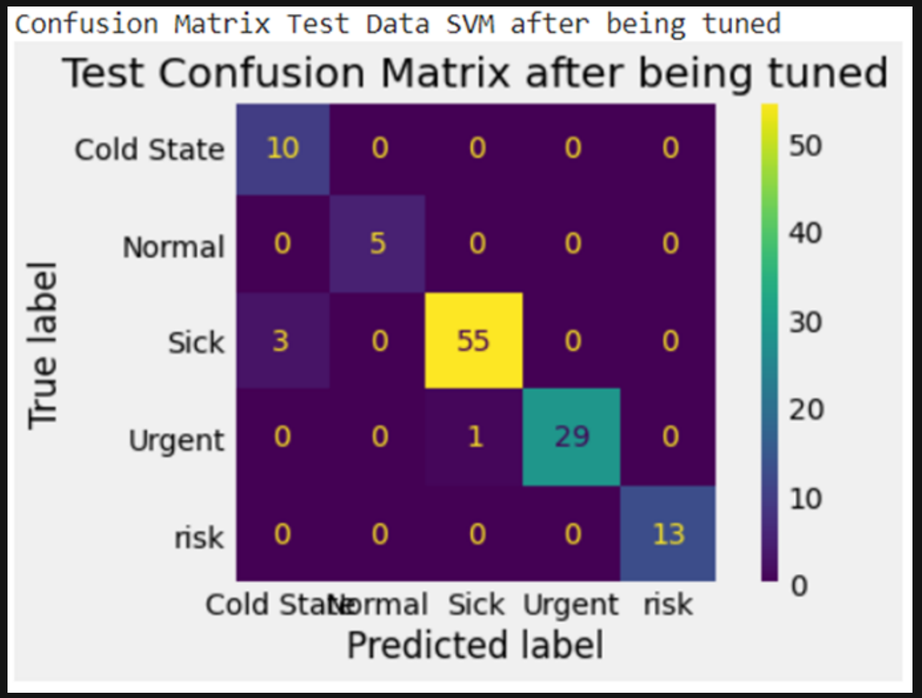
640 rows × 7 columns

- coef0: 0.9
- decision_function_shape: ovo
- degree: 3
- gamma: auto
- kernel: poly
- probability: True
- accuracy: 0.98

Classification Report Test Data SVM after being tuned

	precision	recall	f1-score	support
Cold State	0.77	1.00	0.87	10
Normal	1.00	1.00	1.00	5
Sick	0.98	0.95	0.96	58
Urgent	1.00	0.97	0.98	30
risk	1.00	1.00	1.00	13
accuracy			0.97	116
macro avg	0.95	0.98	0.96	116
weighted avg	0.97	0.97	0.97	116

Confusion Matrix Test Data SVM after being tuned



MODEL EVALUATION

- Accuracy: 0,97
- Precision: 0,77; 1,00; 0,98; 1,00; dan 1,00
- Recall: 1,00; 1,00; 0,95; 0,97; dan 1,00
- f-1 score: 00,87; 1,00; 0,96; 0,98; dan 1,00
- Confusion matrix: 4 data yang salah diprediksi dan masuk ke dalam kelas lain