Understanding COVID-19 Death Risk Factors:

A Case Study at Pratama Hospital Yogyakarta

By

Biliarto Sastro Cemerson

Data Science Project

Overview

Introduction

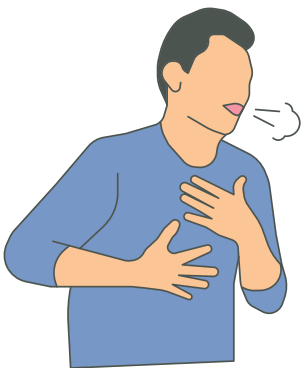**Methodology**
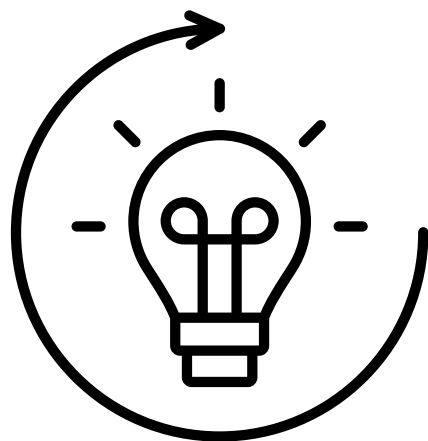
Result
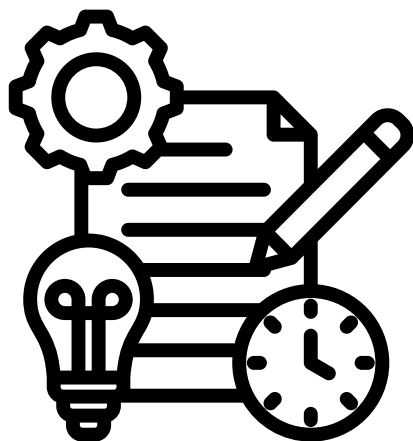
Introduction

Methodology

Result

Introduction

Although global **COVID-19 mortality rates** have declined due to **widespread vaccination** and **improved treatments**, the virus remains a **serious threat**, especially for patients with **pre-existing conditions**.
(Moss et al., 2022).

Anisa & Rifai (2022) emphasized that **shortness of breath**, **sex**, **age**, and **comorbidities** significantly affect the **likelihood of death** in COVID-19 patients.

Studies Ilpaj & Nurwati, (2020) and Shobri et al., (2021) identified **comorbidities** and **age** as **significant factors** influencing COVID-19 mortality.
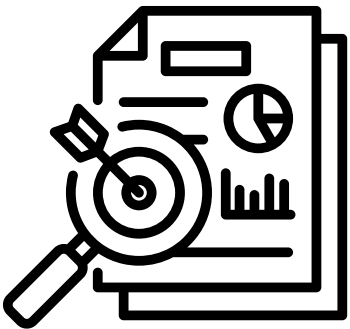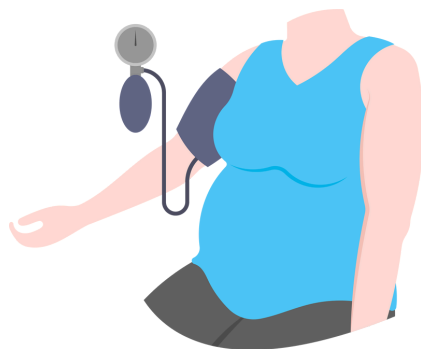
Therefore, this study aims to **identify key mortality risk factors** among COVID-19 patients at **Pratama Hospital Yogyakarta,** based on patient demographics and clinical characteristics.

Additional research (Nugraha et al., 2021; Mariyam et al., 2022) highlights the impact of **demographic** and clinical characteristics—such as **age**, **sex**, and **specific comorbidities**—on patient outcomes.
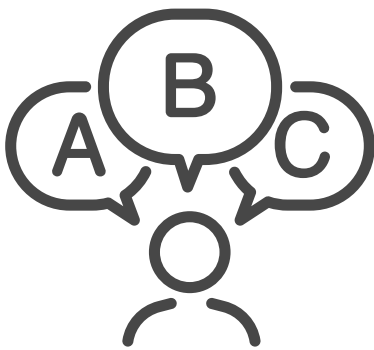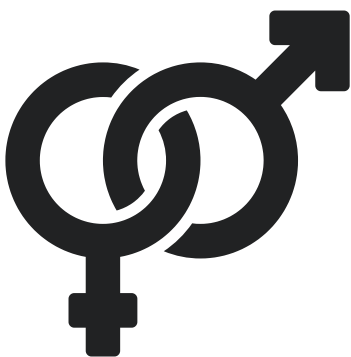
The variables used in this analysis include: **sex**, **age**, **COVID-19 status**, **treatment type**, **clinical symptoms** (shortness of breath, eating disorders), **comorbidities** (diabetes, hypertension), and **patient outcomes** (recovered vs deceased).

Introduction   Methodology   Result

Methodology

Analysis Stages

Tools and Libraries

Data Exploration

Data Preprocessing

Model Evaluation
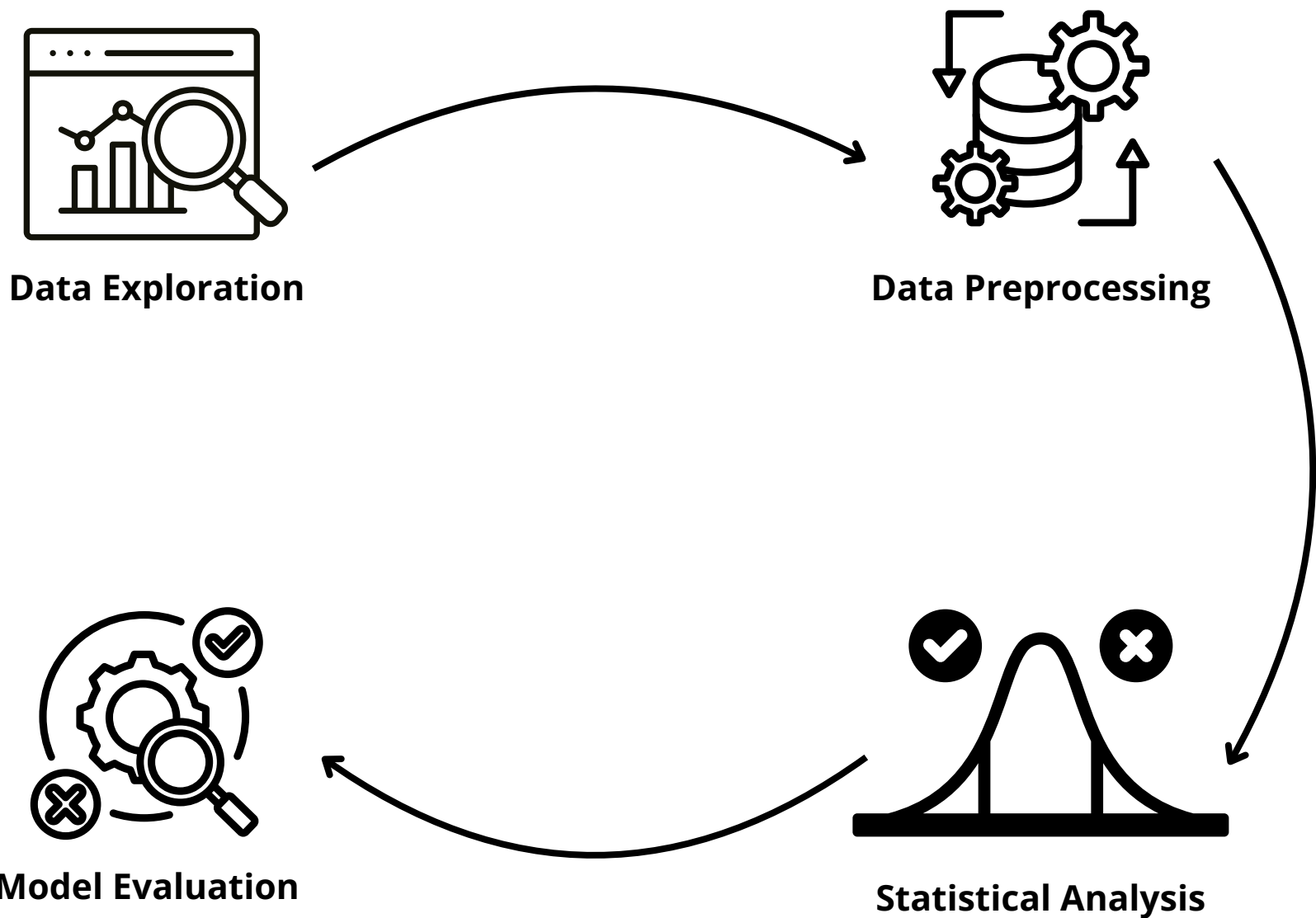
Statistical Analysis

tidyverse

ROSE

nnet

VGAM

caret

Detail Analysis

Data Preprocessing **(Spliting, and Handling Imbalance Class)**

Statistical Analysis (**Independent Test, Parameter Estimation, Simultaneous Test, Partial Test, Binary Logistic Regression Model, Odds Ratio Interpretation, Goodness-of-Fit Test.**

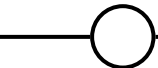Model Evaluation **(Accuracy, Precision, Recall, F1-Score)**

Result

Data

580
Patient

The data consists of **580** patients treated **from 2021 to 2023** at **Pratama hospital Yogyakarta.**

Distribution of Patient Sex

**Female
50.9%**

**Male
49.1%**

The distribution **by sex** appears to be **almost the same**.
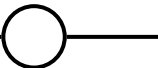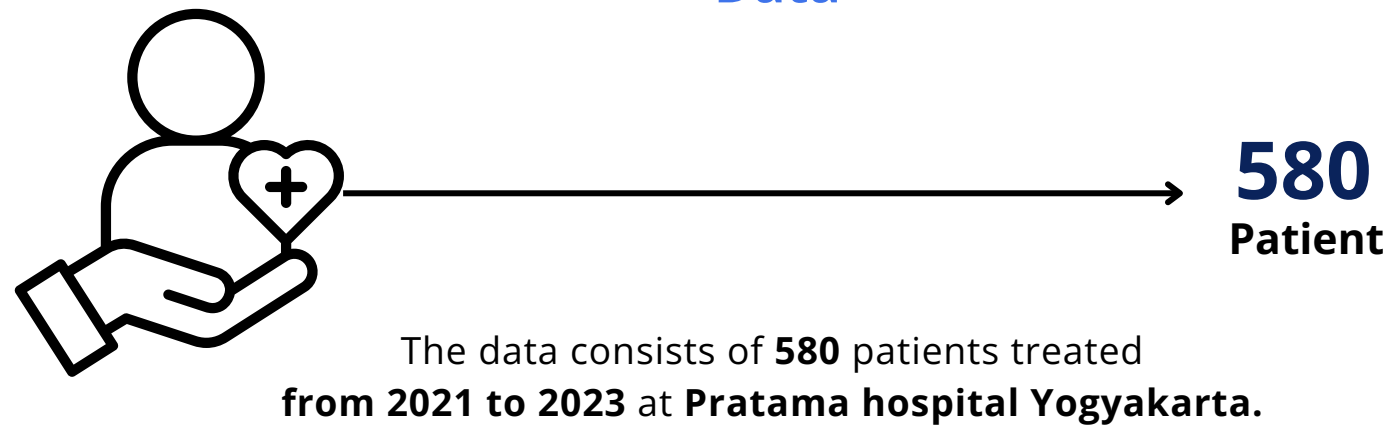
Distribution of Patient Outcome

**Deceased
34.3%**

28

552

Patients who have **recovered** are **more dominant** compared to those who have **deceased**.

**Recovered
65.7%**
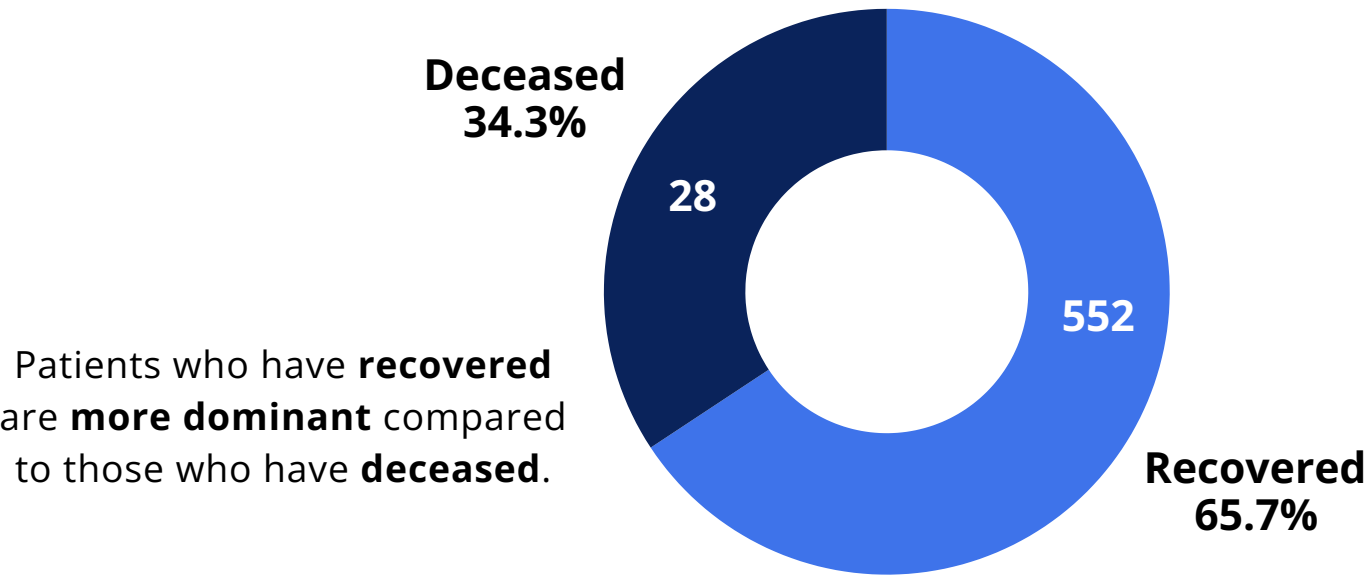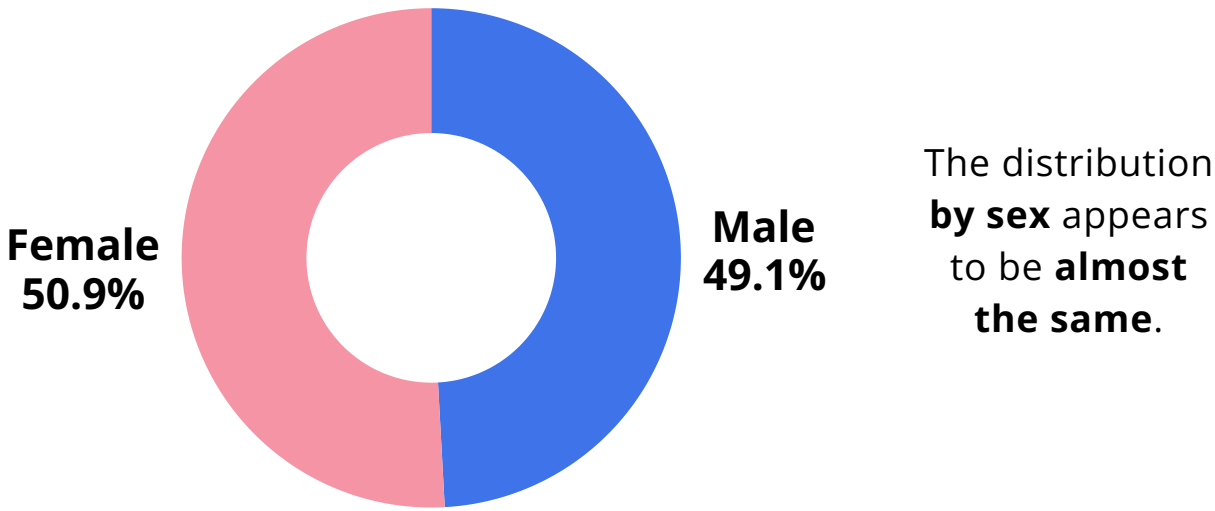
Distribution of COVID-19 Status
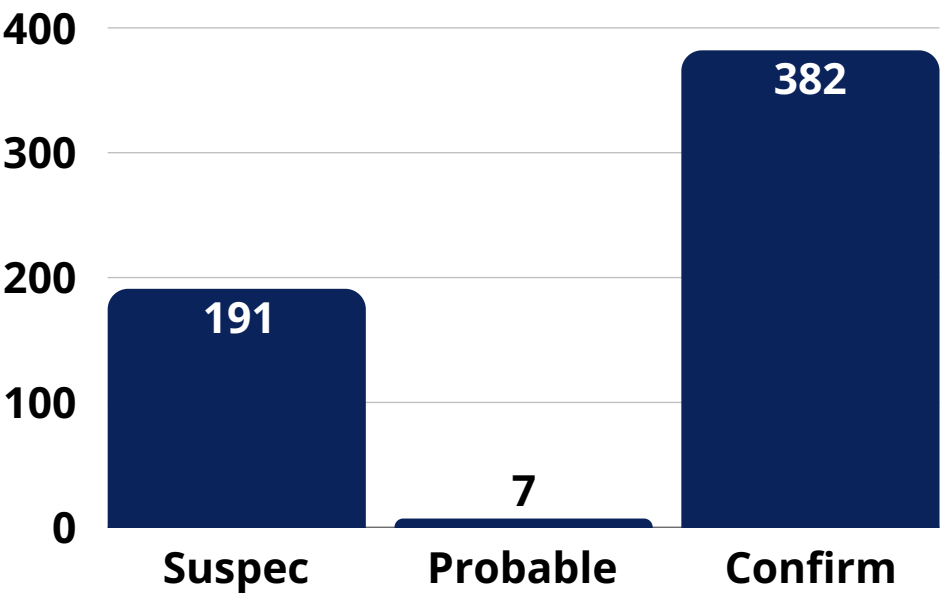
It appears that the **confirmed COVID-19 status dominates**, followed by the suspects.

400

300

200

191

100

7

382

0

Suspec   **Probable**   **Confirm**

Introduction   Methodology   Result

Result

Spliting Data (70:30)

Train data will be used for **statistical analysis** and **training the model**.

Test data to **evaluate** the **model's performance** in predictions.

Test 30%
174

Train 70%
406

● Recovered   ● Deceased

389
17
163
11
Train   Test

target column in train data is imbalance
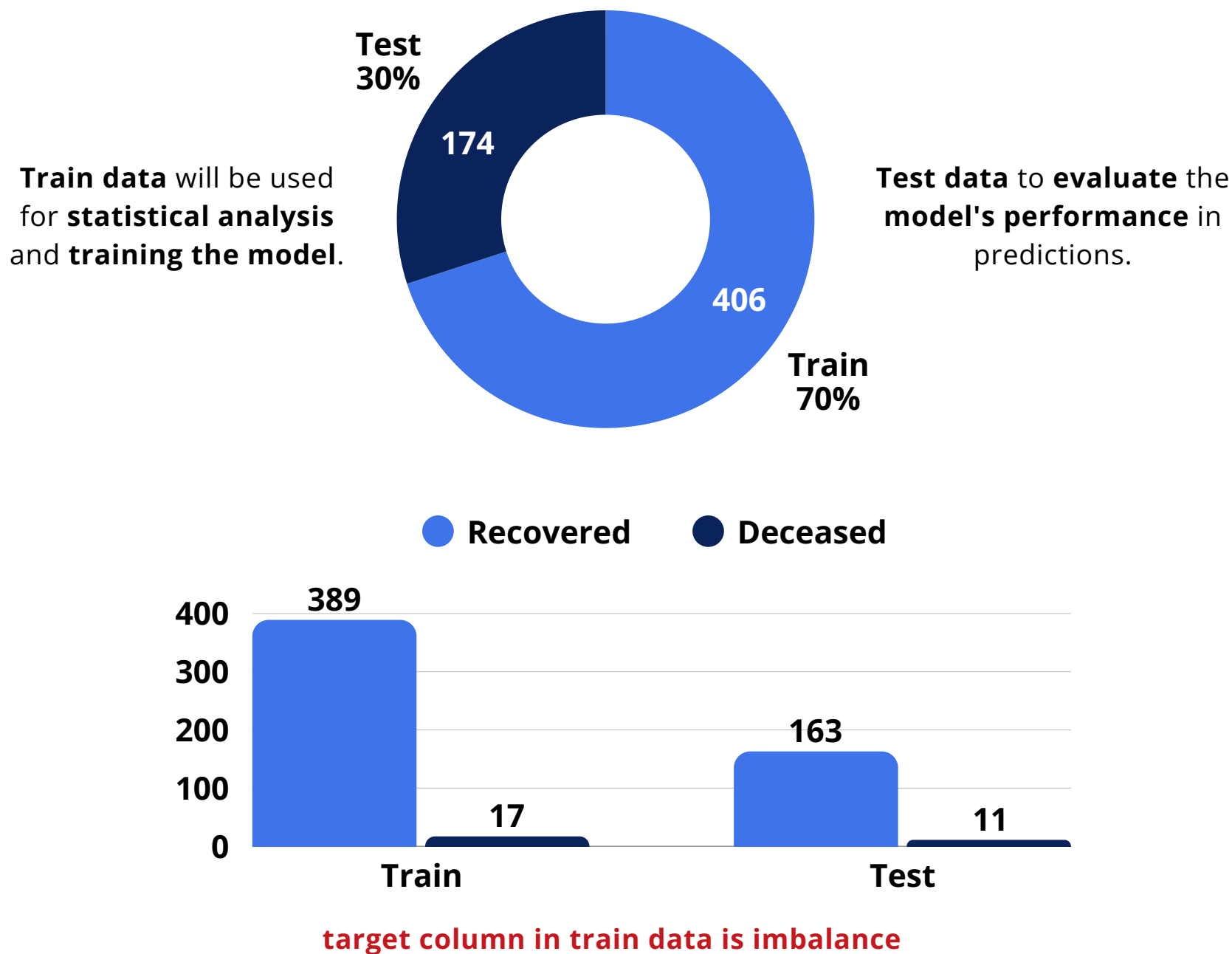
Handling Imbalance Class using ROSE

● Recovered   ● Deceased

250
200
150
100
50
0

200   206

Train

data is more balanced, and ready for analysis

Introduction — Methodology — **Result**
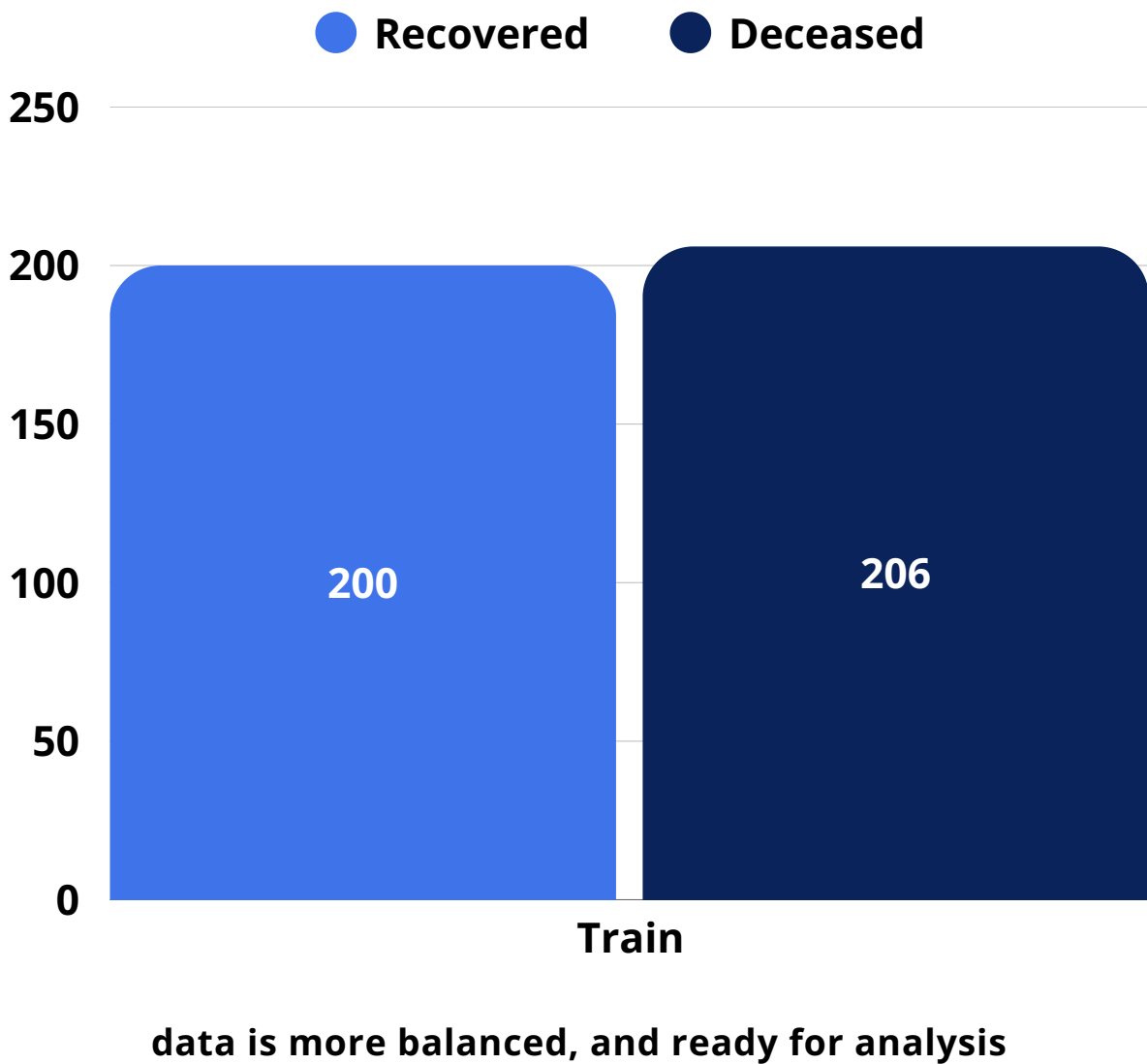
Result

Independet Test

Variable	Notation	p-value	Decision
Sex	X1	0.3874	Fail Reject H0
Age	X2	0.0000	Reject H0
COVID-19 Status	X3	0.0000	Reject H0
Treatment Type	X4	0.0000	Reject H0

Independent Test

Variable	Notation	p-value	Decision
Shortness of Breath	X5	0.0016	Reject H0
Eating Disorders	X6	0.0000	Reject H0
Diabetes	X7	0.0078	Reject H0
Hypertension	X8	0.0000	Reject H0

The **p-value** for **sex** was **0.3874 (> 0.05)**,
indicating **no significant association** with patient outcomes

Therefore, variable **sex was excluded** from further analysis.

Result

Simulatneous Test

Partial Test

	G	Chi-Square	Decision
	317.0345	12.5195	Reject H0

The **G value** from the simultaneous test was **317.03**,
which is **greater than** the critical value $\chi^2(6, 0.05)$ = **12.52**.

Therefore, **H$_0$ is rejected**,
indicating that **at least one variable significantly affects** the patient outcome.

Variable	Category	p-value	Decision
Age	> 50 Year	0.0000	Reject H0
COVID-19 Status	Confirm	0.0000	Reject H0
Treatment Type	Inpatient Care	0.0001	Reject H0
Shortness of Breath	Yes	0.0411	Reject H0
Eating Disorders	Yes	0.0000	Reject H0

Only variables with statistically significant results in the partial test are shown.

Introduction   Methodology   Result

Result

Odds Ratio for Significant Variable

Variable	Category	Odds Ratio
Age	> 50 Year	11.4585
COVID-19 Status	Confirm	2.9522
Treatment Type	Inpatient Care	55.1651
Shortness of Breath	Yes	0.3746
Eating Disorders	Yes	66.5443

Patients aged **over 50** are **11.46 times more likely to die** than recover.

Patients with **confirmed COVID-19** status are **2.95 times more** likely to **die**.

Patients who received **outpatient care** are **55.17 times more** likely to **die**.

Patients experiencing **shortness of breath** are **2.67 times less** likely to **die**.

Patients with **eating disorders** are **66.54 times more** likely to **die** than recover.

Introduction

Methodology

Result

Result

Model evaluation on Test Data

Actual Data	Prediction Data	
	Recovered	Deceased
Recovered	99	64
Deceased	5	6

Confusion matrix of the **predicted** logistic regression model on the **test data**.

Accuracy: = 6+99/174 = 60.3%

Recall = 6/6+64 = 8.6%

Precision = 6/6+5 = 54.5%

F1-Score = 14.9%

The model **correctly** predicted **60.3%** of the **total cases**.

Among all patients **predicted** as deceased, **54.5%** were actually **deceased**.

The model identified **only 8.6%** of the **actual death cases**, indicating **poor sensitivity**.

The **balance** between precision and recall is **low**, showing **weak performance** on the **minority class**.

Recommendation

Pratama Hospital Yogyakarta

Enhance monitoring and care for high-risk patients, especially those aged >50, with confirmed COVID-19, outpatient treatment, shortness of breath, or eating disorders.

Strengthen nutritional support, as eating disorders significantly increase mortality risk.

Improve emergency and inpatient care facilities to manage rapid clinical deterioration effectively.
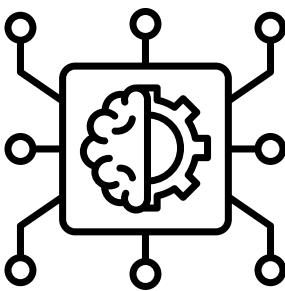
Future Research

Expand data coverage by including **longer time periods** and data from **multiple hospitals** for broader generalizability.

Incorporate additional predictors such as **smoking history**, **prior hospitalizations, or exposure** to positive cases.

Explore alternative models like **Naive Bayes**, **Random Forest,** or **SVM** to improve predictive performance and compare with logistic regression.

Introduction

Methodology

Result

Thank You

If you want to see the full analysis, check out my notebook on GitHub

https://github.com/billycemerson/Intern-Project