

Predictive Modeling for Heart Disease Assessment

Presented by: Aya EL HAJJ

Johnny CHREIM

Academic Supervisors Dr Malak KHREISS, PhD

Dr Boulous KHOUEIRY,

Company Supervisors Dr Eng. Riad ASSAF

Mr. Francis EL HELOU



Outline



INTRODUCTION



PROBLEM
STATEMENT



PROJECT
DESIGN



TOOLS USED IN
THE PROJECT



METHODOLOGY



RESULTS



Introduction

- The problem of Heart Disease
- Project Scope and Statement
- Project Objectives

Problem Statement



Using machine learning algorithm



Develop a predictive modeling system



Choose the right machine learning algorithm

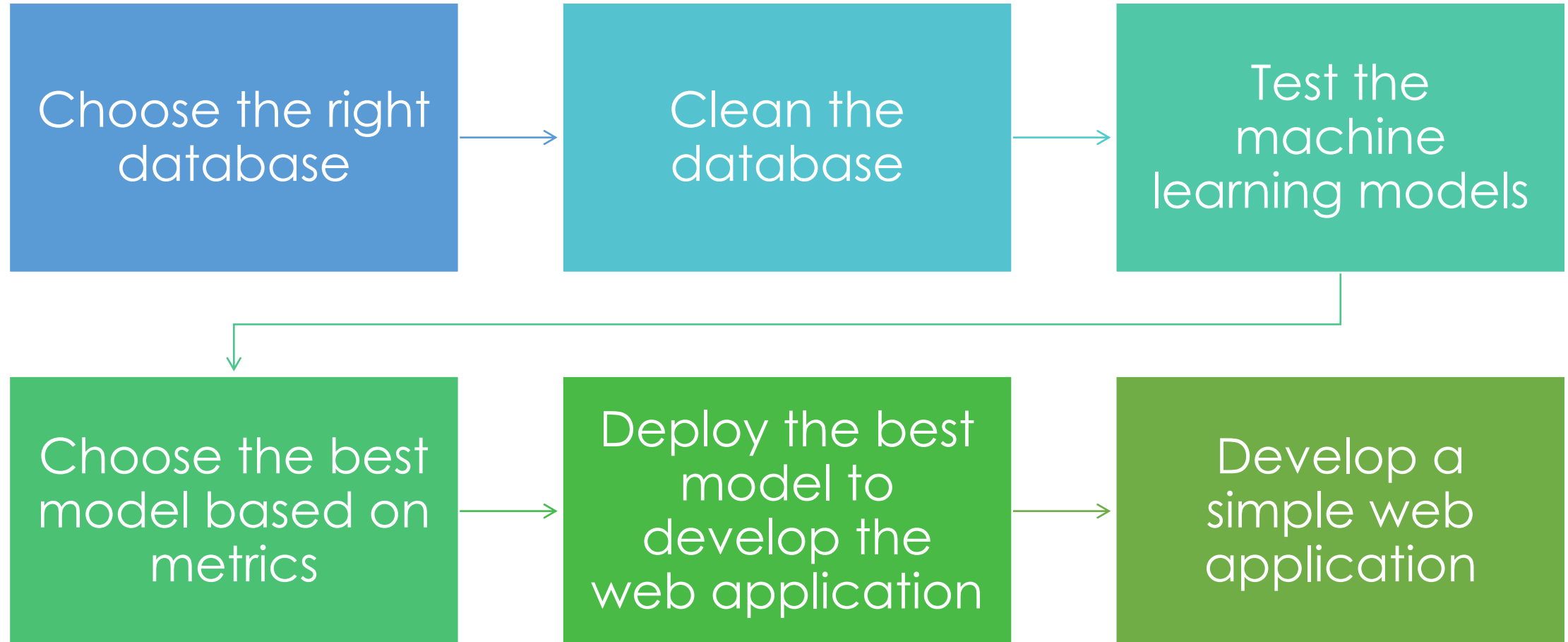


Integrate the model into a web app



Deploy the app for patients and doctors

Project Design





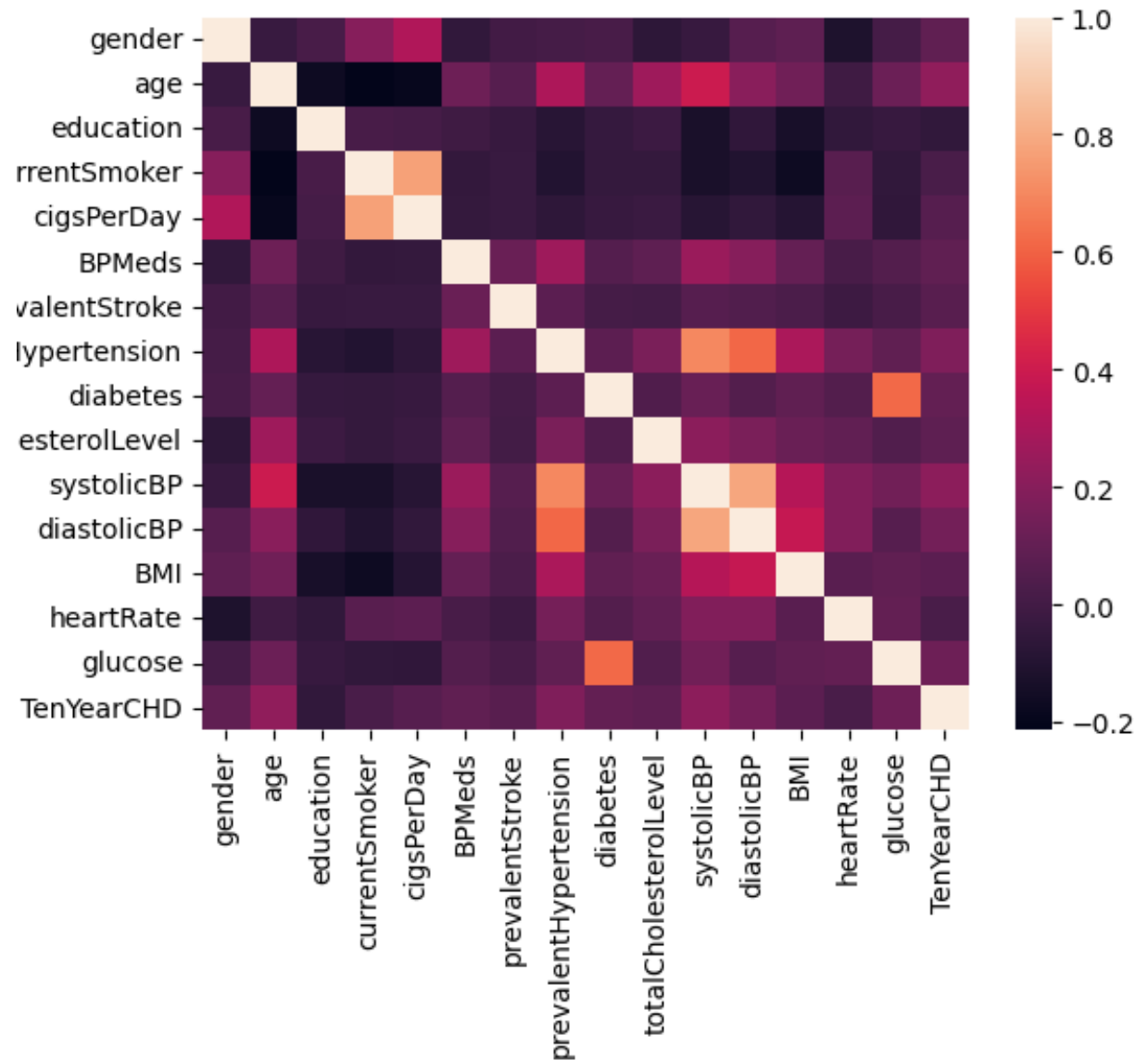
Tools used in the project

- Python:
 - Pandas for data cleaning
 - Numpy for data cleaning
 - Matplotlib for visualization
 - Seaborn for visualization
 - Sklearn for machine learning
 - Streamlit for app development
- Microsoft office word for the report
- Microsoft office power point for the presentation

Methodology:

Correlation matrix:

No correlation between
target variable and
education → remove
education feature



Methodology

Exploratory Data Analysis (EDA)



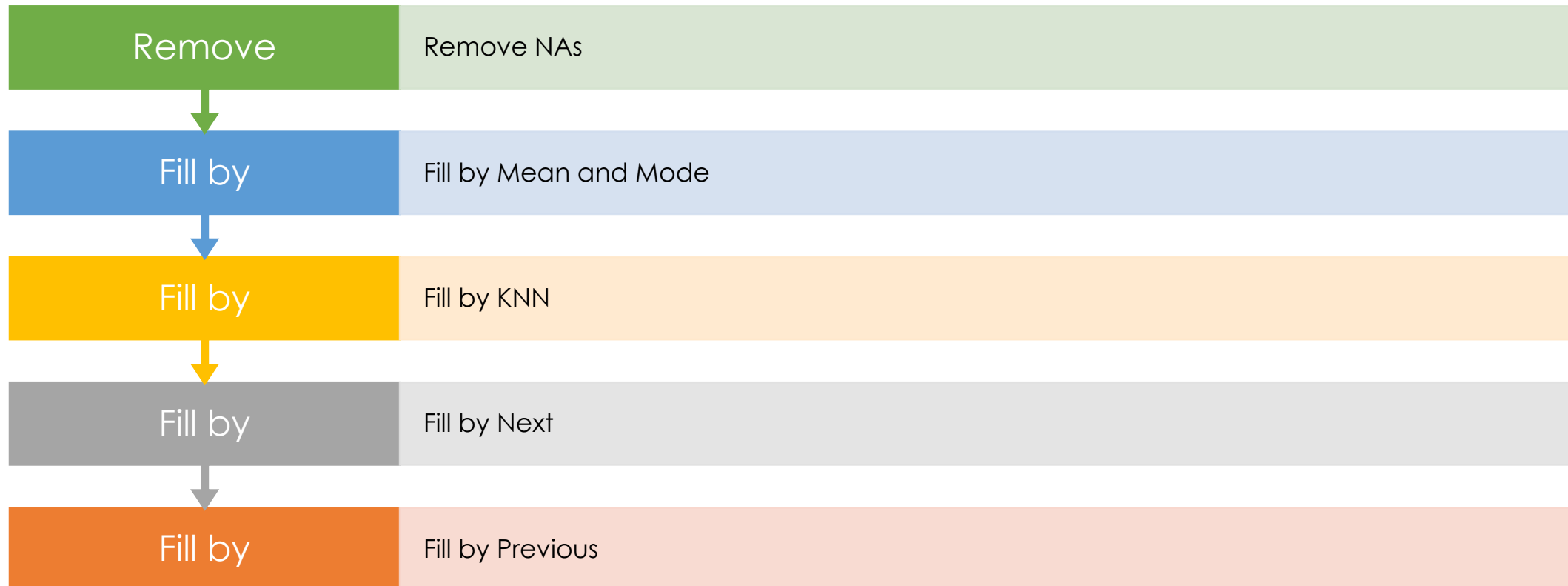
Dataset shape

Boxplots and
histograms for
numerical
features

Bar plots for
categorical
features

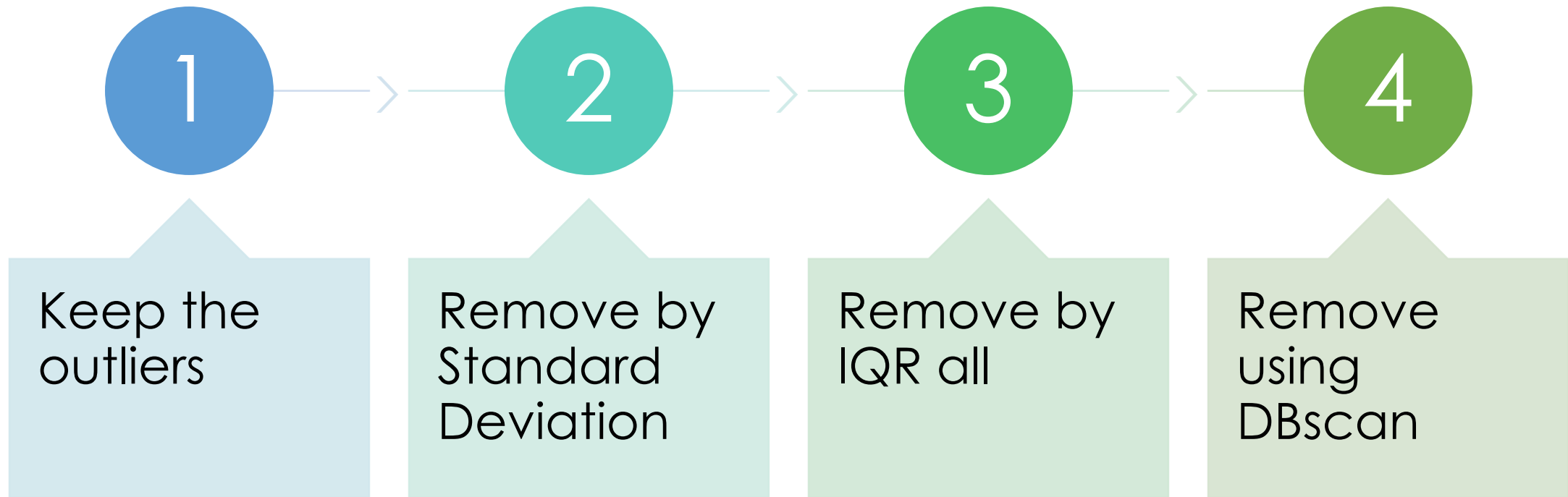
Methodology

Data Cleaning: Dealing with missing data



Methodology

Data Cleaning:
Dealing with outliers



Methodology

Data Cleaning:
Dealing with outliers



Tested each function

Remove by IQRAll → loss of all target
variable 1 value

We will not be using the function

Methodology

Data Cleaning:
Features Selection

Select K best features

Select percentile of the features

Select based on variance threshold

Dimensionality reduction

Keep all the features

Methodology

Data Cleaning:
Balancing

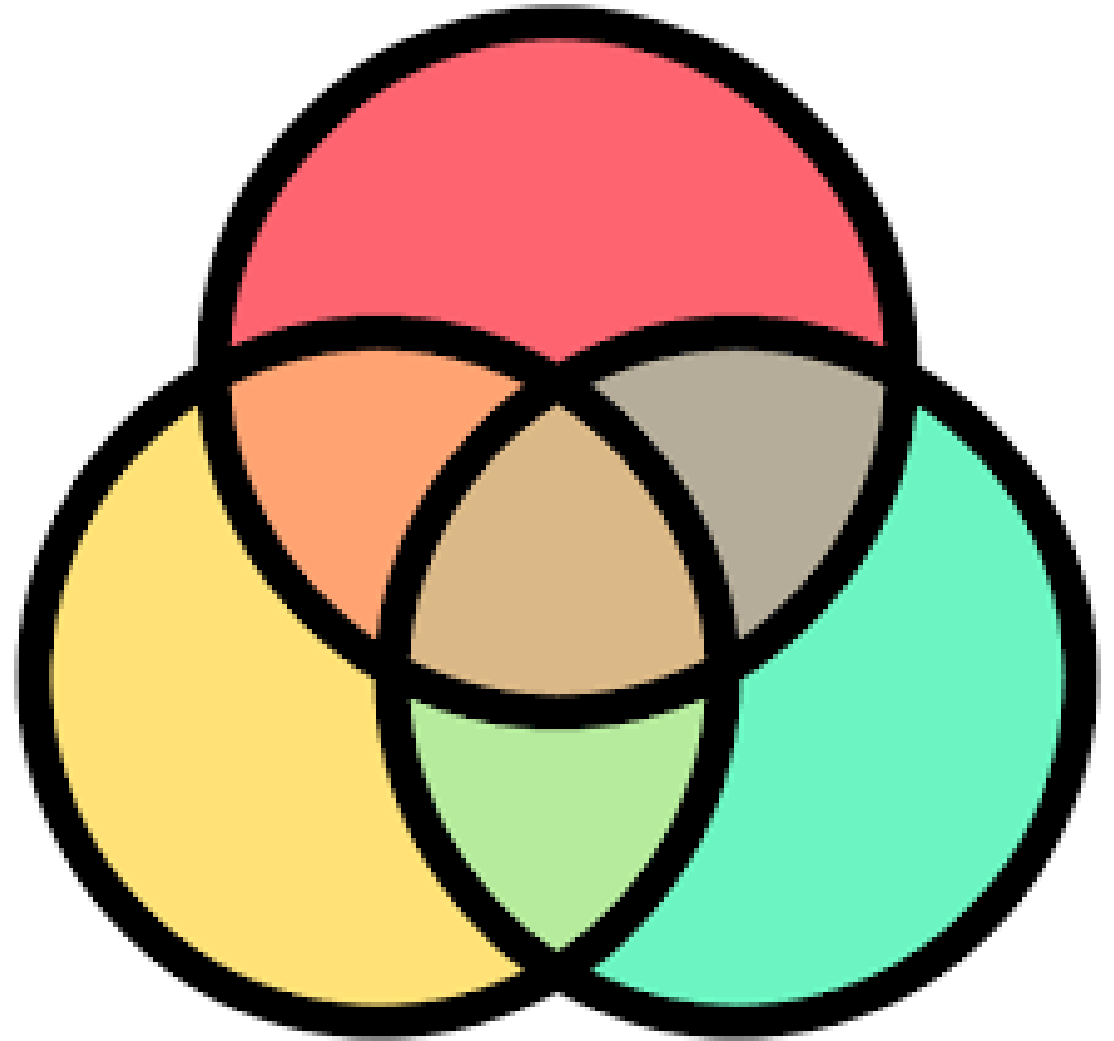


Over Sampling

Under Sampling

Methodology

Generate all possible
data frames for
modeling



Methodology

Modeling:

The below model will be tested on all the generated data frames

KNN (K nearest neighbor)

Logistic Regression

Decision Tree

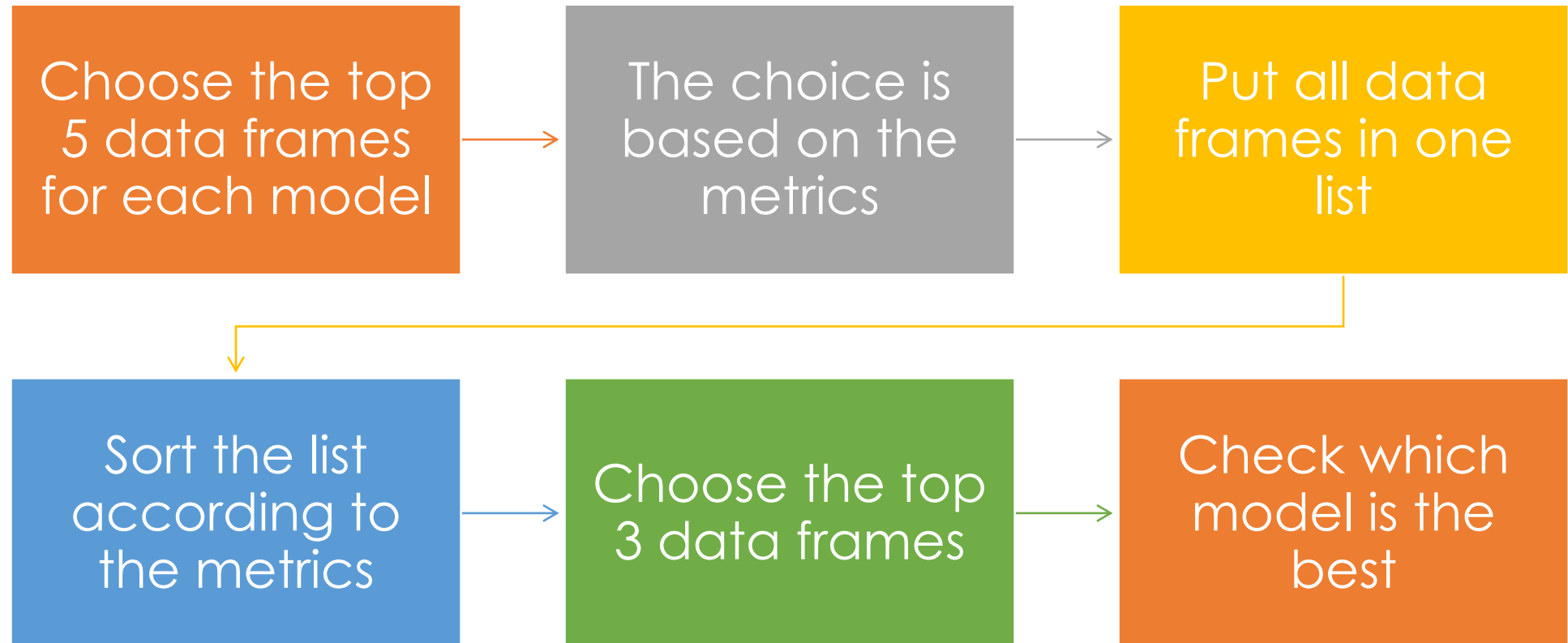
Naïve Bayes

Random Forest

SVM (Support Vector Machine)

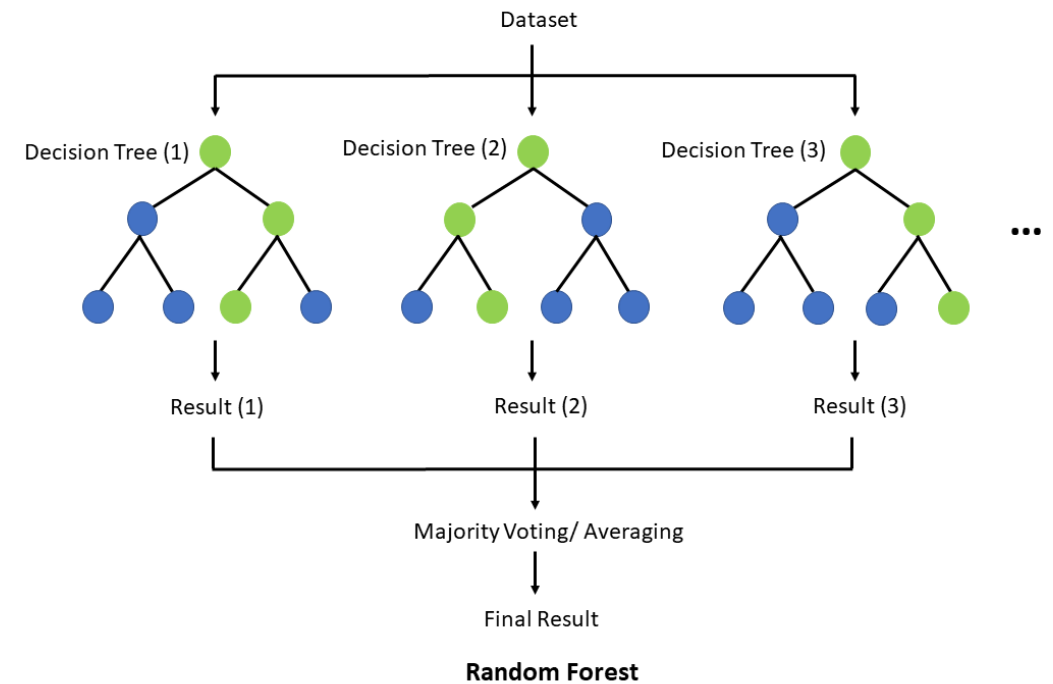
Methodology

Modeling: Choosing the best model



Methodology

Modeling:
Choosing the best model and data frame



Methodology

Deployment



Simple website



Using streamlit



Takes medical records from user



Generate percentage



Percentage is based on number of output 1 in the random forest model



[Try it out !](#)

Results



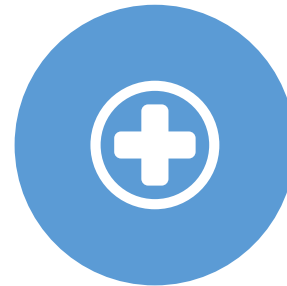
Predictive modeling
system



Can be used by
patients or doctors



Uses Random Forest
algorithm



Help doctors
prioritize patient
treatment



Questions

Thank you for your
attention !