HEART FAILURE PREDICTION USING PYSPARK

Pham Thuy Dung - 20521214

AGENDA

- Introduction
- Heart Failure Pediction
- Process Flow Diagram
- PySpark Overview
- Machine Learning
- Data analysis and visualization
- Results

INTRODUCTION

Heart failure prediction is an essential area of medical research and healthcare, aiming to identify individuals who are at risk of developing heart failure before they experience any symptoms.

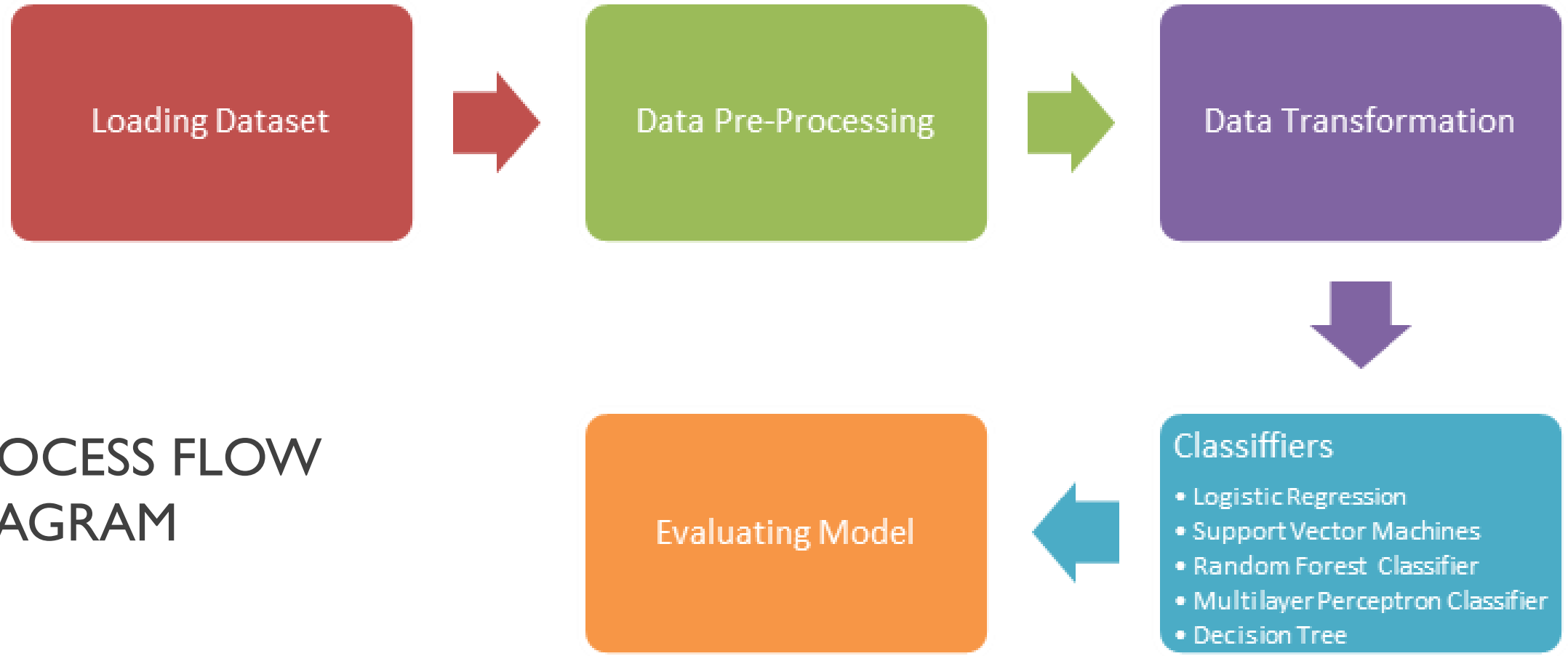
HEART FAILURE

- a medical condition related to the cardiovascular system
- no longer functions efficiently to pump enough blood to meet the body's needs

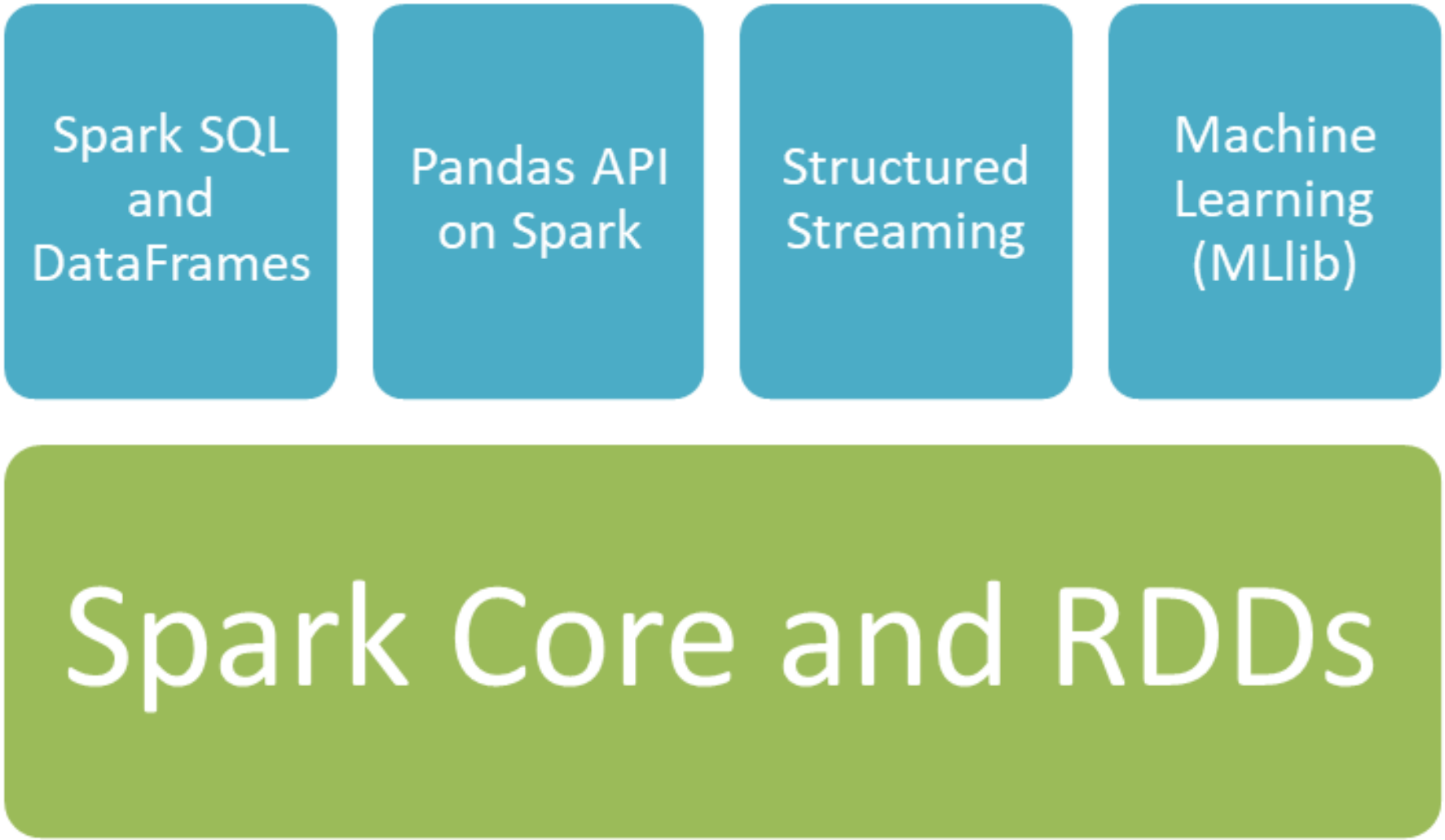
HEART FAILURE PREDICTION

- process of using various medical, clinical, and lifestyle data to assess an individual's risk of developing heart failure

PROCESS FLOW DIAGRAM



OVERVIEW OF PYSPARK



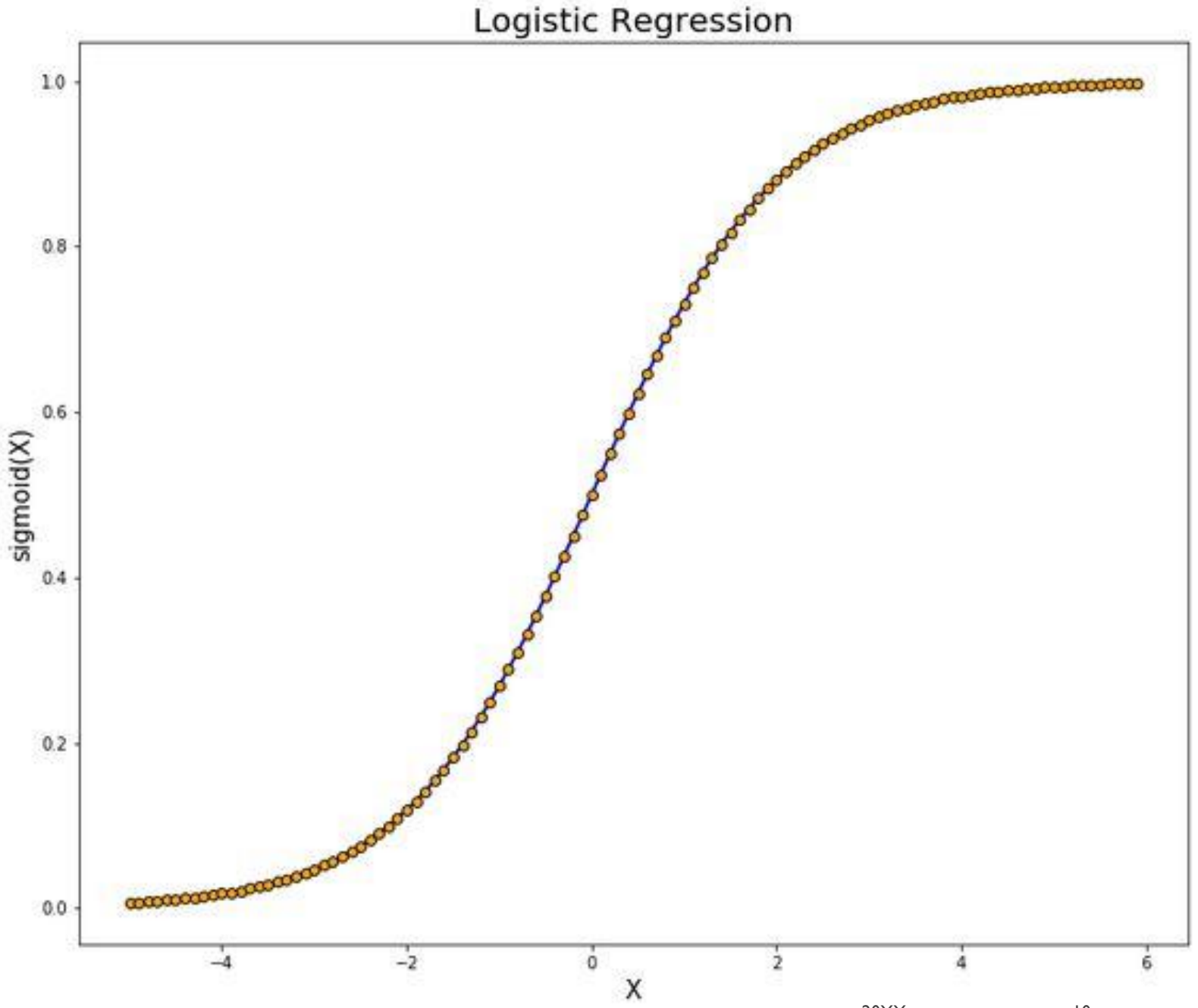
DATASET

Features	Description
Age	age
Sex	Sex(M, F)
ChestPainType	Chest pain type (ASY, NAP, Other)
RestingBP	Resting blood pressure
Cholesterol	Serum cholesrol
FastingBS	Fasting blood sugar
RestingECG	Resting electrocardiogram results (Normal, LVH, Others)
MaxHR	Maximum heart rate achieved
ExerciseAngina	Exercise induced angina
Oldpeak	Oldpeak = ST
ST_Slope	The slope of the peak exercise ST segment
HeartDisease	Target to predict

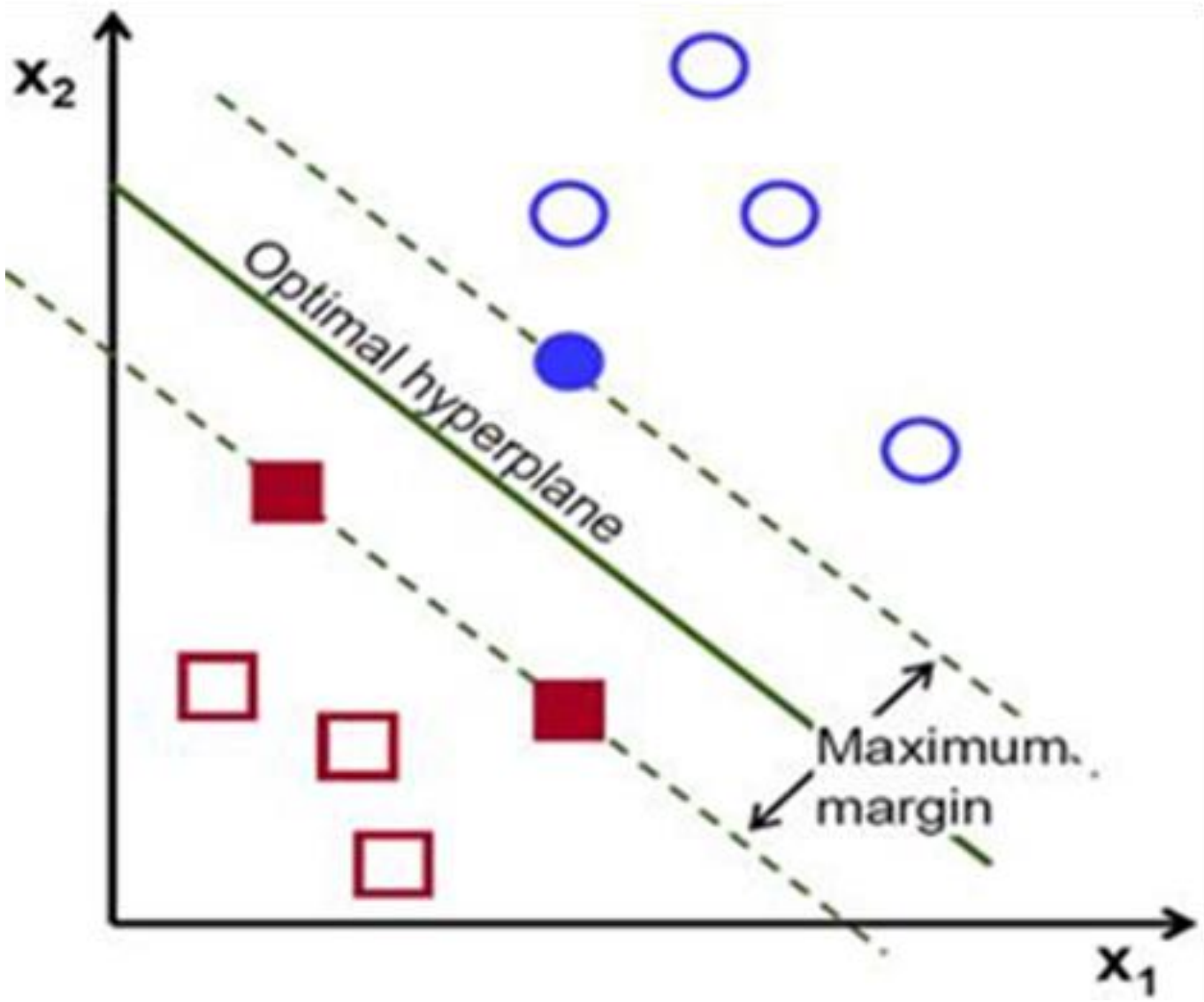
MACHINE LEARNING

LOGISTICS REGRESSION

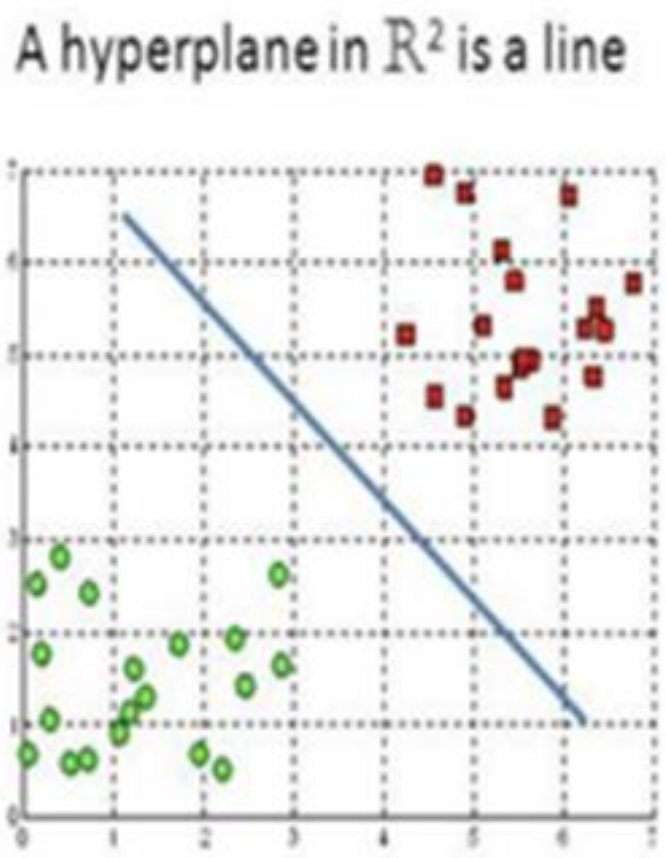
$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}}$$

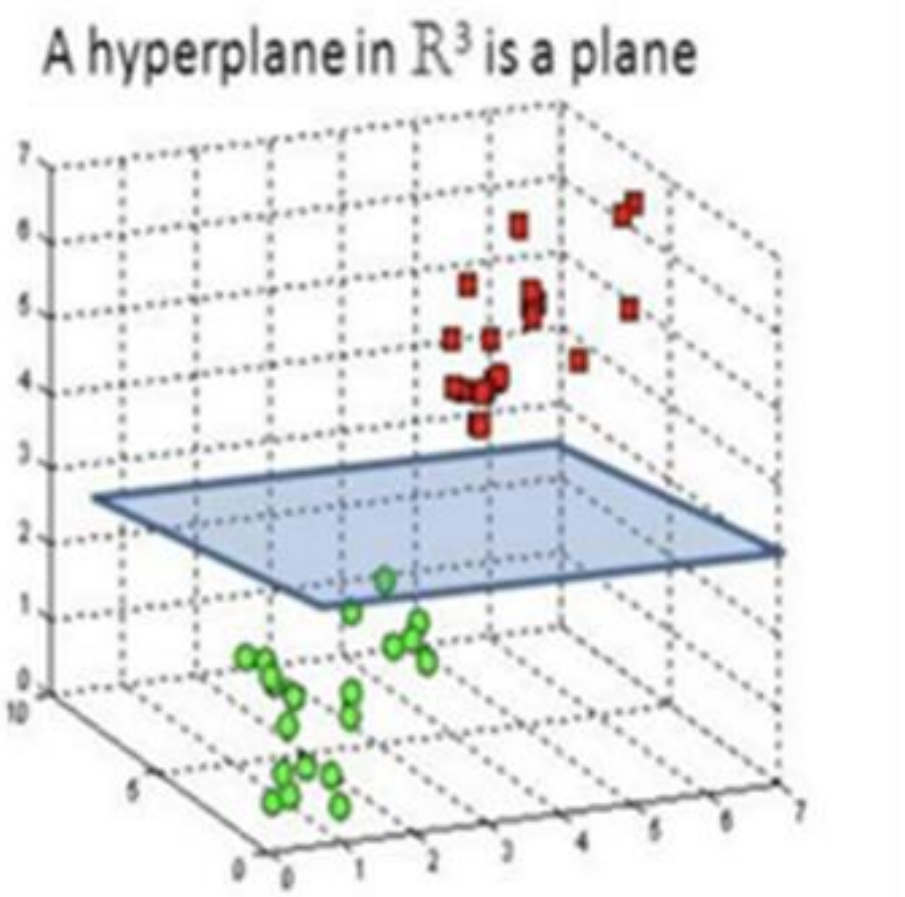


SUPPORT VECTOR MACHINE (SVM)

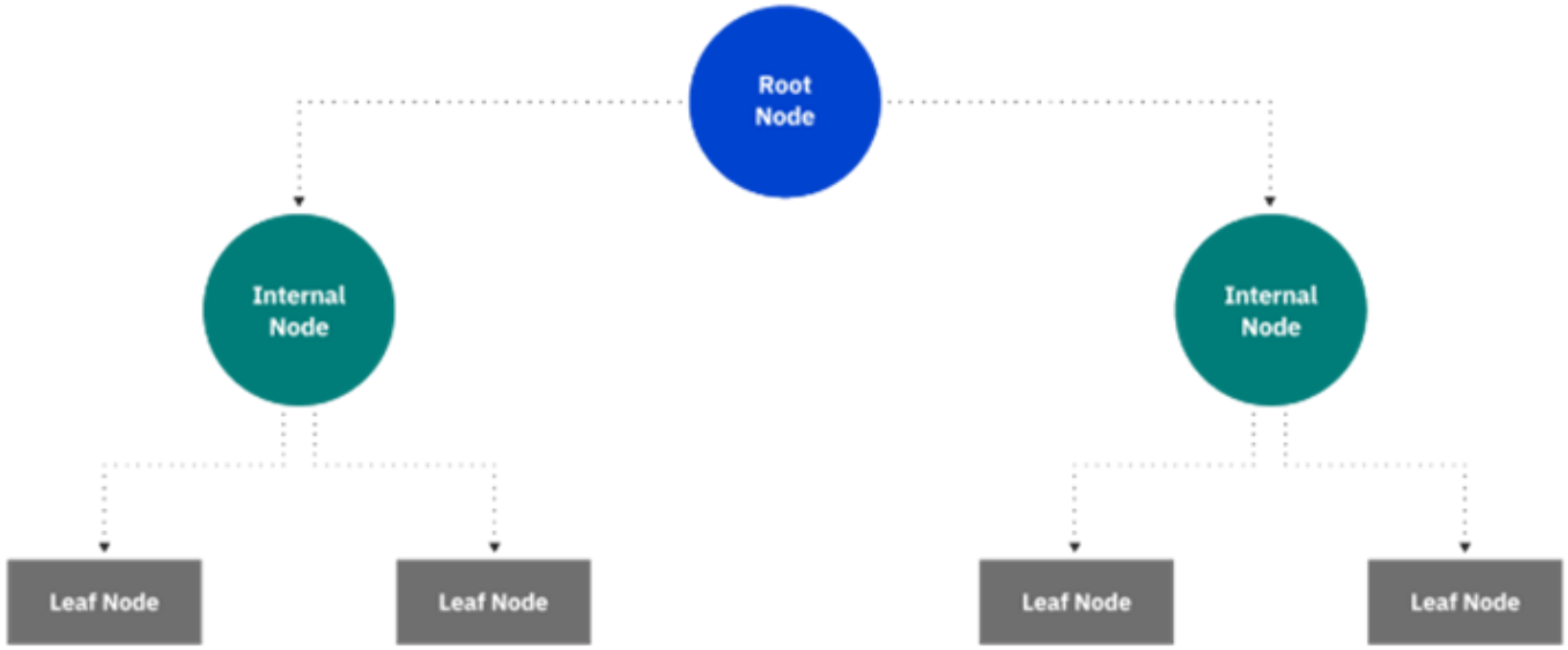


SUPPORT VECTOR MACHINE (SVM)

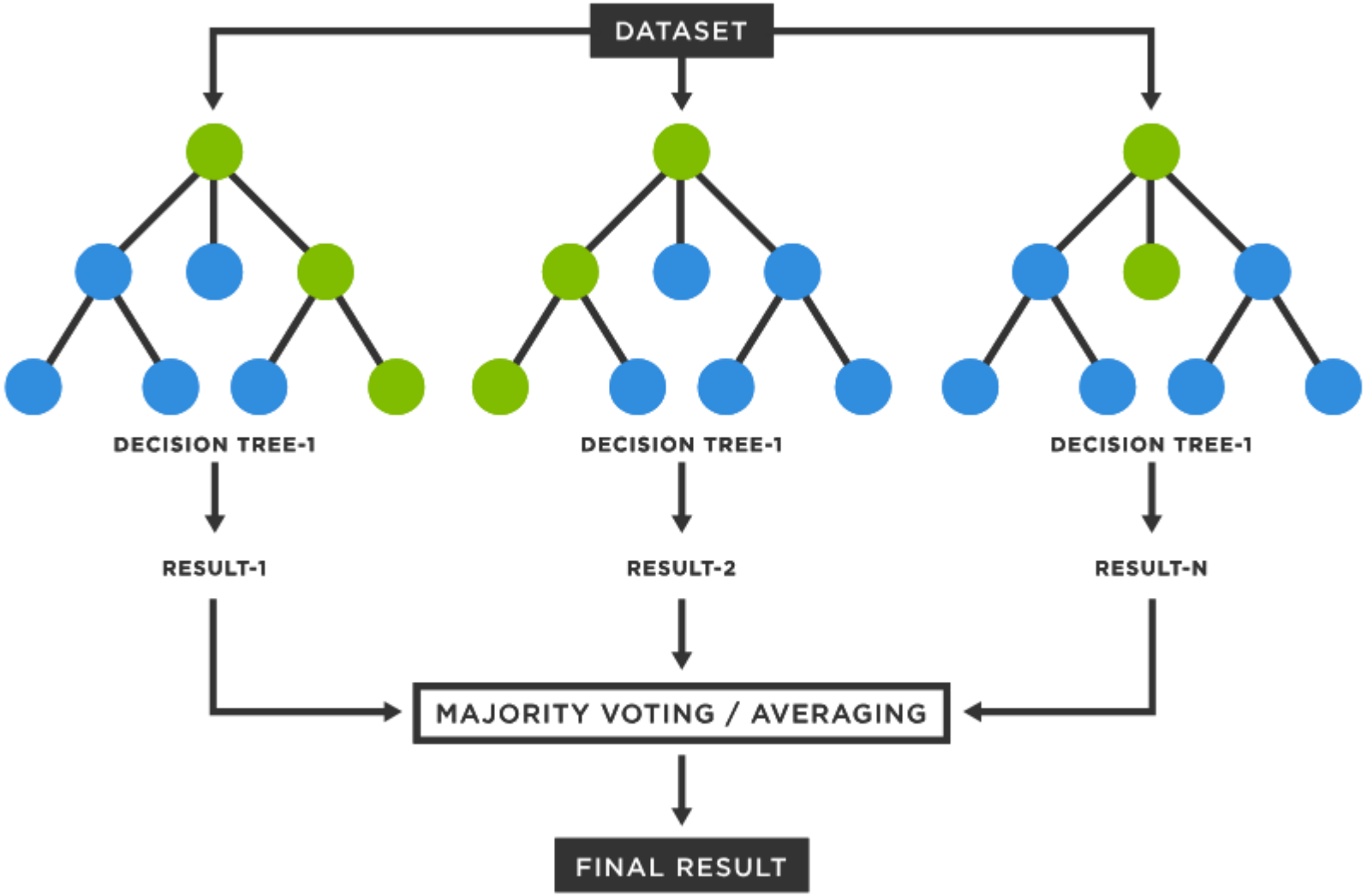




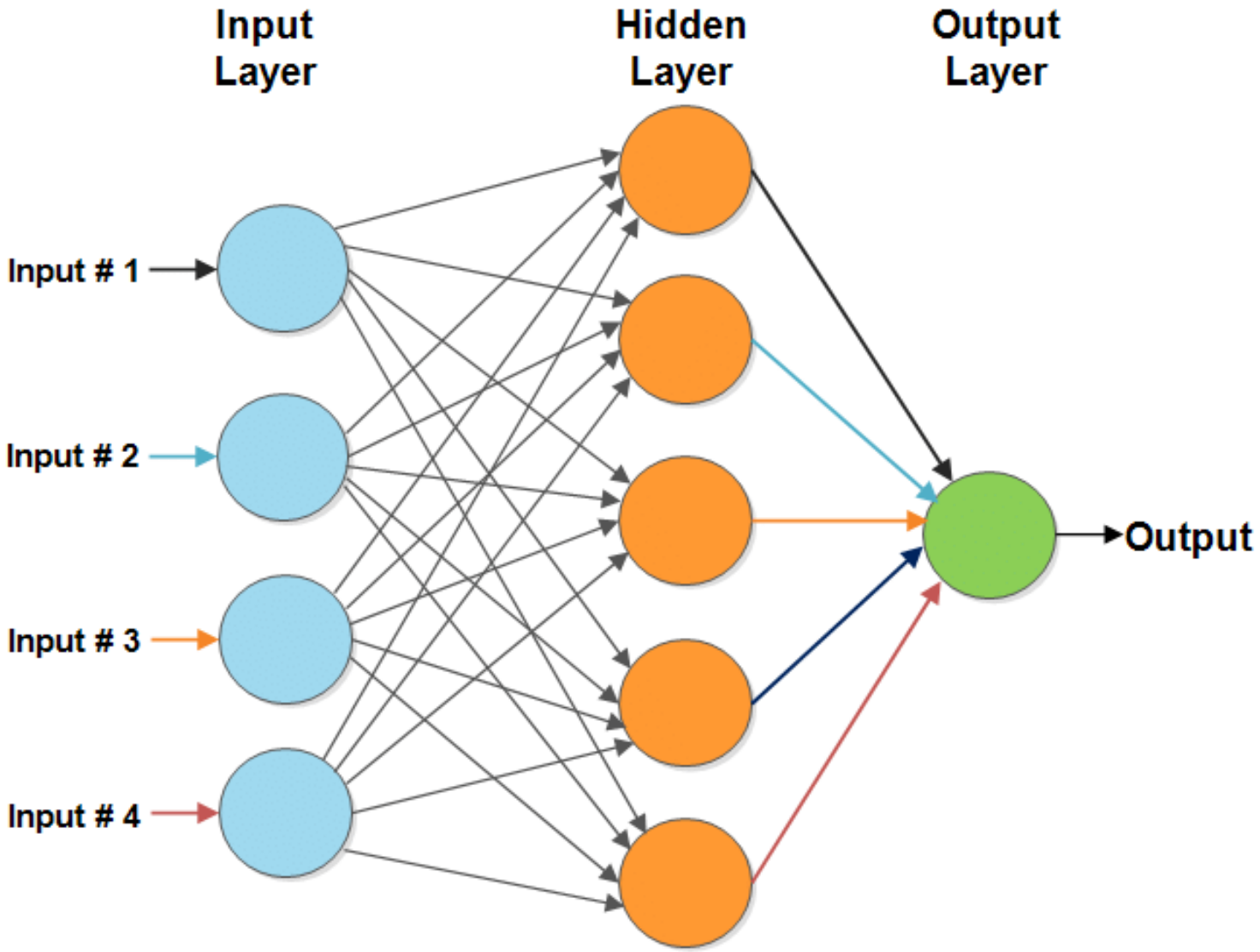
DECISION TREE



RANDOM FOREST

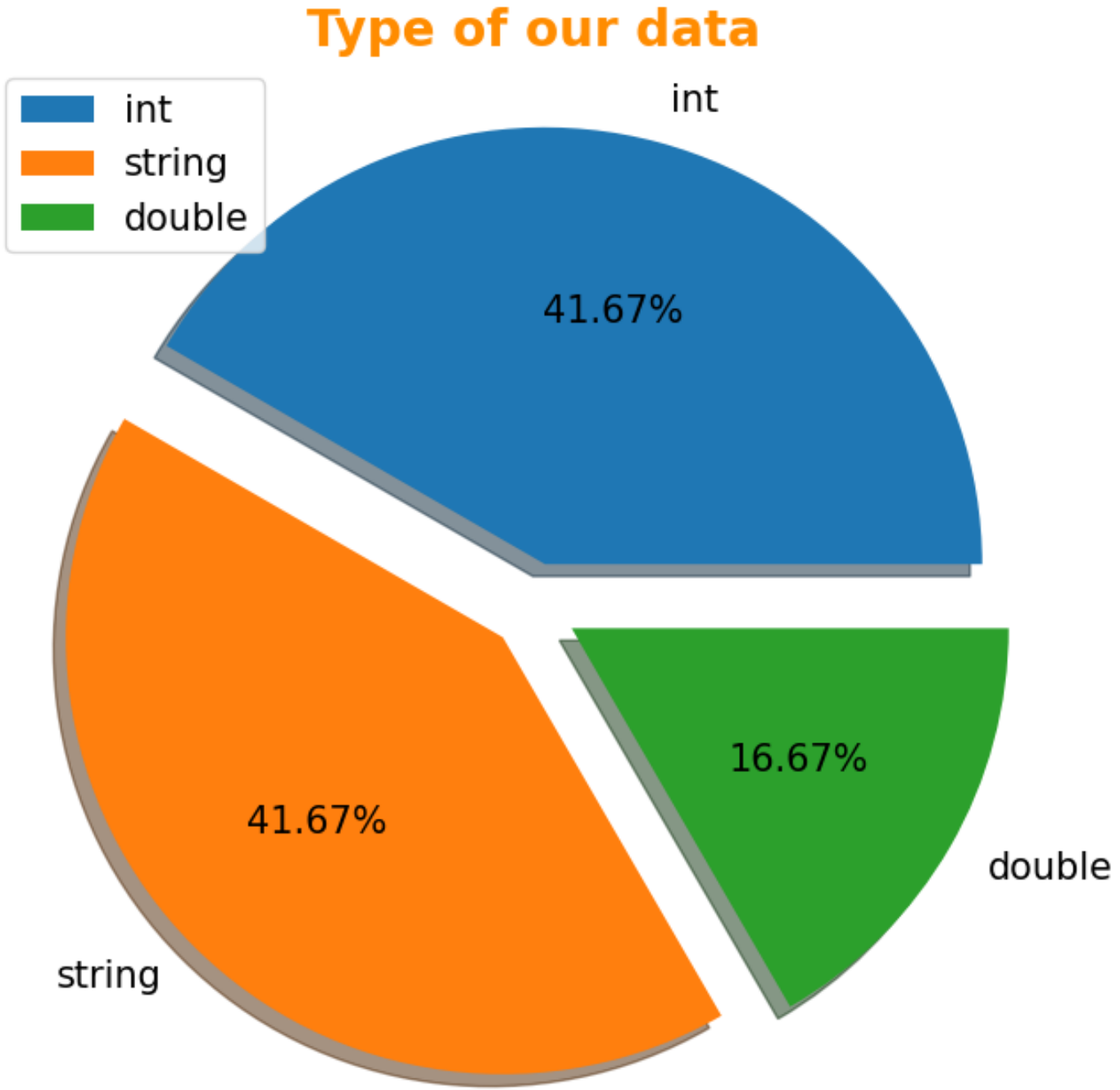


MULTI-LAYER PERCEPTION

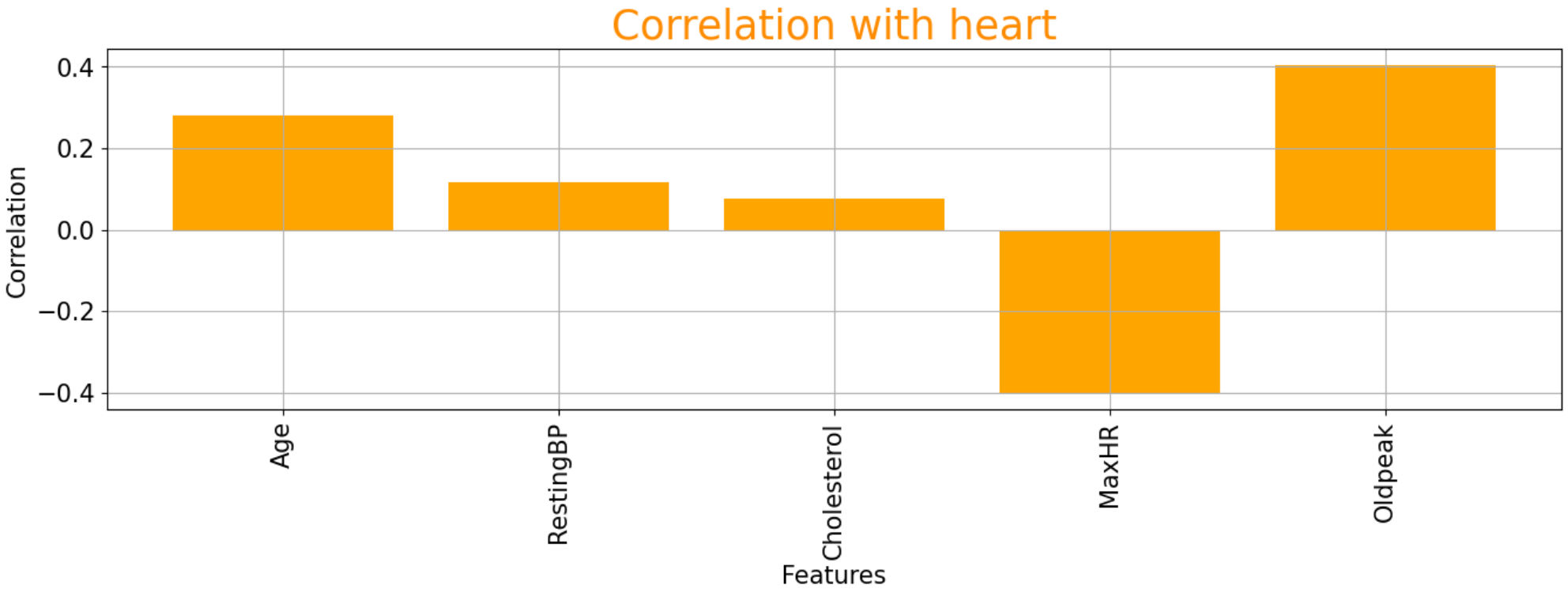


DATA ANALYSIS & DATA VISUALIZATION

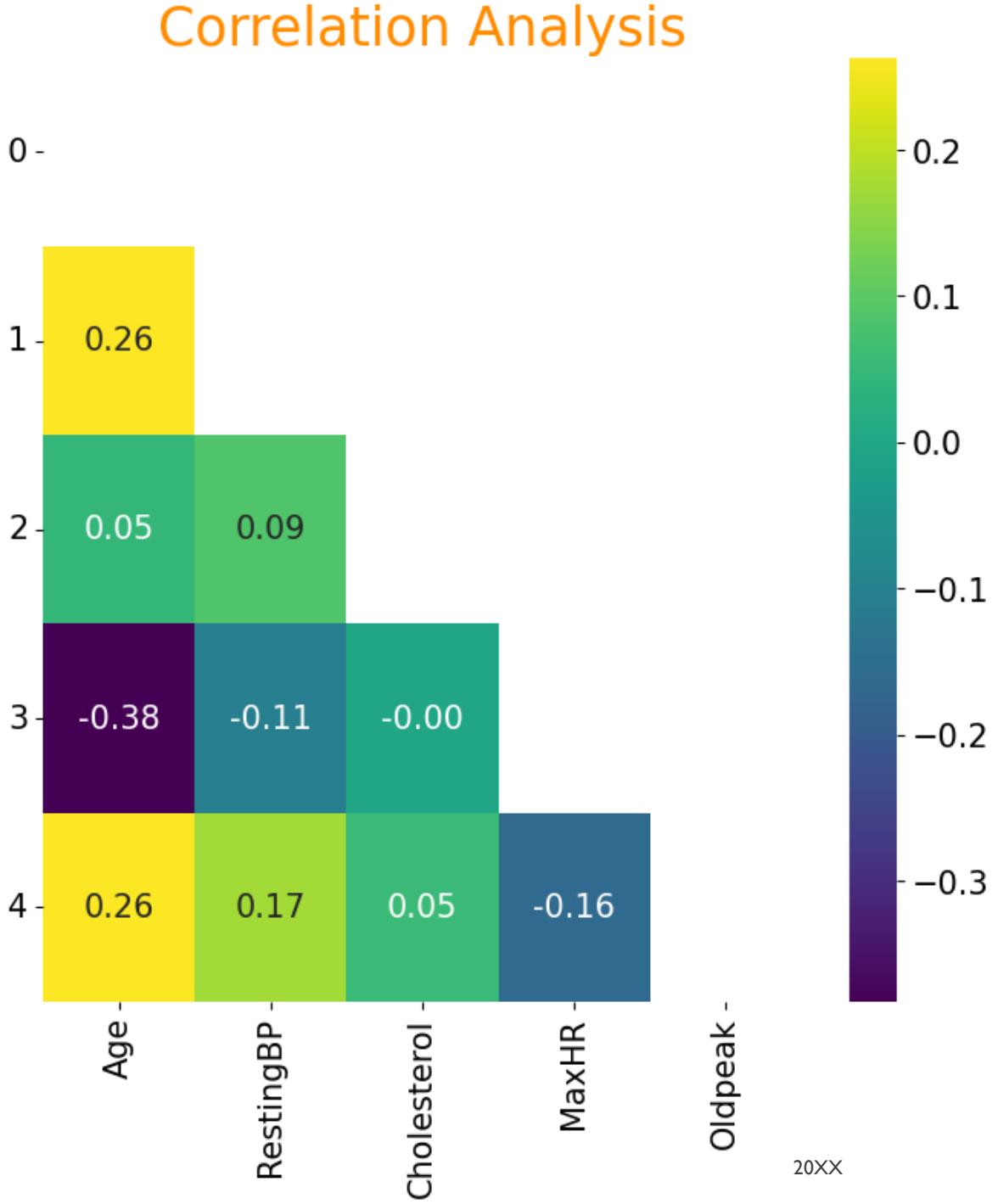
A PIE CHART FOR TYPE OF DATA



A BAR PLOT FOR CORRELATION WITH HEART

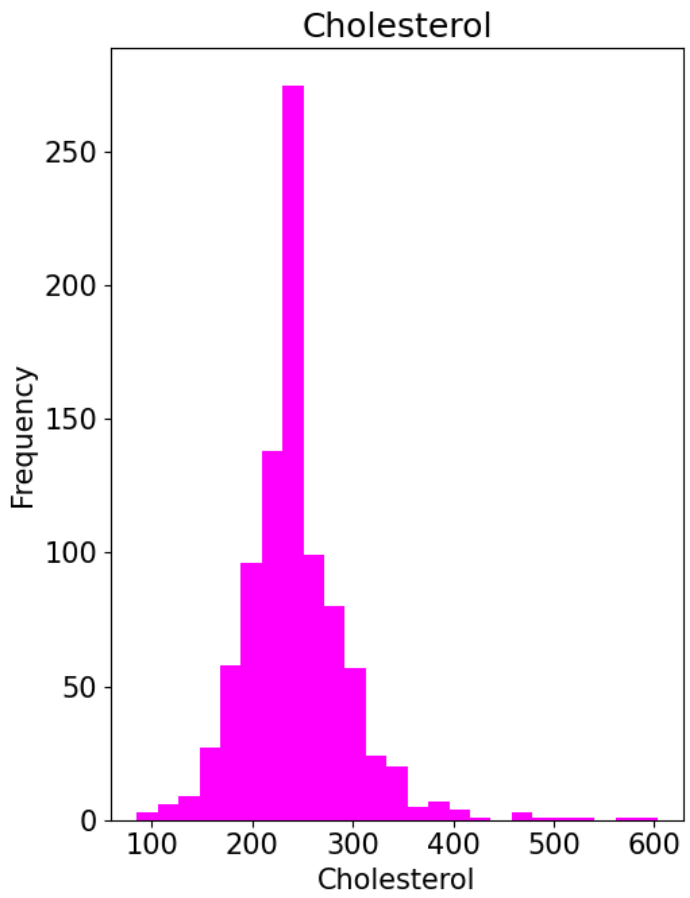


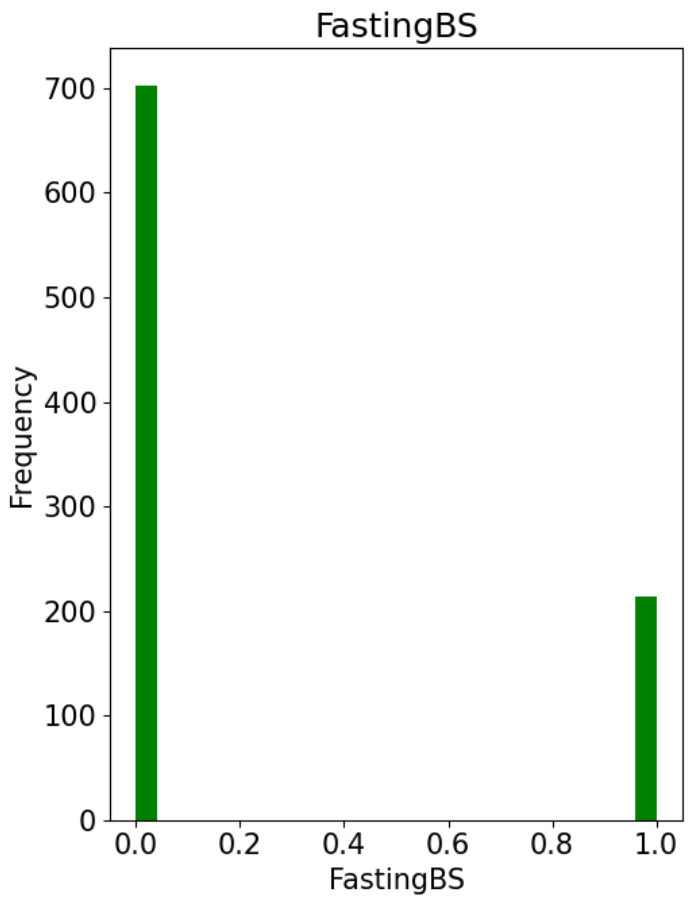
HEATMAP FOR CORRELATION ANALYSIS

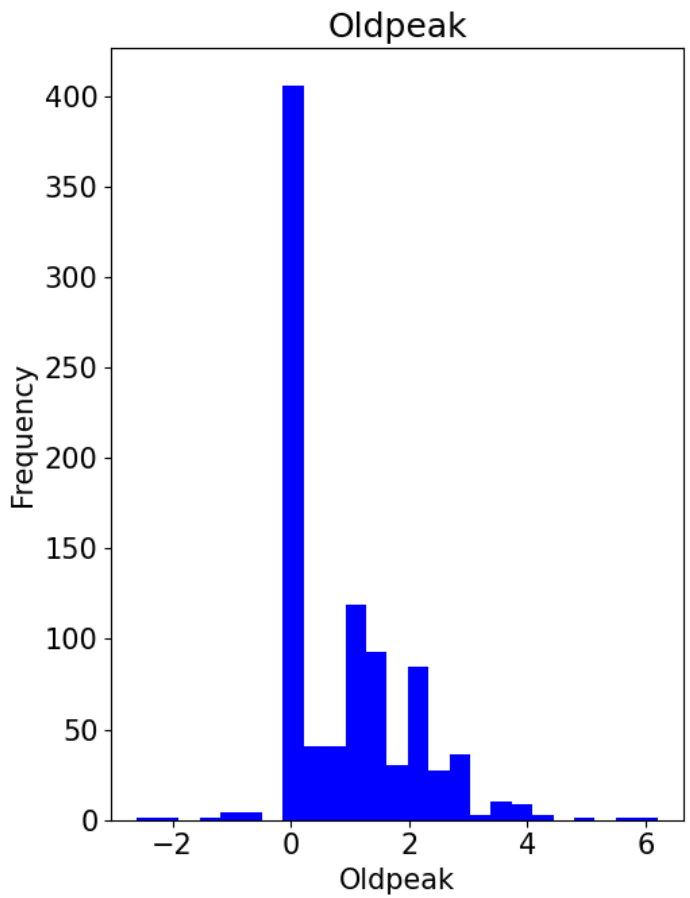


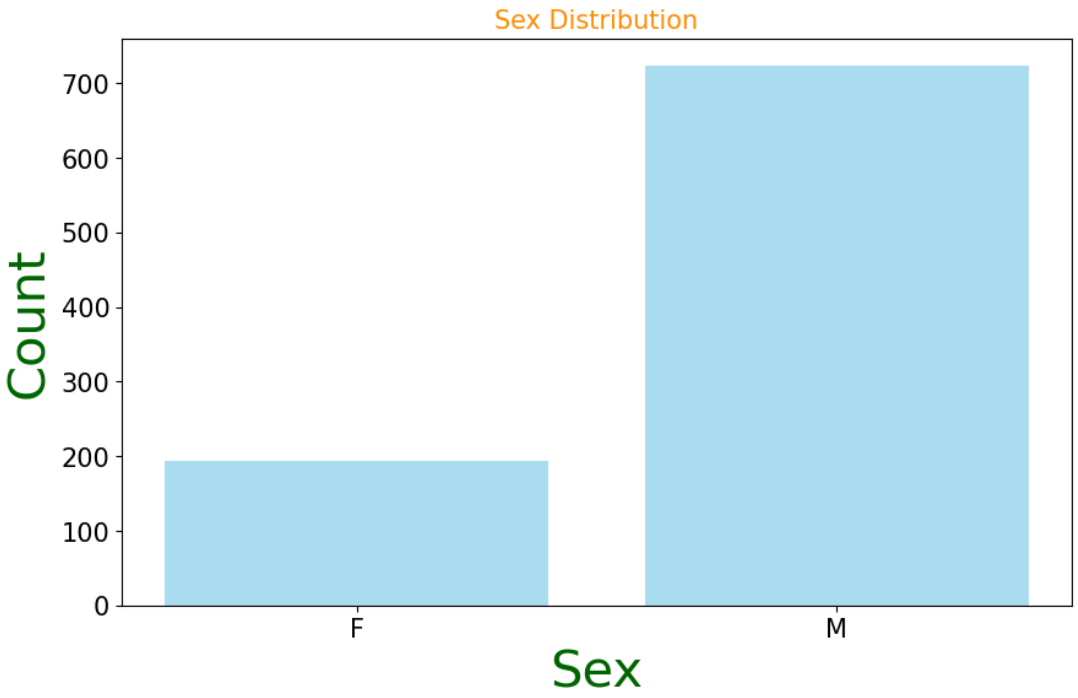
HISTOGRAMS FOR THREE NUMERICAL COLUMNS

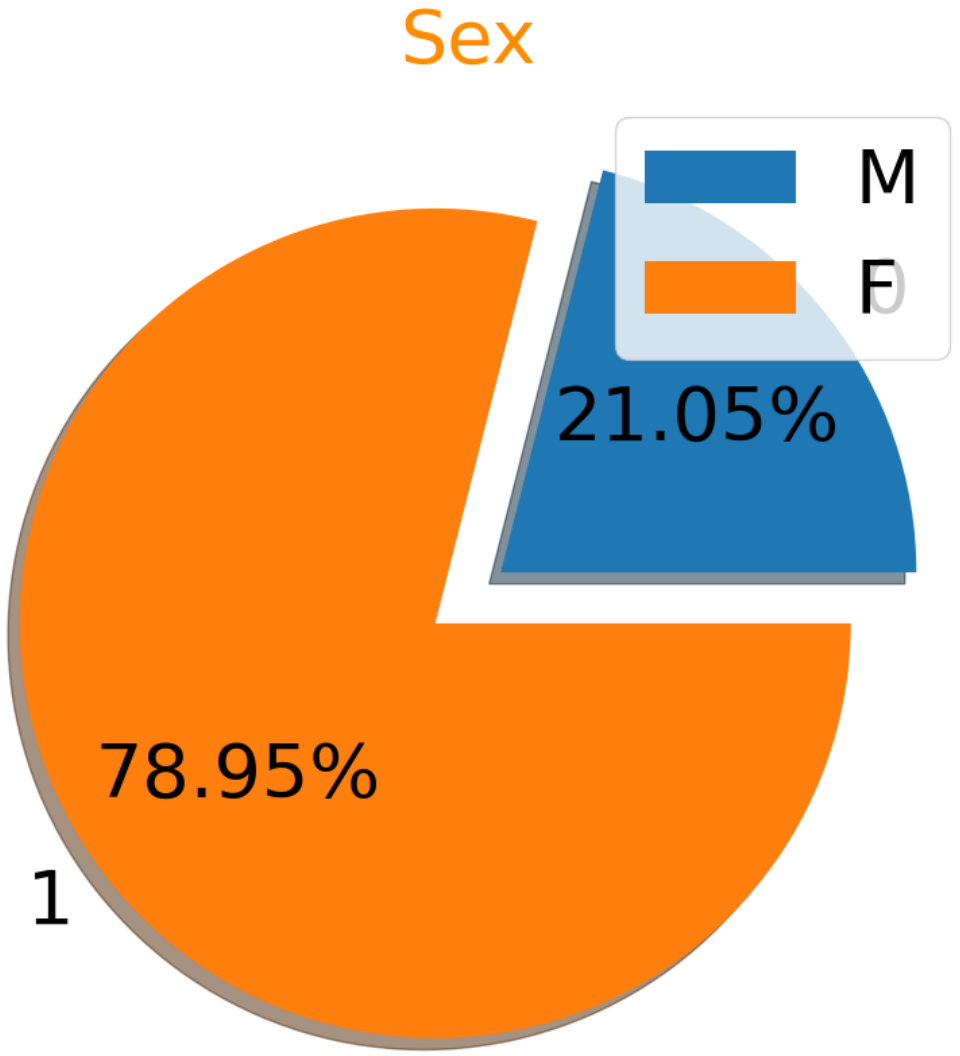
Subtitle





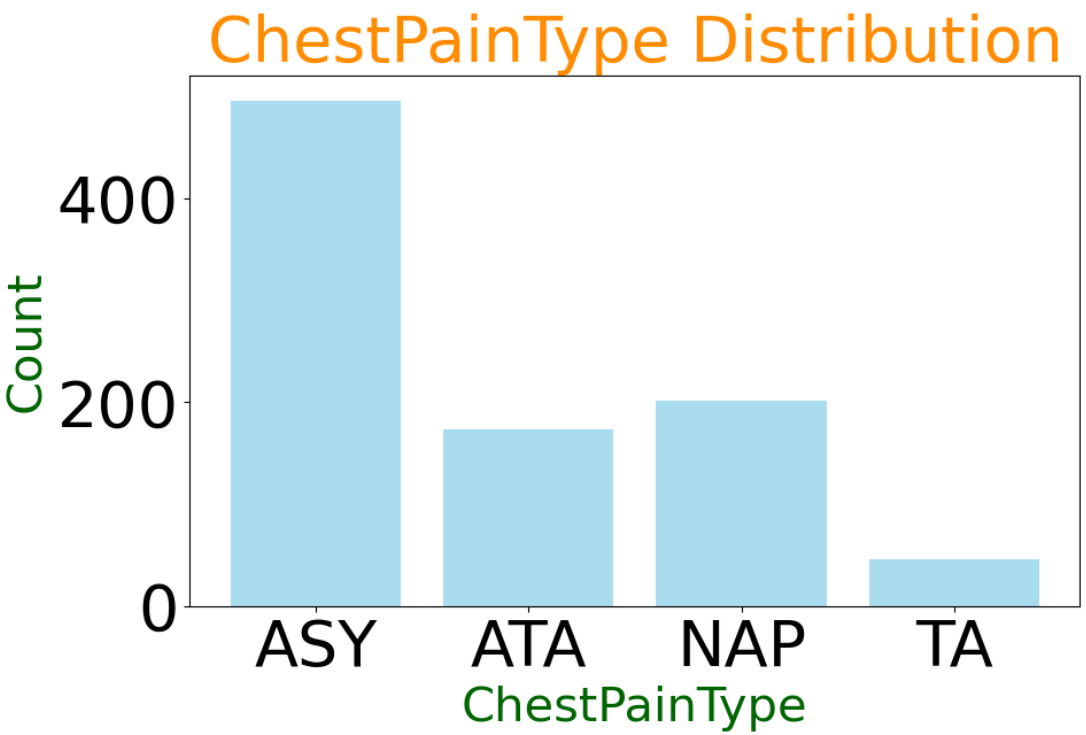


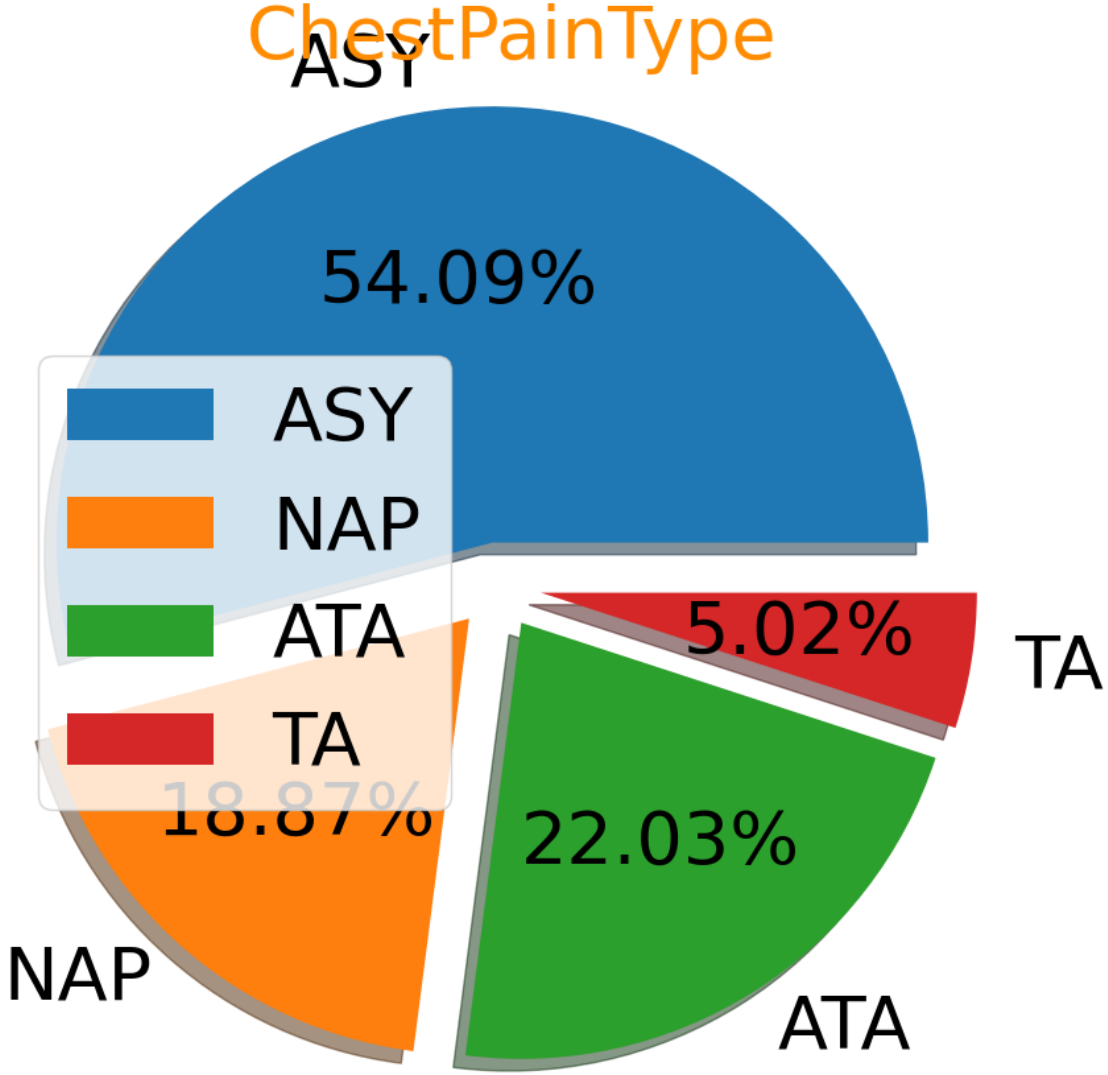




VISUALIZATION OF SEX

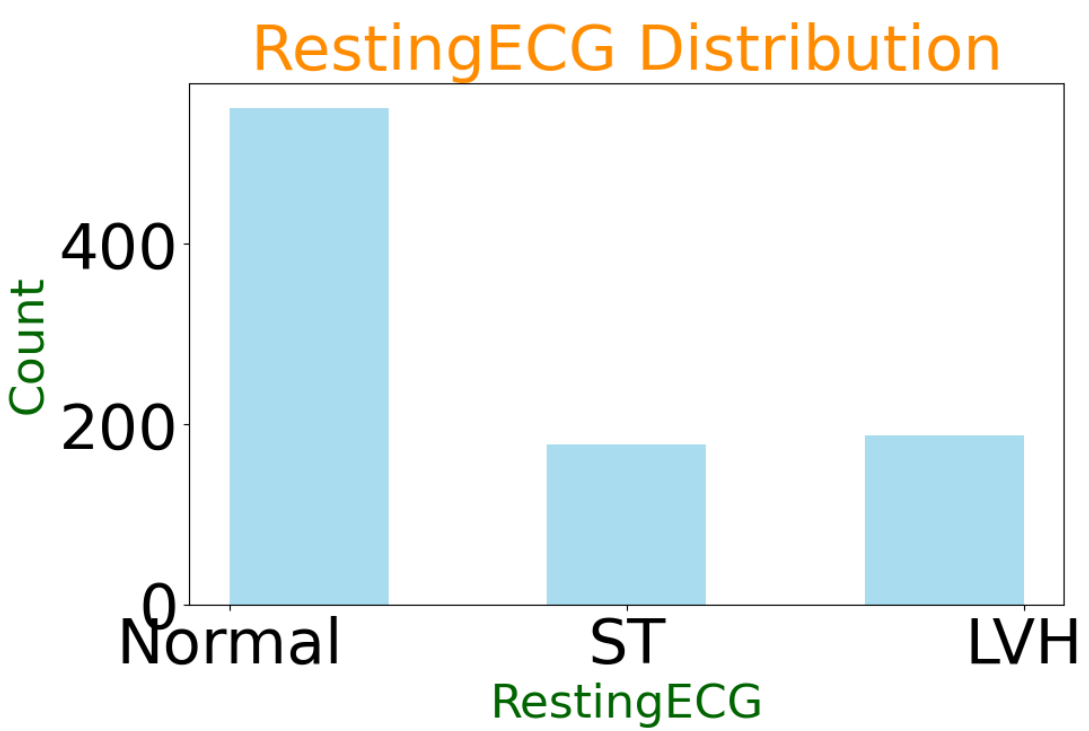
	Sex	Total	Percentage
0	F	193	21.05
1	M	724	78.95

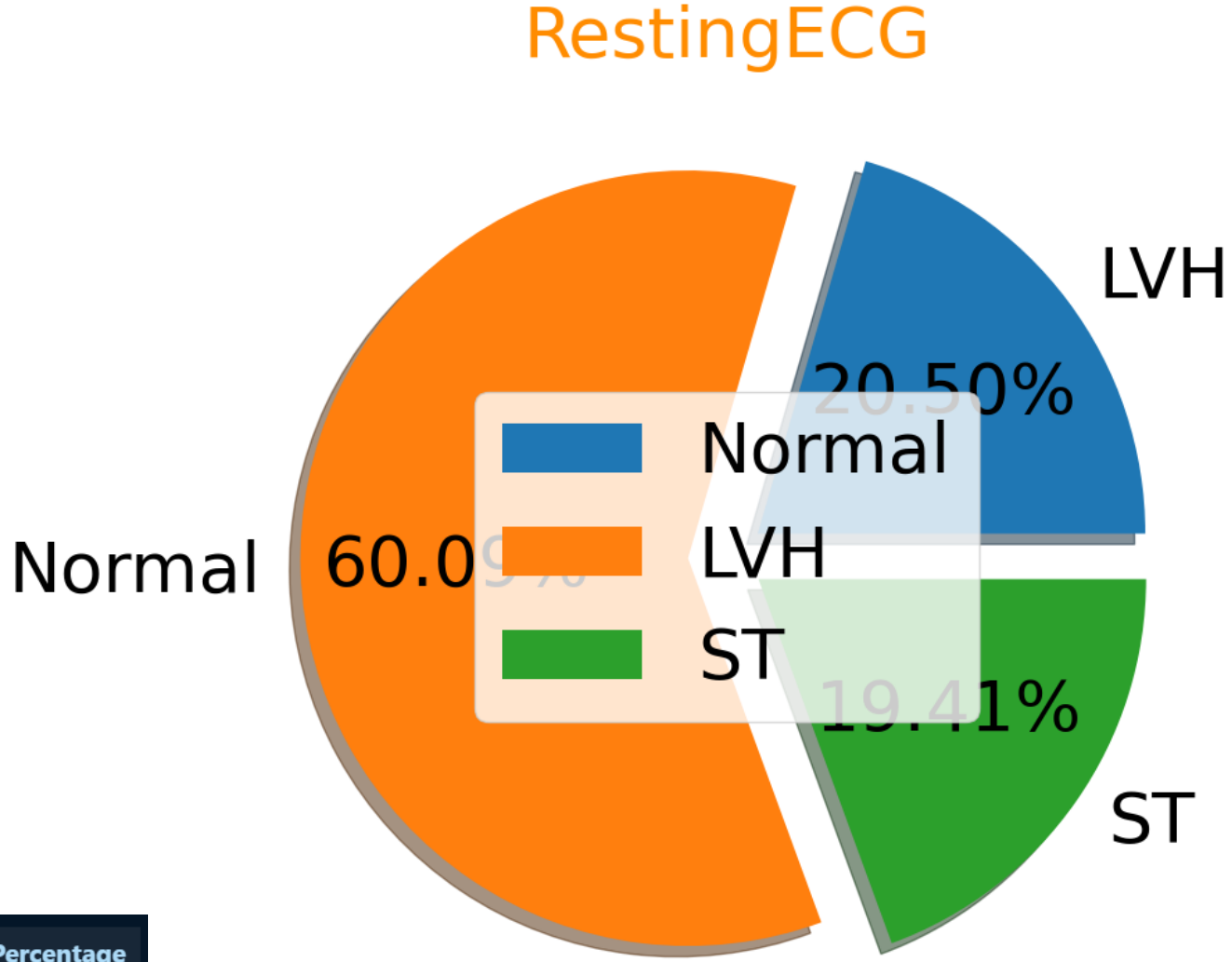




VISUALIZATION OF CHEST PAIN TYPE

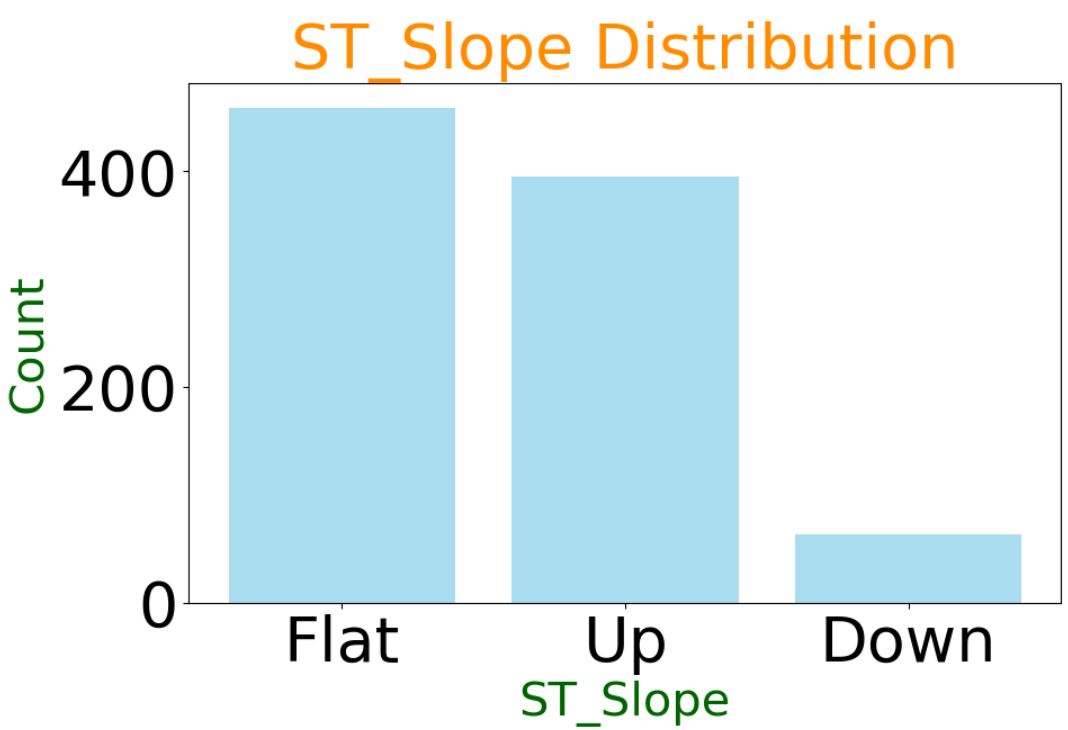
	ChestPainType	Total	Percentage
0	NAP	202	22.03
1	ATA	173	18.87
2	TA	46	5.02
3	ASY	496	54.09

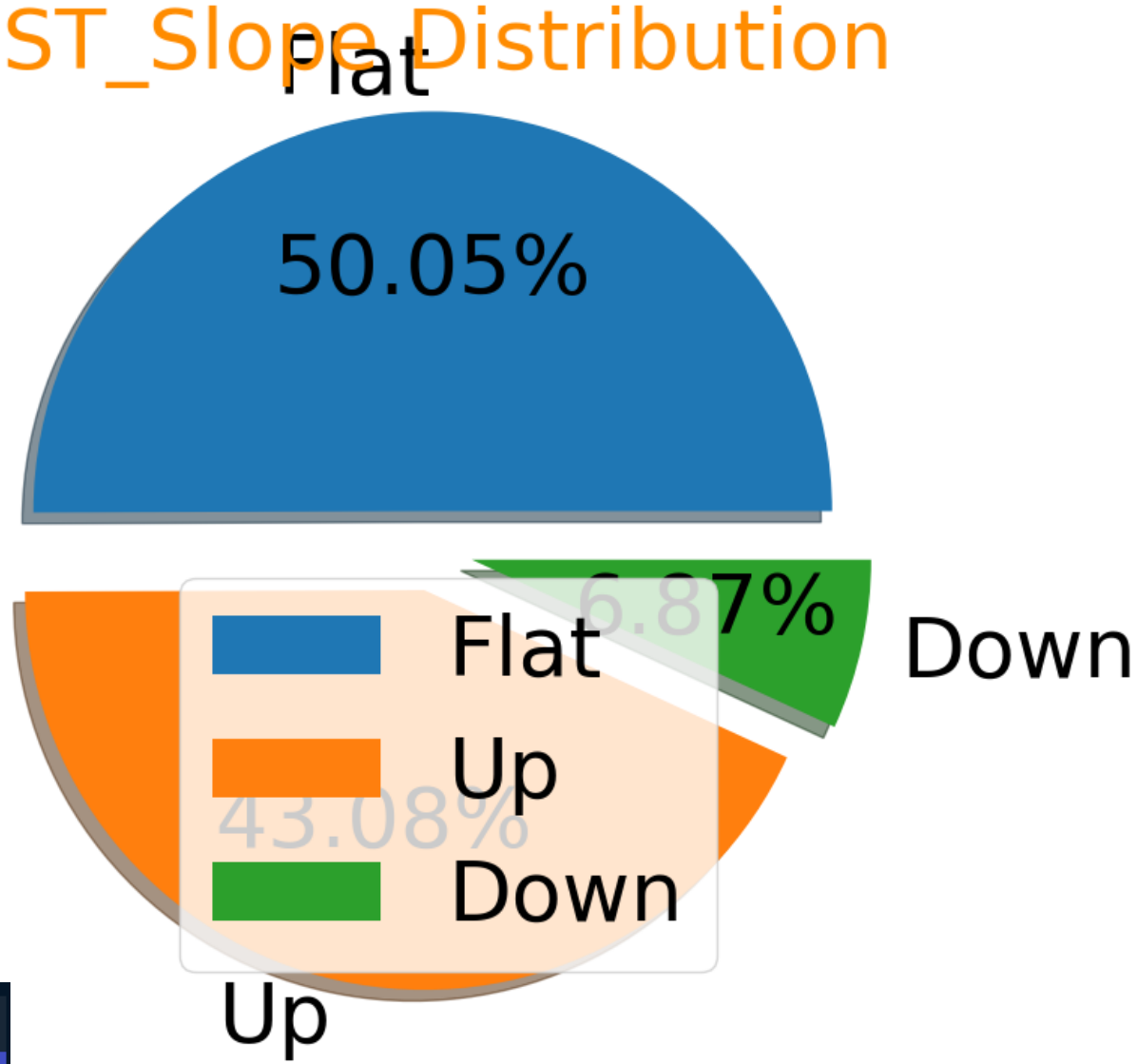




VISUALIZATION OF RESTINGECG

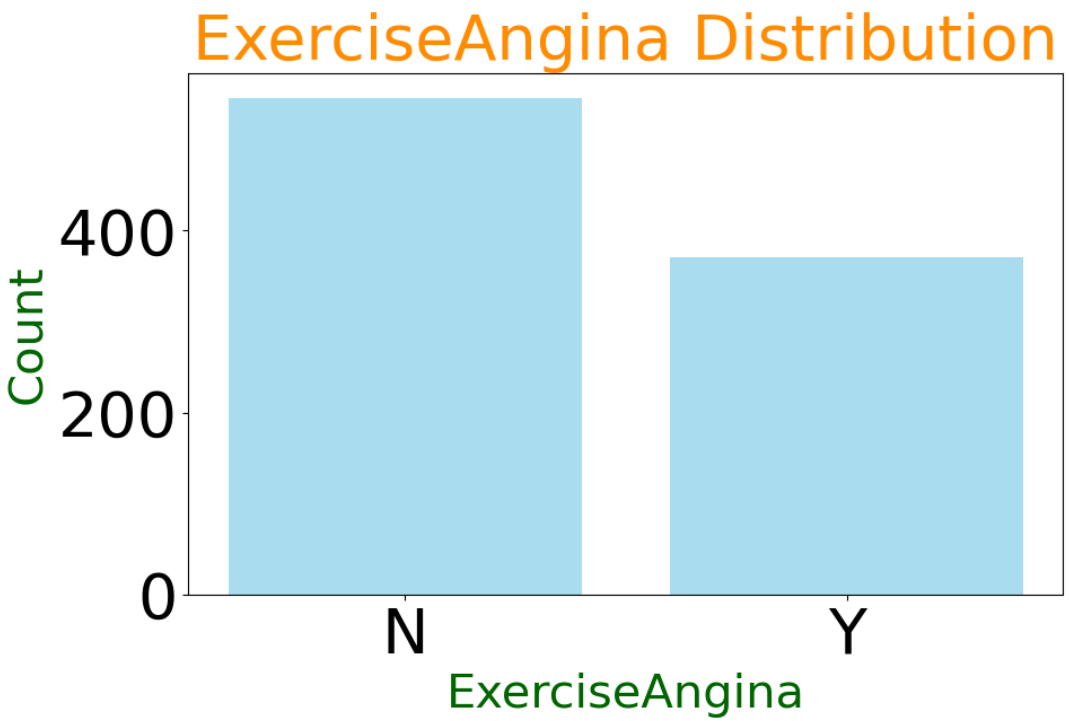
RestingECG	Total	Percentage
Normal	551	60.09
LVH	188	20.50
ST	178	19.41

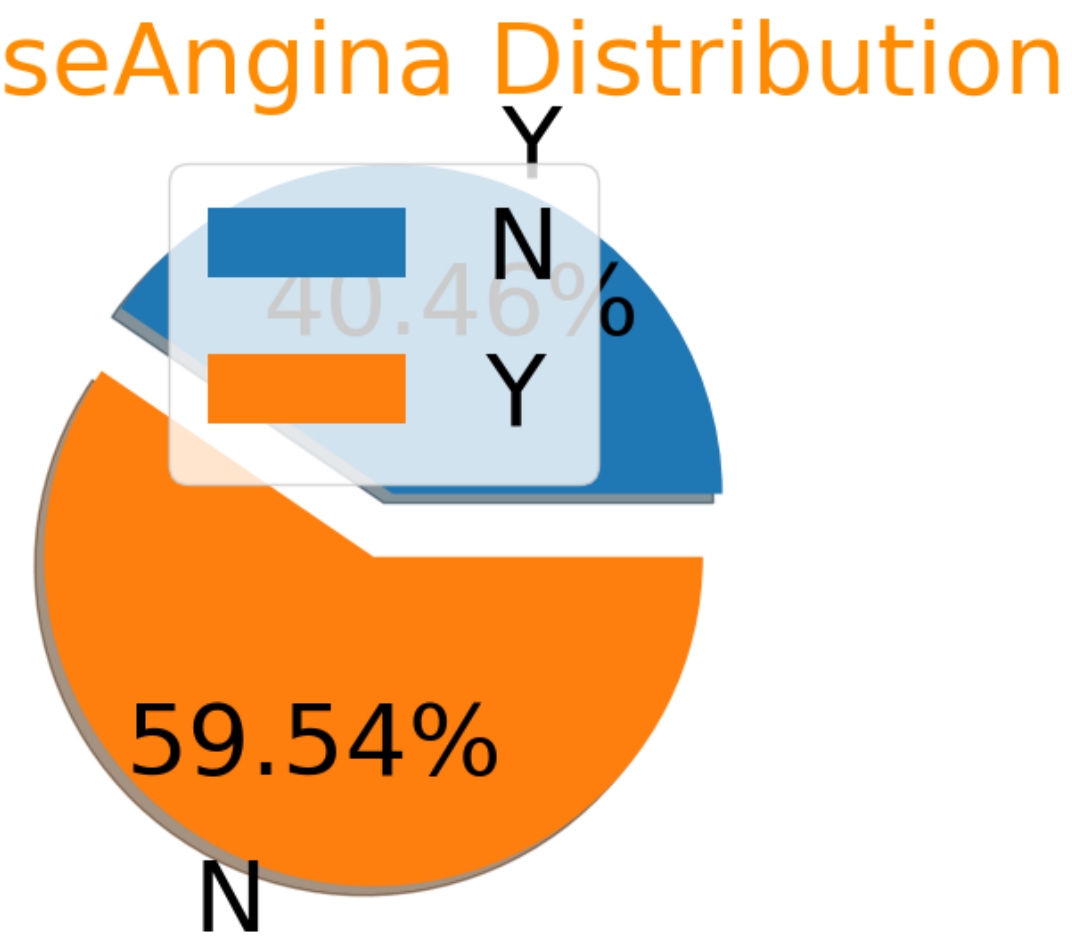




VISUALIZATION OF ST_SLOPE

	Value	Total	Percentage
0	Y	371	40.460000
1	N	546	59.540000





VISUALIZATION OF EXERCISEANGINA

	Value	count	Percentage
0	Flat	459	50.050000
1	Up	395	43.080000
2	Down	63	6.870000

RESULT OF ACCURACY SCORE

Model	Train Accuracy Score	Average Accuracy Score
Logistic Regression Model	84.45	91.35
Support Vector Machines Model	84.46	85.19
Random Forest Classifier Model	83.78	85.14
Multilayer Perceptron Classifier Model	81.08	79.48
Decision Tree Model	81.08	81.16

THANKS FOR LISTENING