Unraveling ML

Uncover the core concepts of Machine Learning

Acknowledgements

A lot of the content is adopted, and modified from the following resources:

- TowardsDataScience Articles
- Medium Blogs
- Kaggle
- SuperDataScience

Why me?

FSM
Smart Manufacturing Simplified

Ural Federal University
named after the first President of Russia B.N.Yeltsin

Machine Learning

Machine Learning Process

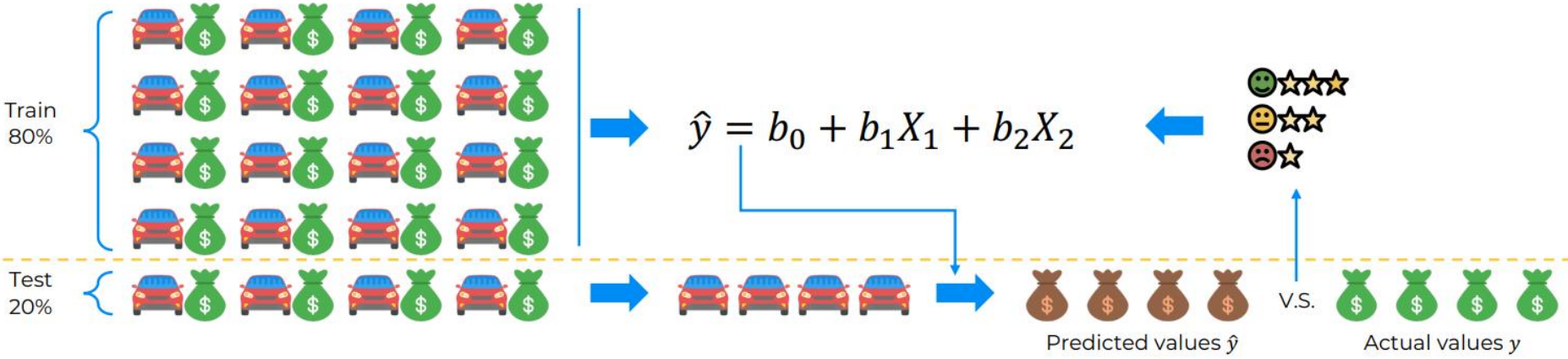
- Import data
- Clean it

- Feature Engineering
- Experimental analysis

- Build a model
- Train that model
- Make predictions

- Calculate performance metrics
- Make a verdict

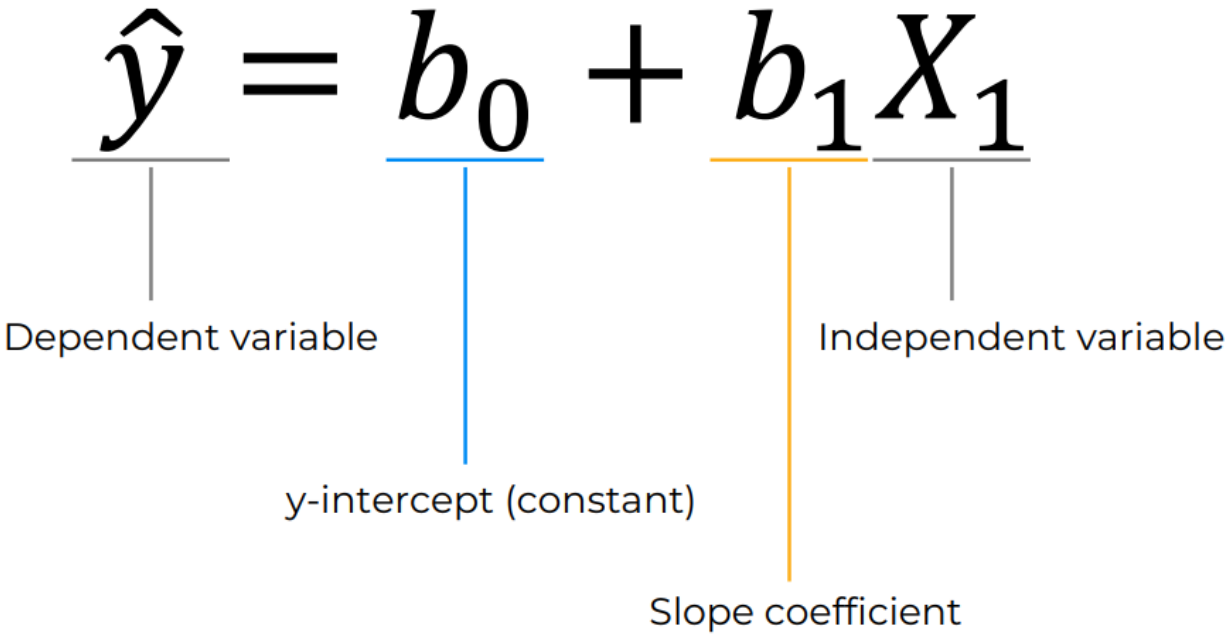
Training Set & Test Set



Exploratory Data Analysis

Regression

Simple Linear Regression



Simple Linear Regression

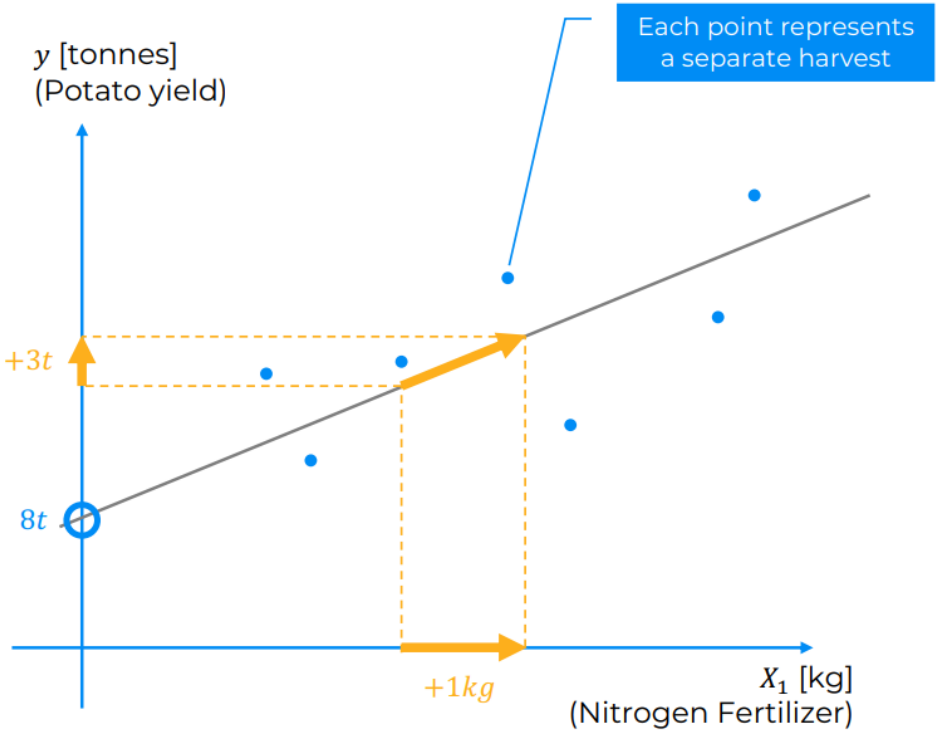
$\sim$

$$\hat{y} = b_0 + b_1 X_1$$

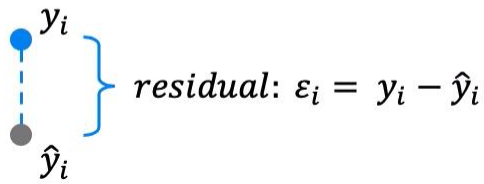
$$Potatoes[t] = b_0 + b_1 \times Fertilizer[kg]$$

$b_0 = 8[t]$

$b_1 = 3[\frac{t}{kg}]$



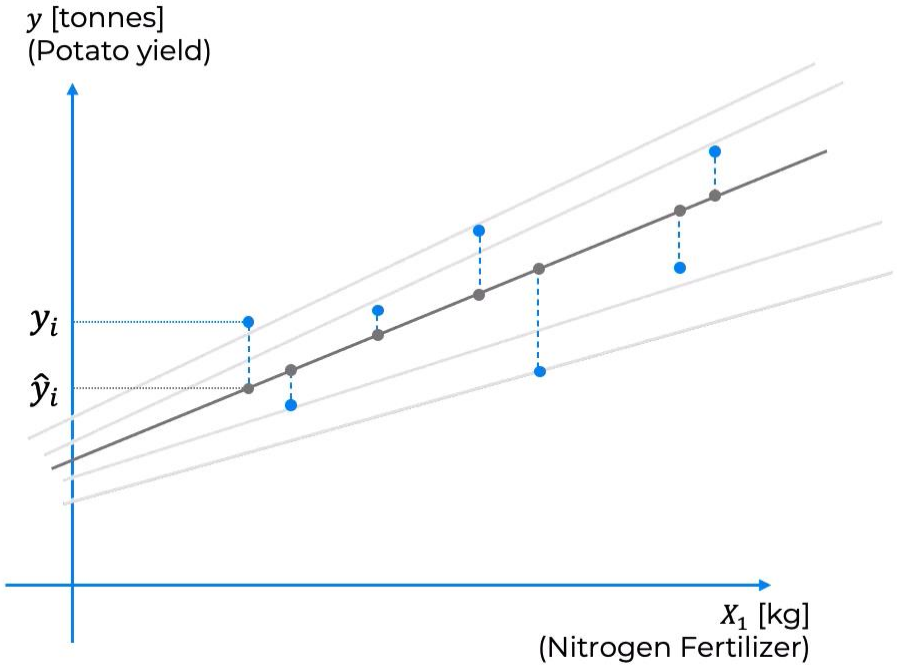
Ordinary Least Squares:



$\hat{y} = b_0 + b_1 X_1$

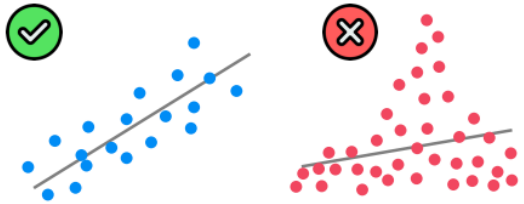
b_0, b_1 such that:

$SUM(y_i - \hat{y}_i)^2$ is minimized

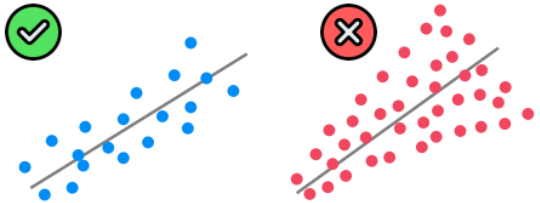


Assumptions of Linear Regression

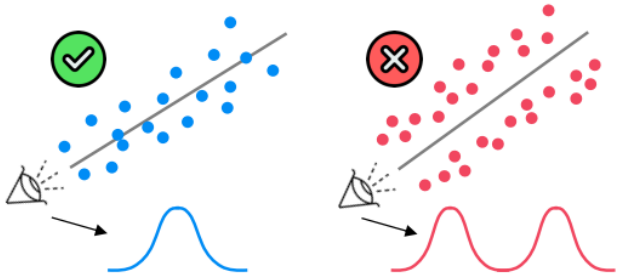
1. Linearity
(Linear relationship between Y and each X)



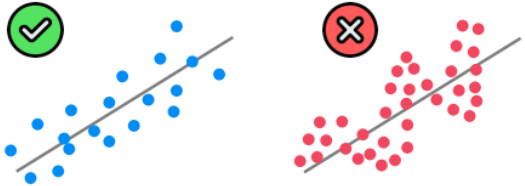
2. Homoscedasticity
(Equal variance)



3. Multivariate Normality
(Normality of error distribution)



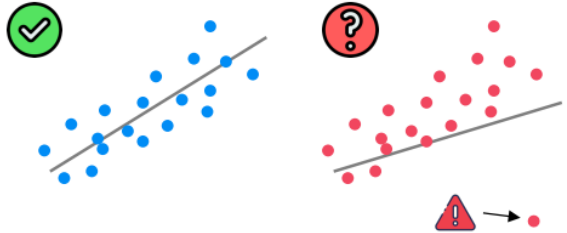
4. Independence
(of observations. Includes “no autocorrelation”)



5. Lack of Multicollinearity
(Predictors are not correlated with each other)

$X_1 \not\sim X_2$ $\quad X_1 \sim X_2$

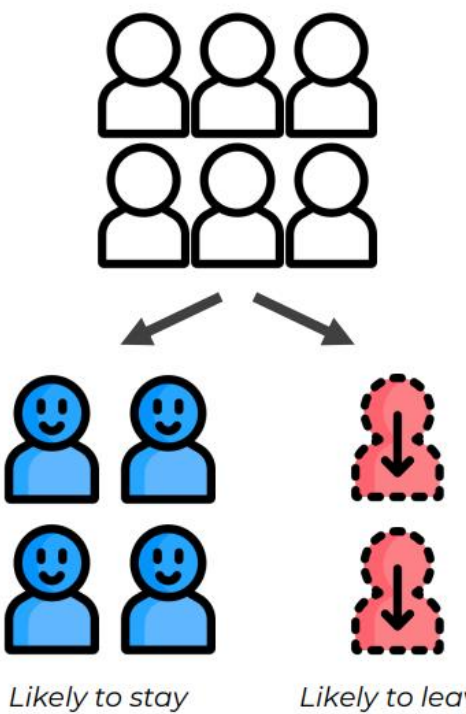
6. The Outlier Check
(This is not an assumption, but an “extra”)

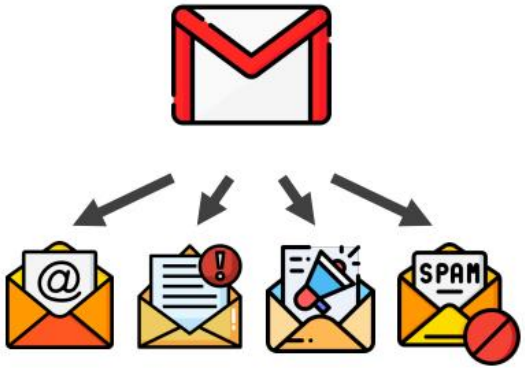


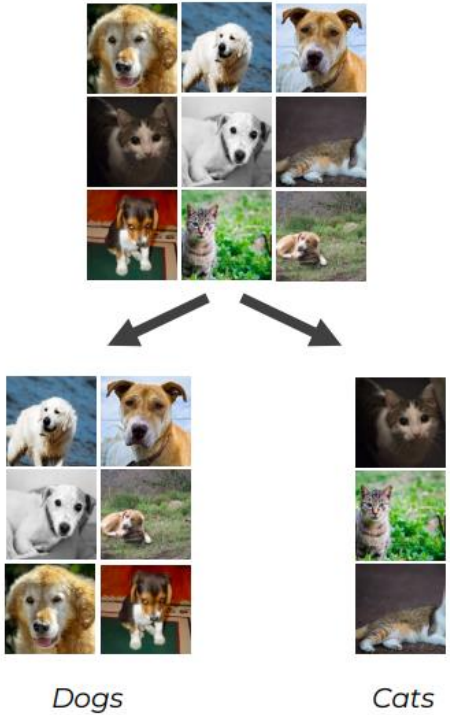
Classification

What is Classification?

identify the category of new observations based on training data







Logistic Regression

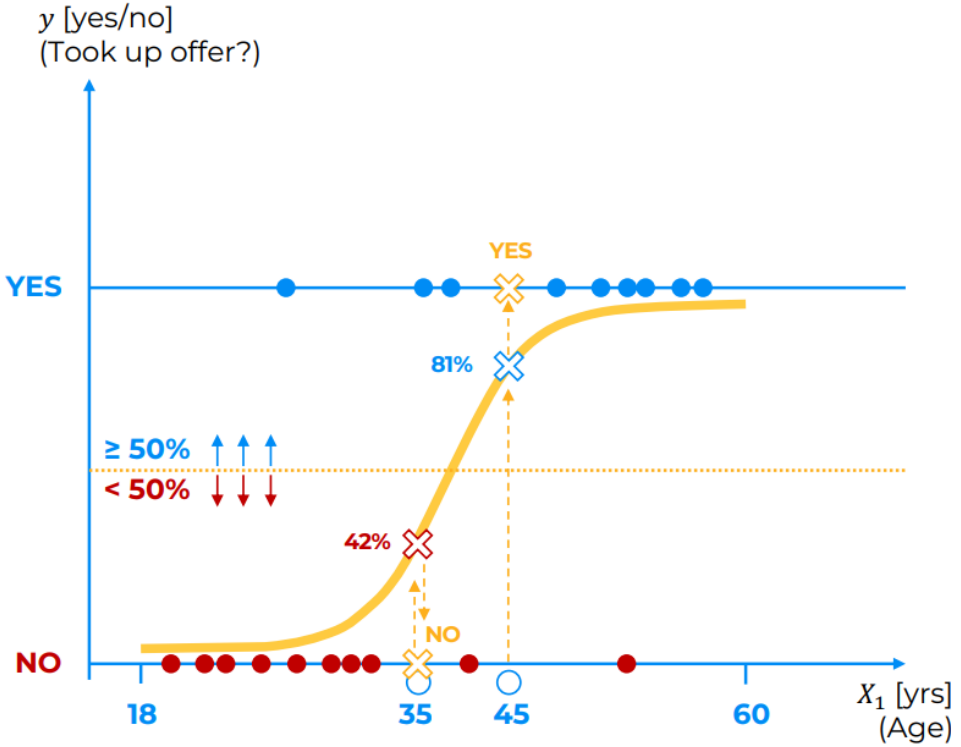
predict a categorical dependent variable from a number of independent variables

~

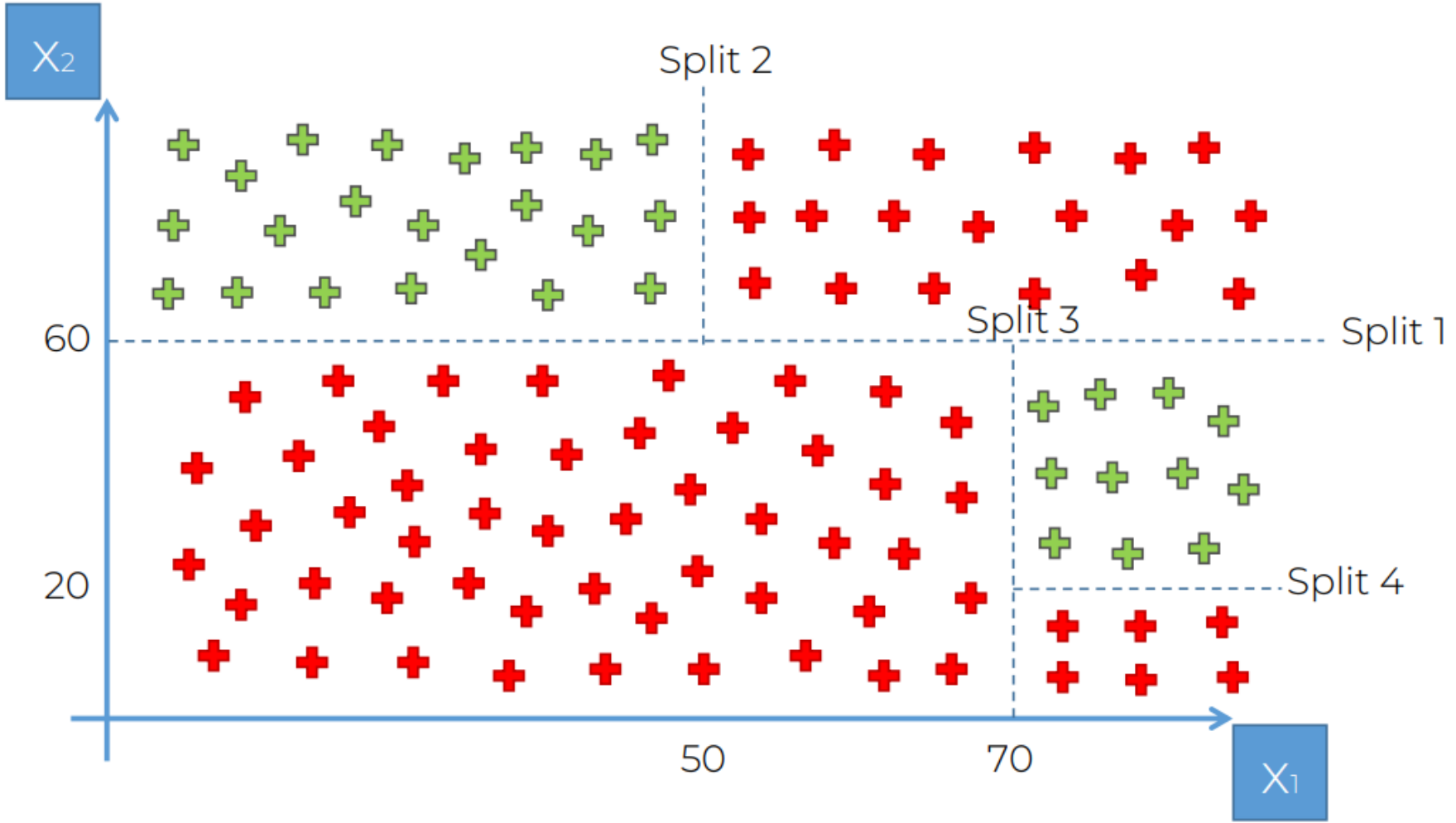
Will purchase health insurance: Yes / No

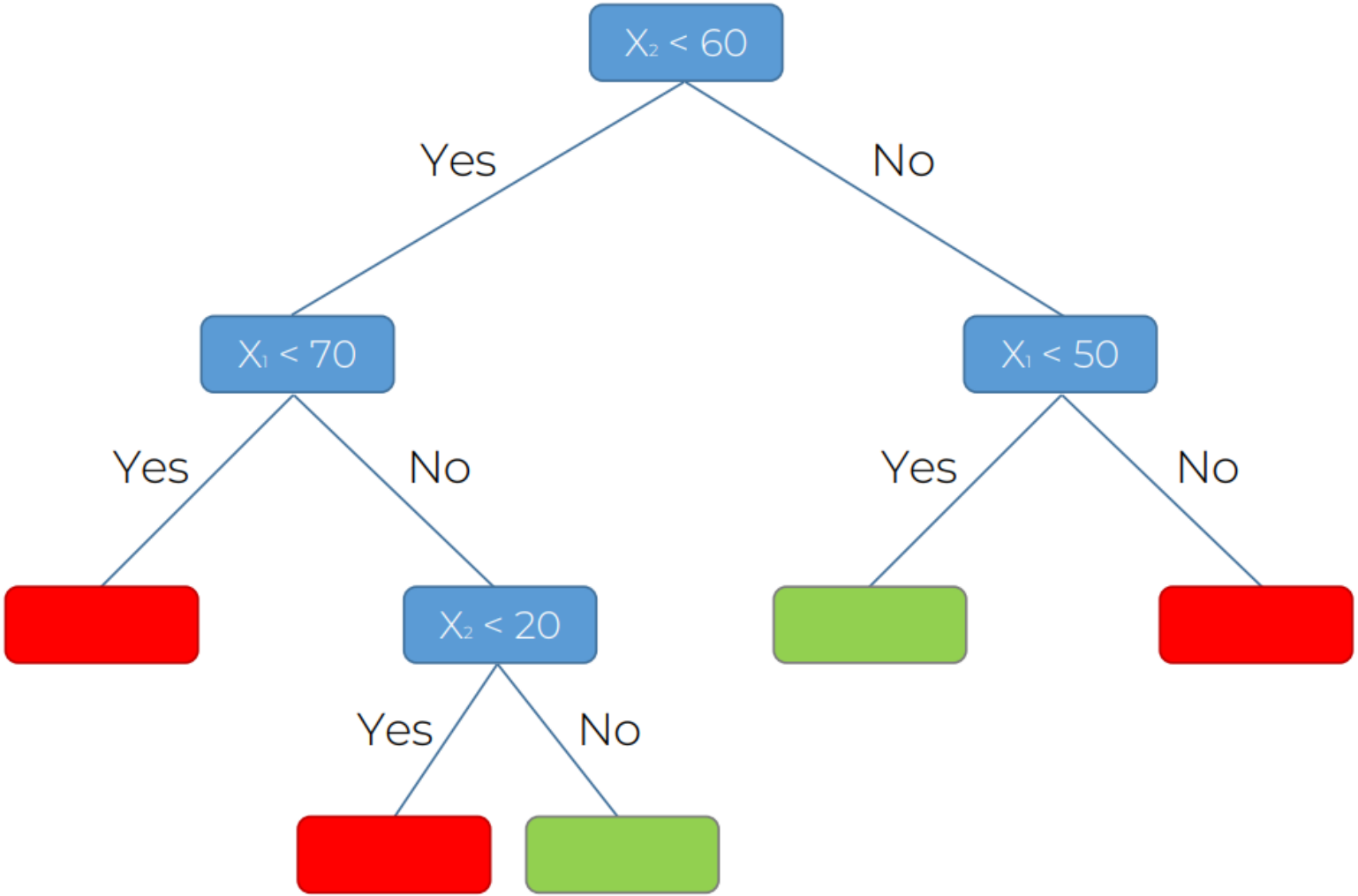
Age

$$\ln \frac{p}{1-p} = b_0 + b_1 X_1$$



Decision Tree


X2
Split 2
60
Split 3
Split 1
20
Split 4
50
70
X1


$X_2 < 60$
Yes
No
$X_1 < 70$
$X_1 < 50$
Yes
No
Yes
No
$X_2 < 20$
Yes
No

Reach out to me

[Keivalya Pandya | LinkedIn]
for connecting and engaging on professional level

[Keivalya Pandya (topmate.io)]
for technical assistance

[Keivalya Pandya (@keivalya)]
for non-technical vlogs