# MACHINE LEARNING FOR BUSY PEOPLE

*10/12/2019 @* AI & SMART TECH


Fabien VAUCHELLES

ZELROS

zelros.com / fabien.vauchelles.@zelros.com / @fabienv


http://bit.ly/ml-aismarttech

# 1

# INTRODUCTION

I WILL TELL YOU THE TRUTH

# JACK KNEW HE WOULD DIE

JACK KNEW HIS FUTURE

# LIFE DECISION PATH

JACK

# LIFE DECISION PATH

JACK

Men

Women

# LIFE DECISION PATH

JACK

Men

Women

Adult

Child

# LIFE DECISION PATH

JACK

Men

Women

Adult

Child

1st class

3th class

# LIFE DECISION PATH

JACK

Men

Women

Adult

Child

1st class

3th class

GAME OVER

70%

# BACK TO MACHINE LEARNING

If you know the passengers list:

- Gender
- Age
- Ticket class
- Does he survived ?

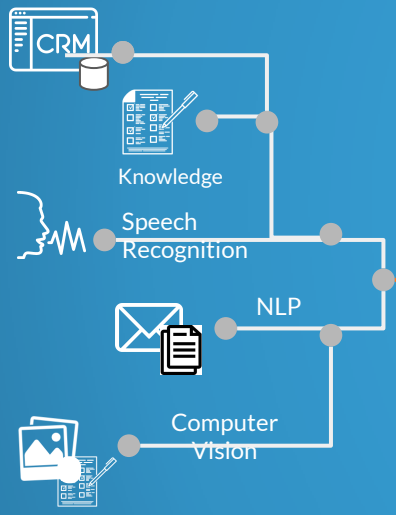
You can create a **Decision Tree** ...

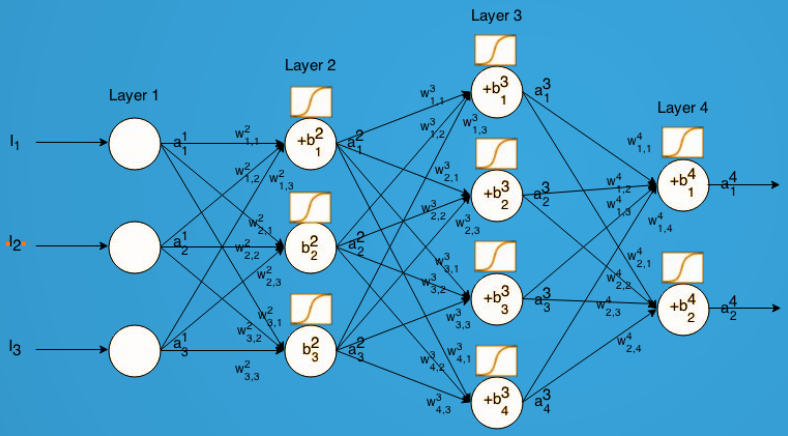
... for this **Supervised** Problem !

REAL LIFE IS DIFFERENT...

# KEEP COOL AND USE MACHINE LEARNING

# ZELROS // FABIEN VAUCHELLES



**Insurer Data**

**Insurance ML / Predictive models**

**AI coach**

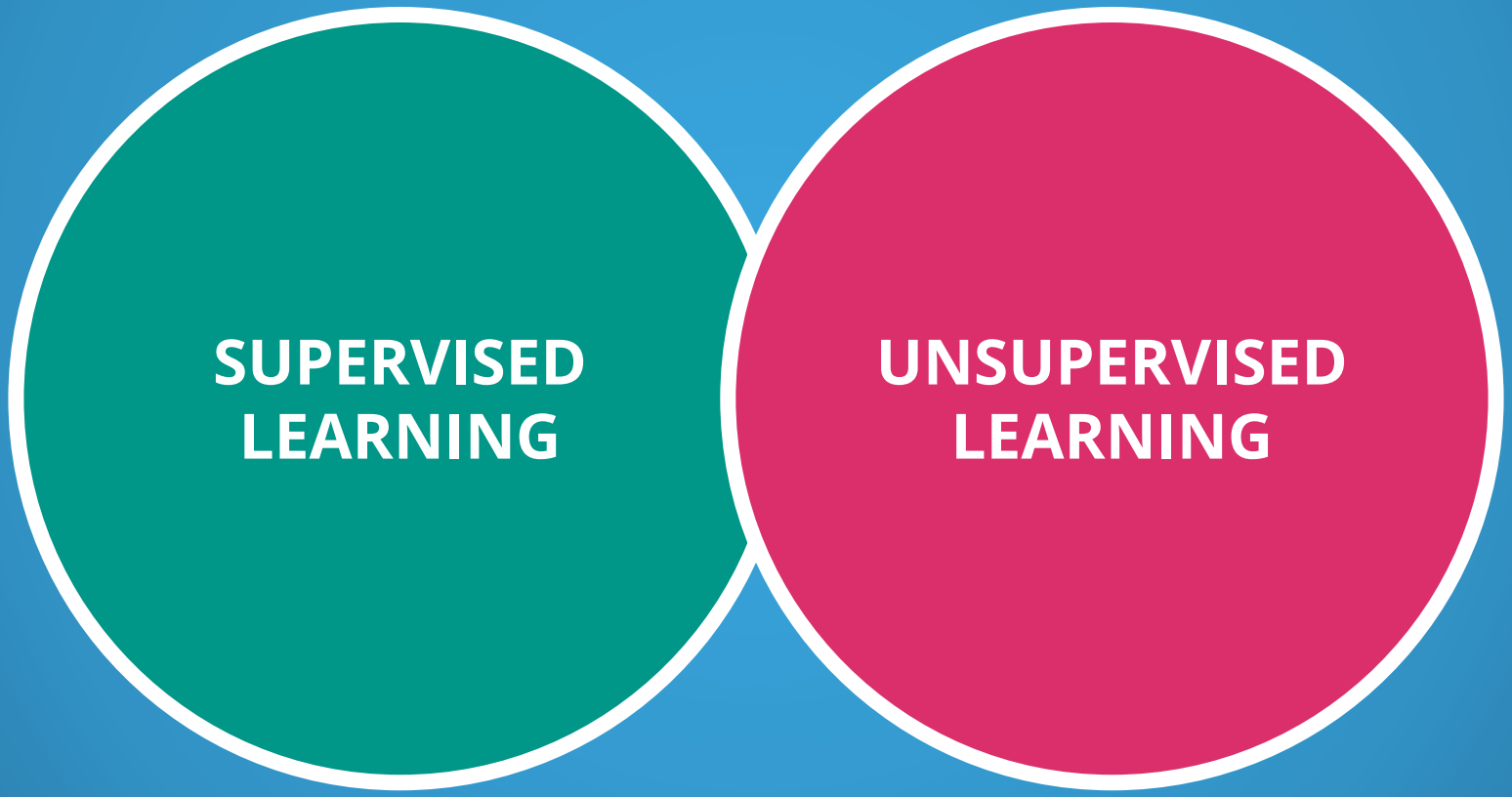
**Insurance Advisor**

# What is Machine Learning

ML SOLVE PROBLEMS WHICH ARE UNSOLVABLE BY CONVENTIONAL METHODS

# What is the type of problem

# TYPE OF PROBLEM



# SUPERVISED LEARNING

JACK

Men

Women

Adult

Child

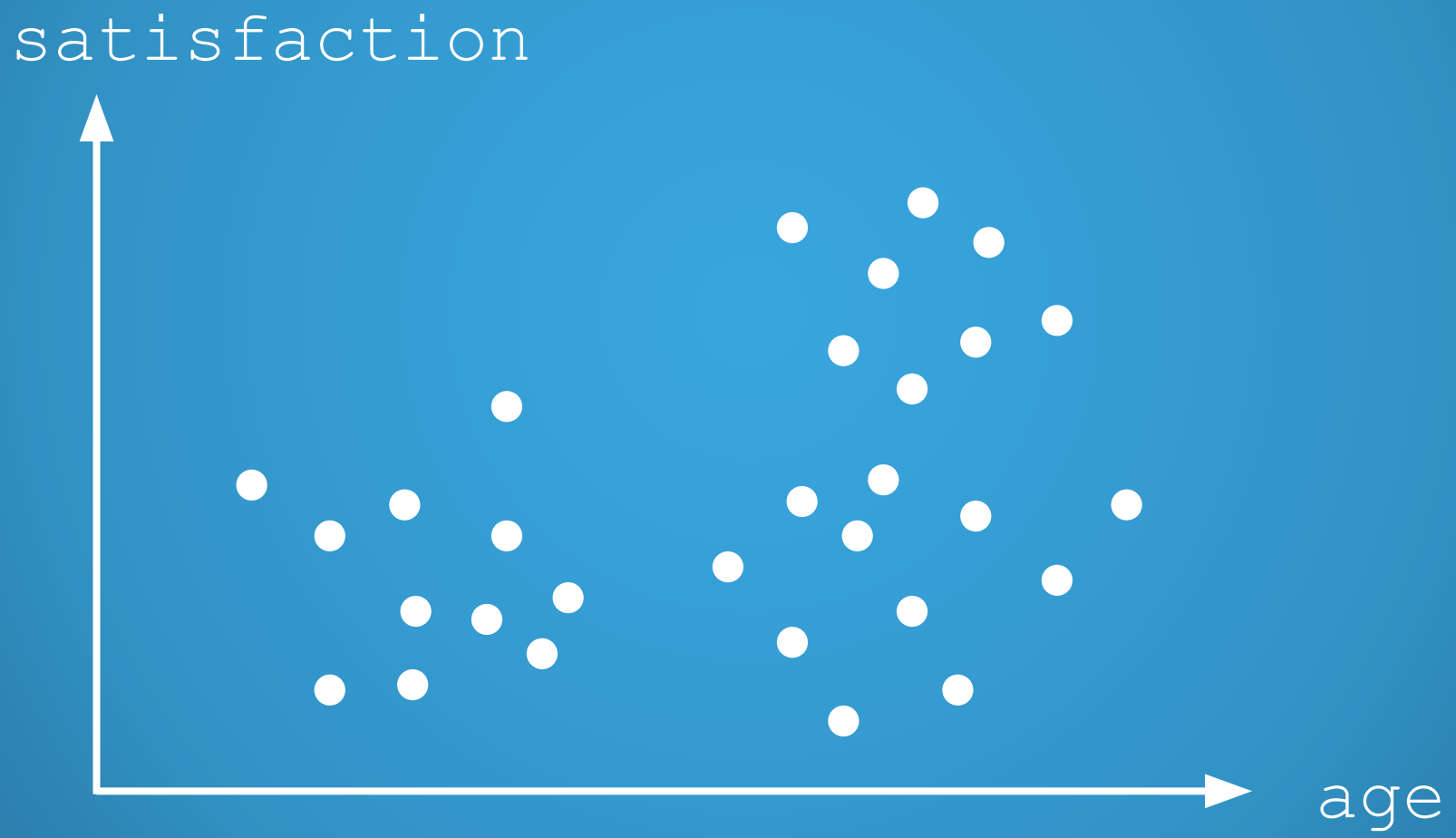
1st class

3th class

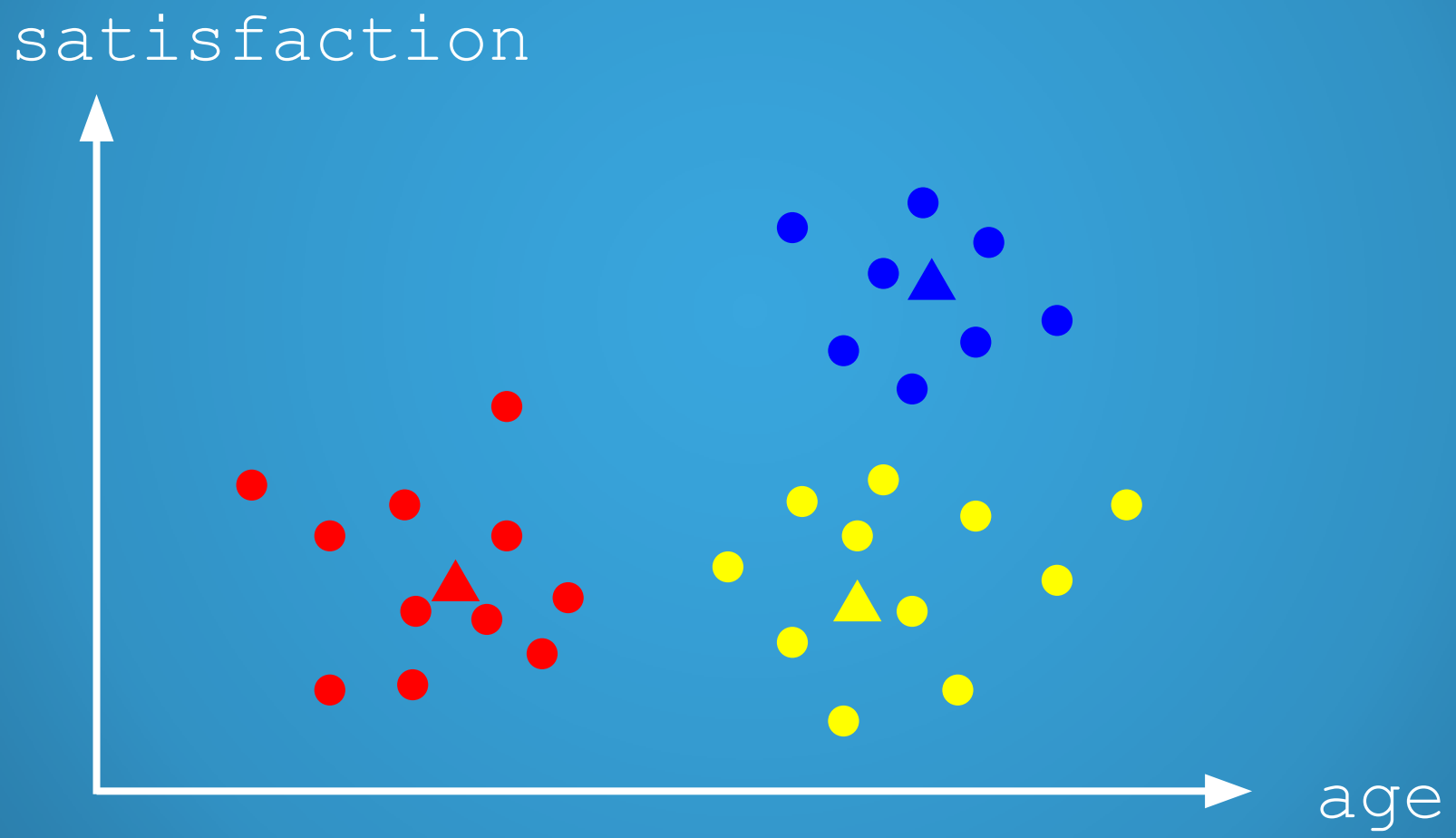
GAME OVER

70%

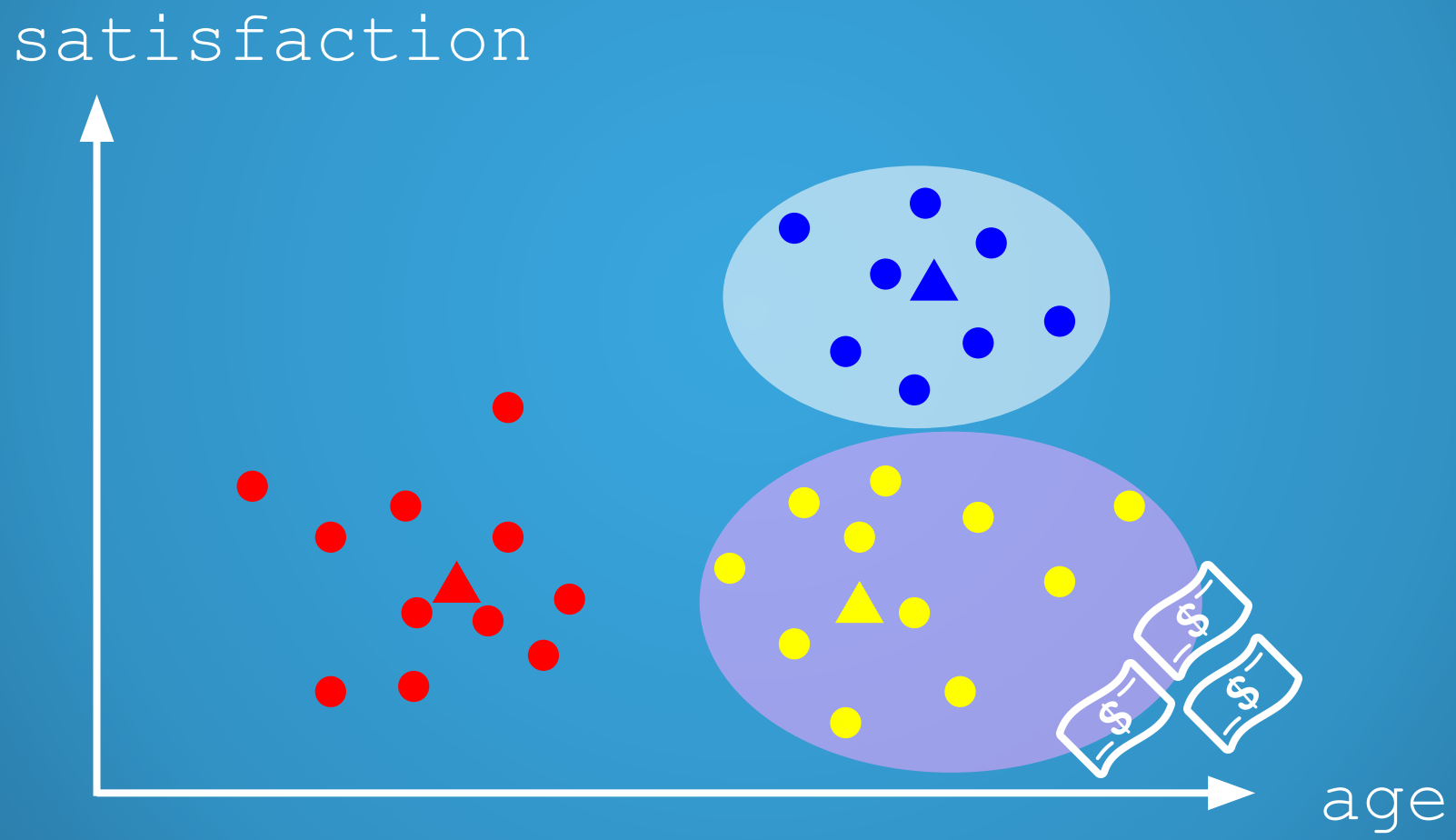
# UNSUPERVISED LEARNING



# UNSUPERVISED LEARNING



# UNSUPERVISED LEARNING



# What is the goal of ML algorithms

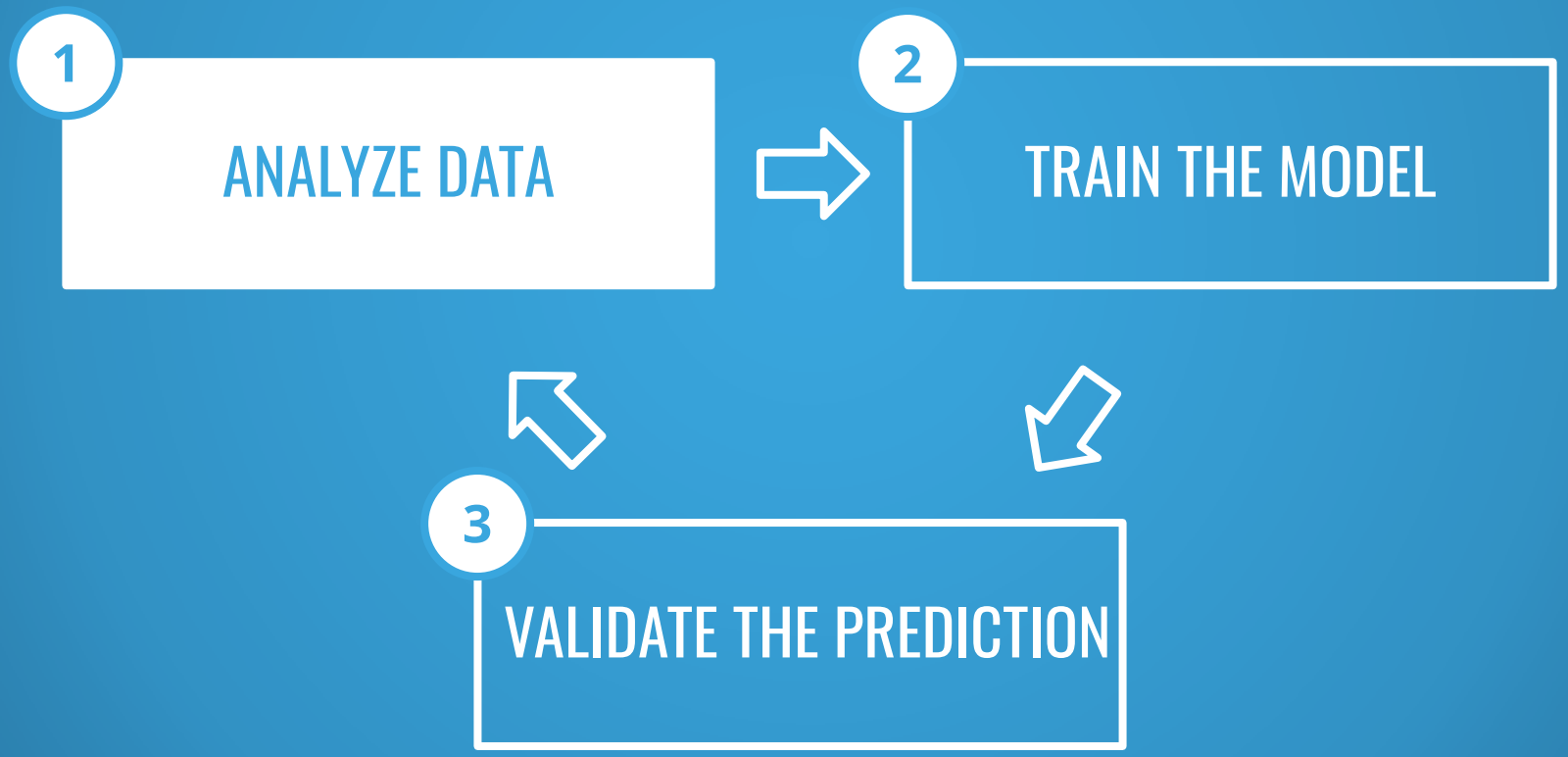
# MINIMIZE THE ERROR

# How to start a ML problem

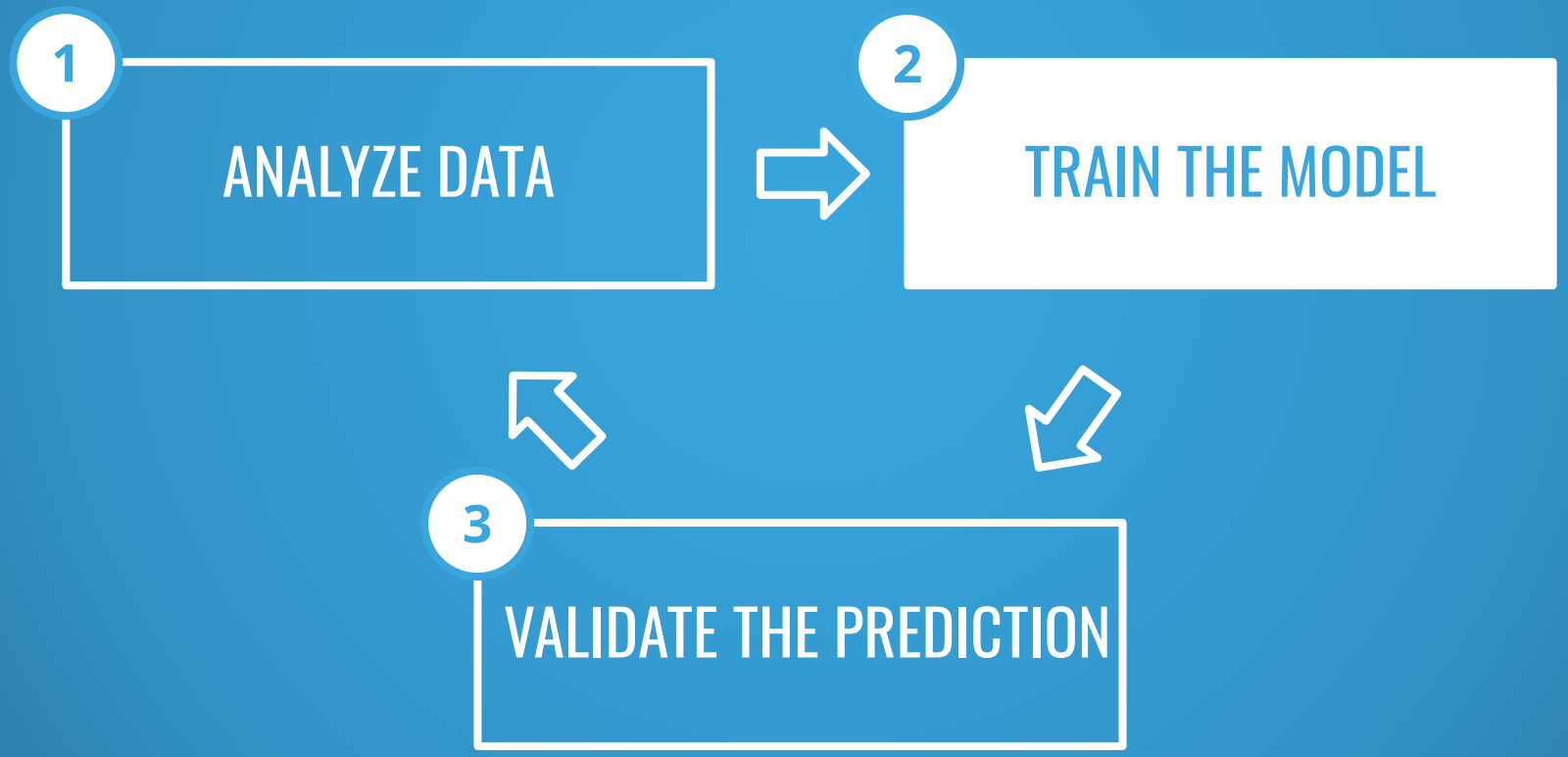
AS DEVELOP

WE USE TEST DRIVEN DEVELOPMENT

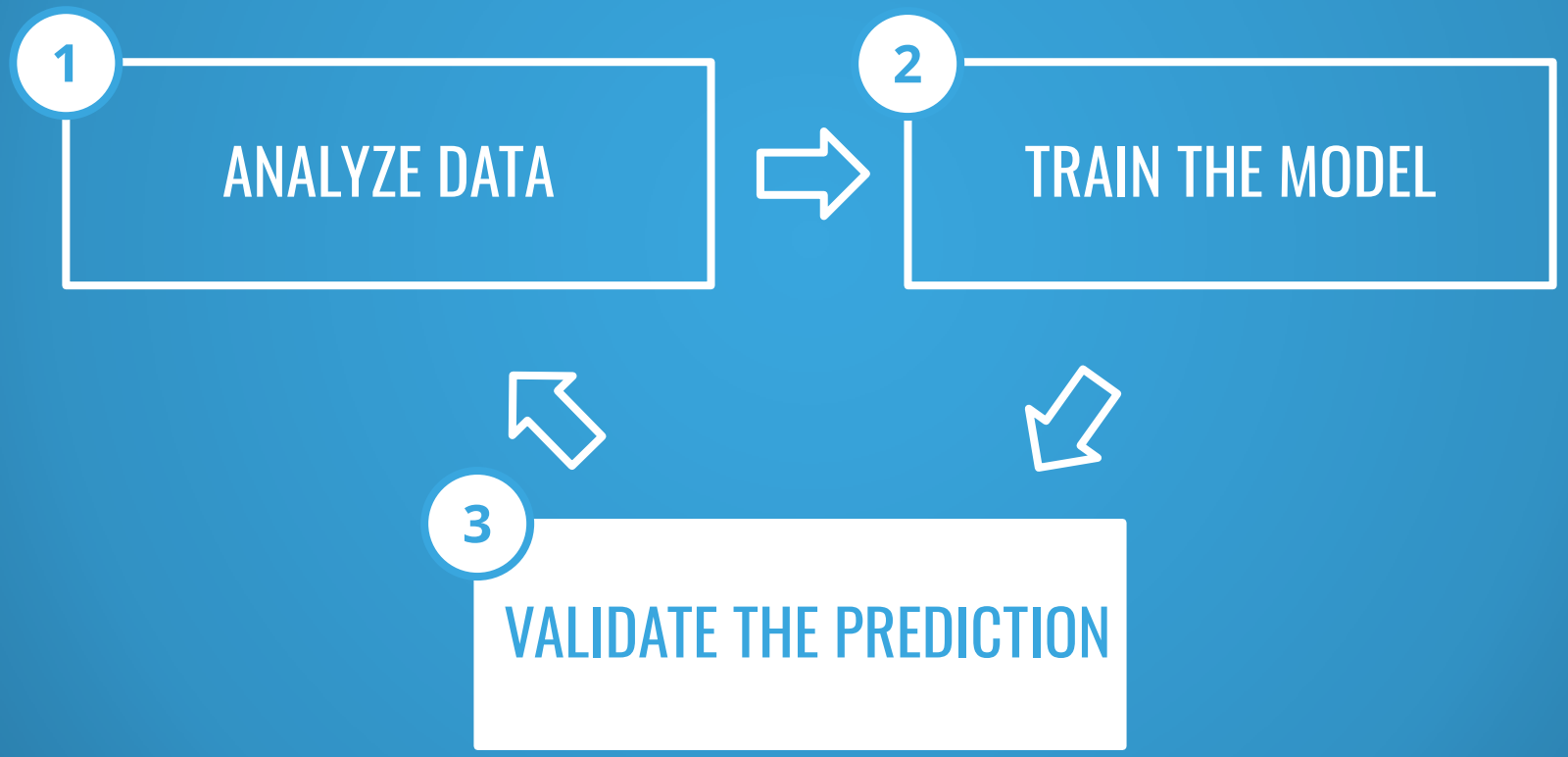
# THE PROCESS OF DATA SCIENCE



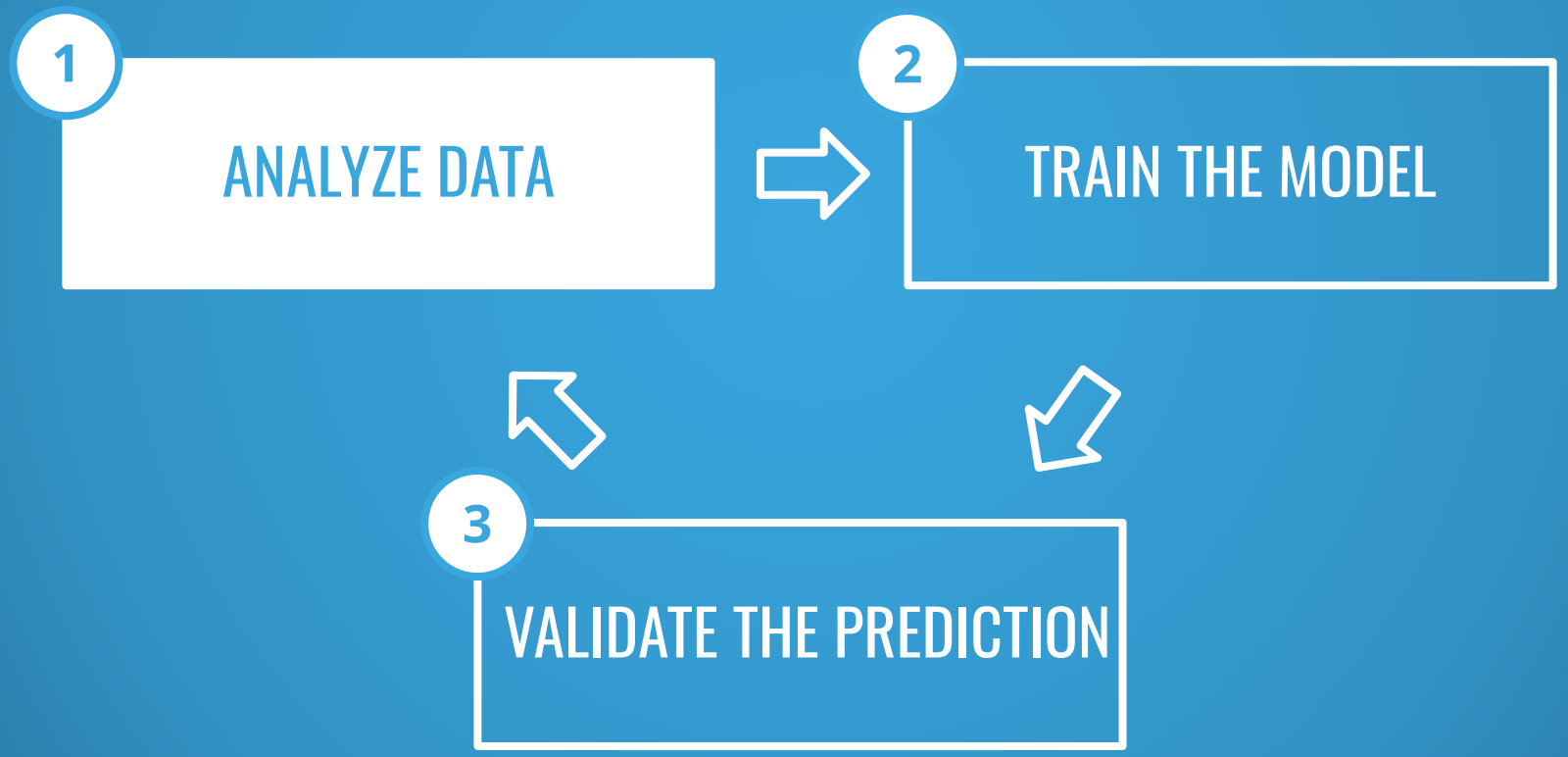
# THE PROCESS OF DATA SCIENCE



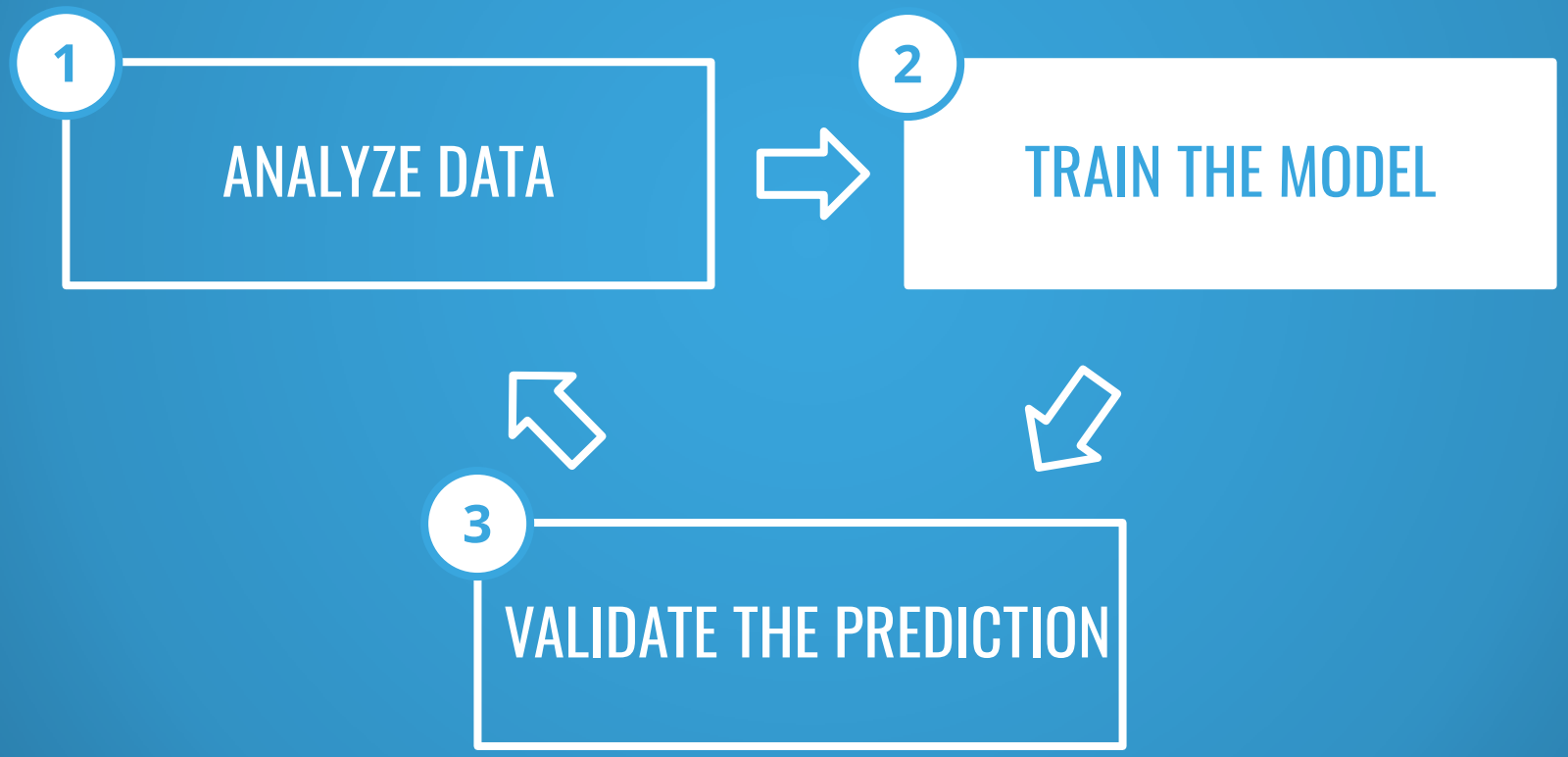
# THE PROCESS OF DATA SCIENCE



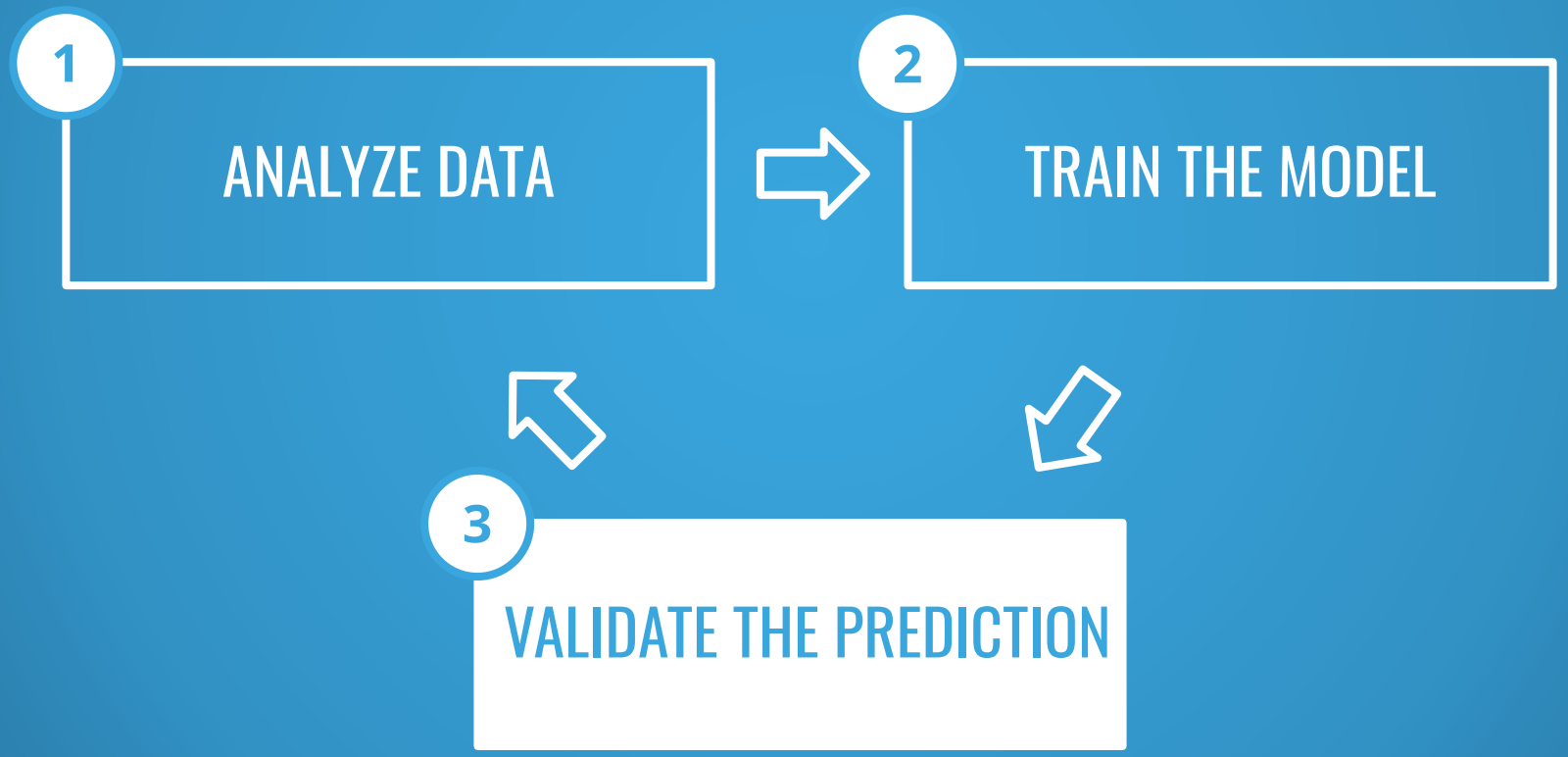
# THE PROCESS OF DATA SCIENCE



# THE PROCESS OF DATA SCIENCE



# THE PROCESS OF DATA SCIENCE



# 2

# ANALYZE DATA

# Do you speak Data

# DATA LANGUAGE

*Dataset*

| NAME | AGE | CLASS | DIED ? |
|---|---|---|---|
| John | 23 | 3 | Yes |
| Marry | 31 | 1 | No |
| Henry | 23 | 2 | Yes |
| Nicolas | 41 | 1 | No |
| Anna | 18 | 3 | Yes |

# DATA LANGUAGE

*Feature*

| NAME | AGE | CLASS | DIED ? |
|---|---|---|---|
| John | 23 | 3 | Yes |
| Marry | 31 | 1 | No |
| Henry | 23 | 2 | Yes |
| Nicolas | 41 | 1 | No |
| Anna | 18 | 3 | Yes |

# DATA LANGUAGE

*Target*

| NAME | AGE | CLASS | DIED ? |
|---|---|---|---|
| John | 23 | 3 | Yes |
| Marry | 31 | 1 | No |
| Henry | 23 | 2 | Yes |
| Nicolas | 41 | 1 | No |
| Anna | 18 | 3 | Yes |

# DATA LANGUAGE

| NAME | AGE | CLASS | DIED ? |
|---|---|---|---|
| John | 23 | 3 | Yes |
| Marry | 31 | 1 | No |
| Henry | 23 | 2 | Yes |
| Nicolas | 41 | 1 | No |
| Anna | 18 | 3 | Yes |

*Sample*

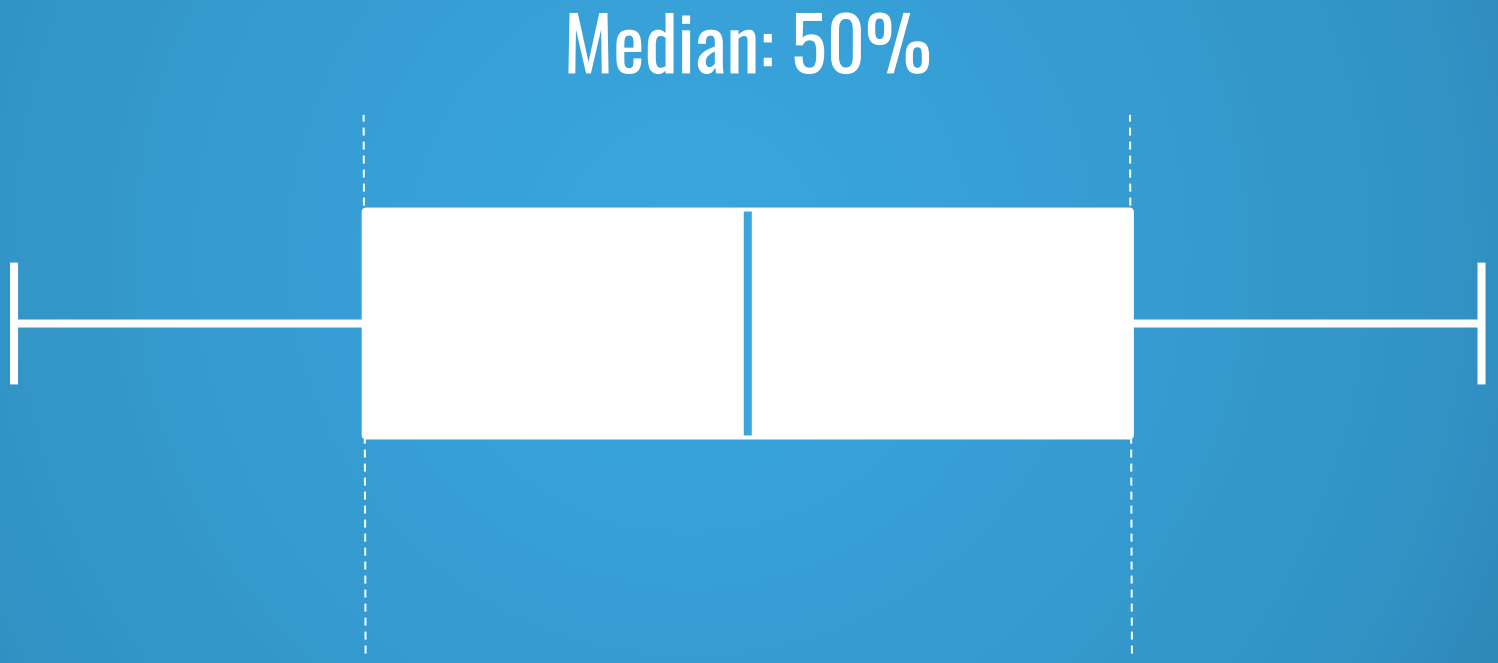
DEMO TIME !

# How to visualize a feature

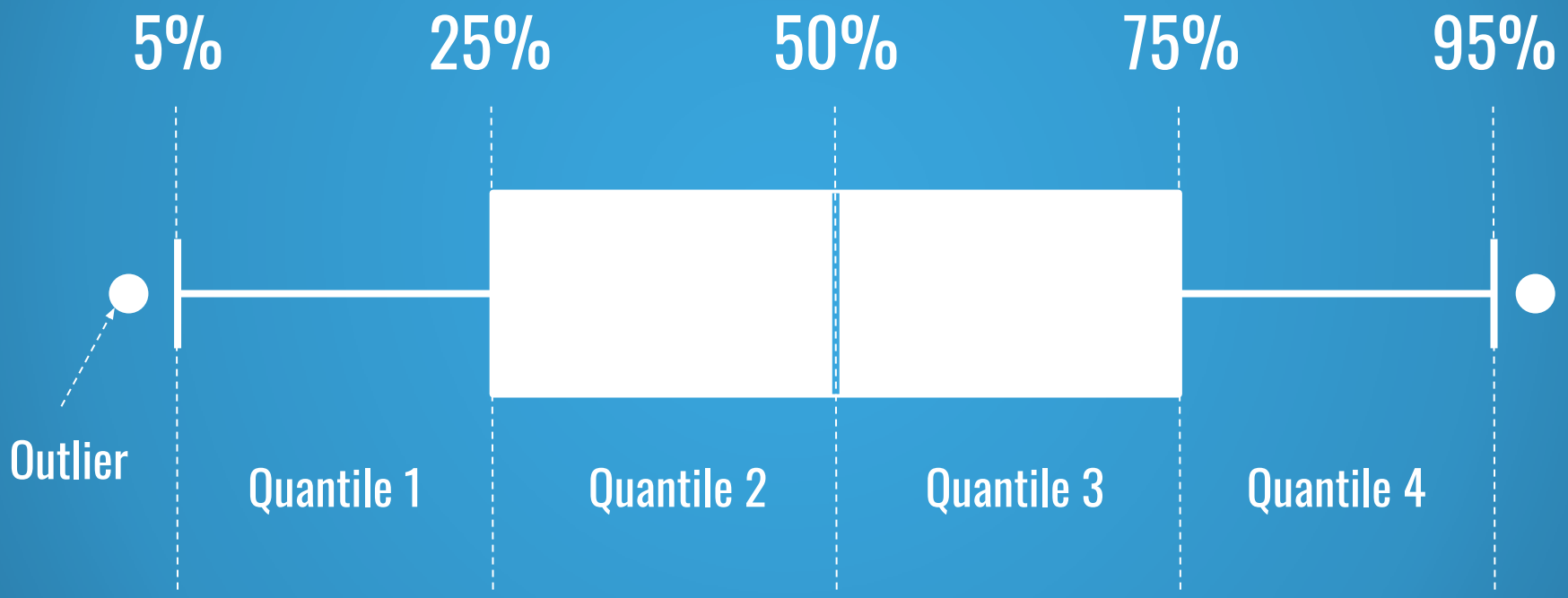
# EXCEL WON’T HELP YOU ON BIG DATA

YOU LOVE STATISTIC METHODS

# BOX PLOT



# BOX PLOT


5%
25%
50%
75%
95%
Outlier
Quantile 1
Quantile 2
Quantile 3
Quantile 4

SENORGIF.COM
DEMO
TIME !

# Why do we care of missing data

# ALGORITHMS DON’T LIKE MISSING DATA

# MISSING VALUES

10

**Nan**

20

**Nan**

30

# MISSING VALUES

| | | |
|---|---|---|
| 10 | | |
| ~~Nan~~ | ⇨ | 10 |
| 20 | | 20 |
| ~~Nan~~ | ⇨ | 30 |
| 30 | | |

BAD !!!

# MISSING VALUES

10

**Nan**

20

**Nan**

30

# MISSING VALUES

| | | |
|---|---|---|
| 10 | | 10 |
| **Nan** | ⇨ | **0** |
| 20 | | 20 |
| **Nan** | ⇨ | **0** |
| 30 | | 30 |

BAD !!!

# MISSING VALUES

Fill empty value with median: 20

| | | |
|---|---|---|
| 10 | | 10 |
| **Nan** | ⇨ | **20** |
| 20 | | 20 |
| **Nan** | ⇨ | **20** |
| 30 | | 30 |

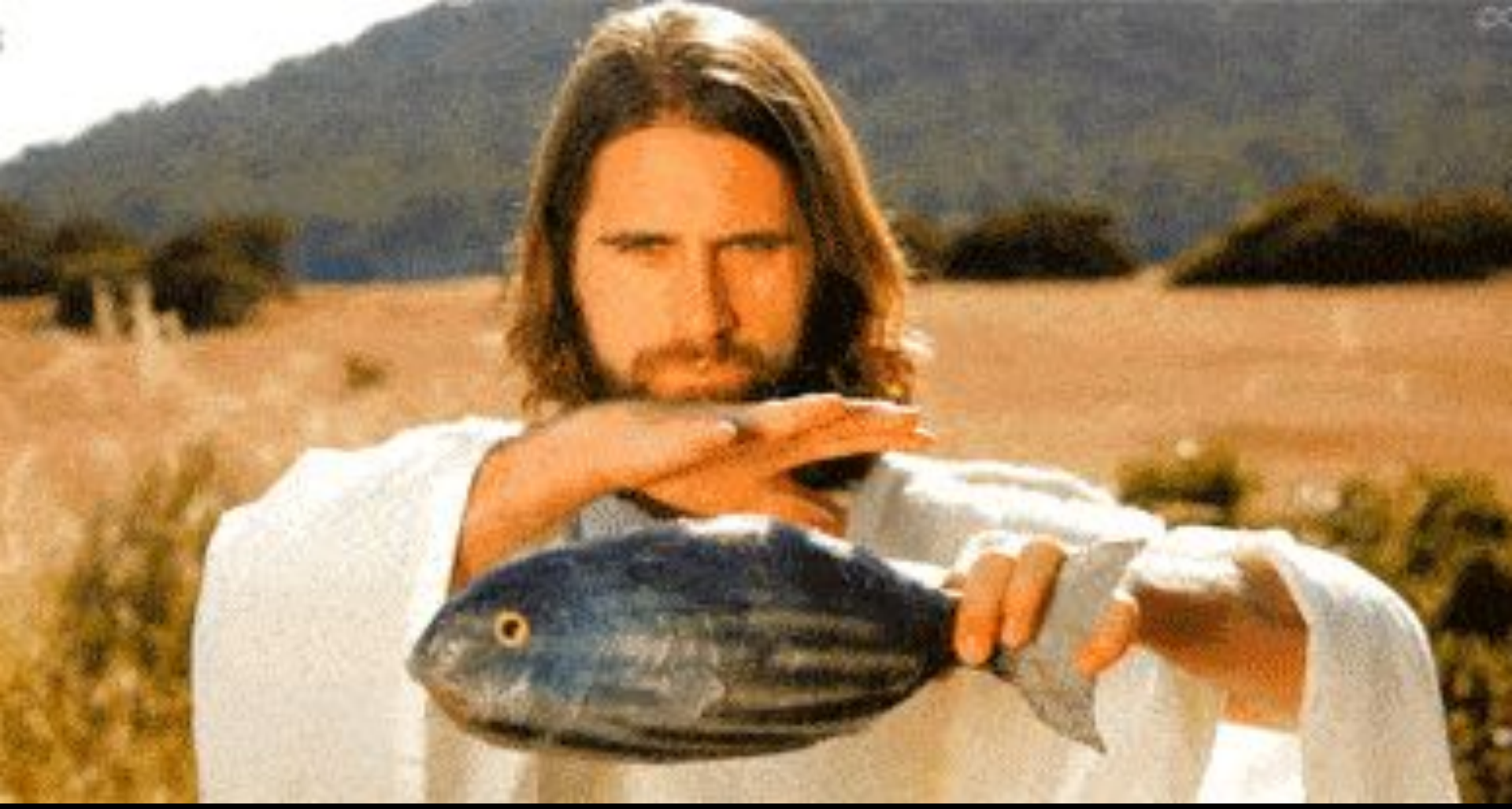
GOOD !!!

DEMO TIME !

# 3

# CREATE FEATURES

# Why do we create Artificial Features



# ARTIFICIAL FEATURES HELP ALGORITHMS TO HAVE BETTER PREDICTION

# UNDERSTAND CATEGORIES

| NAME | GENDER |
|---|---|
| John | male |
| Marry | female |
| Henry | male |
| Nicolas | male |
| Anna | female |

I DON'T UNDERSTAND TEXT !

# UNDERSTAND CATEGORIES

| NAME | GENDER |
|---|---|
| John | male |
| Marry | female |
| Henry | male |
| Nicolas | male |
| Anna | female |

| NAME | GENDER |
|---|---|
| John | 1 |
| Marry | 2 |
| Henry | 1 |
| Nicolas | 1 |
| Anna | 2 |

LabelEncoder

2 > 1 !!!

# UNDERSTAND CATEGORIES

| NAME | GENDER |
|---|---|
| John | male |
| Marry | female |
| Henry | male |
| Nicolas | male |
| Anna | female |

| NAME | GENDER_MALE | GENDER_FEMALE |
|---|---|---|
| John | 1 | 0 |
| Marry | 0 | 1 |
| Henry | 1 | 0 |
| Nicolas | 1 | 0 |
| Anna | 0 | 1 |

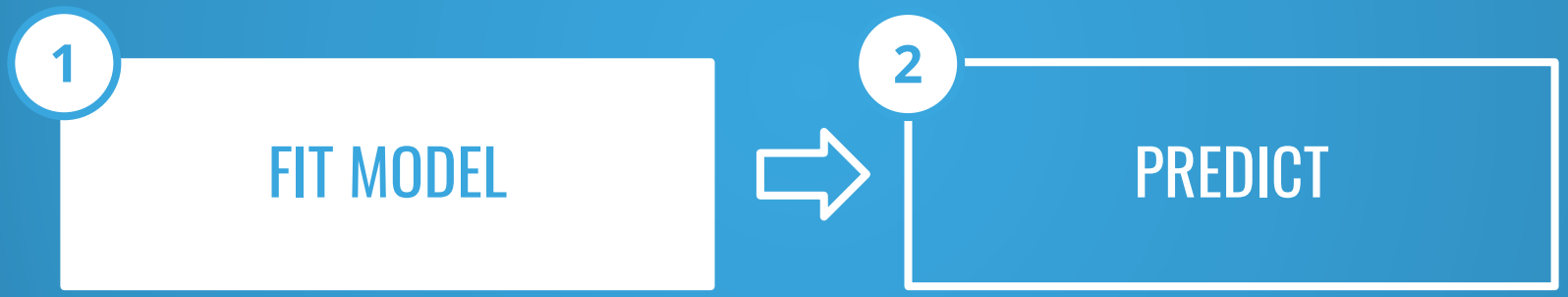
OneHotEncoder

DEMO TIME !

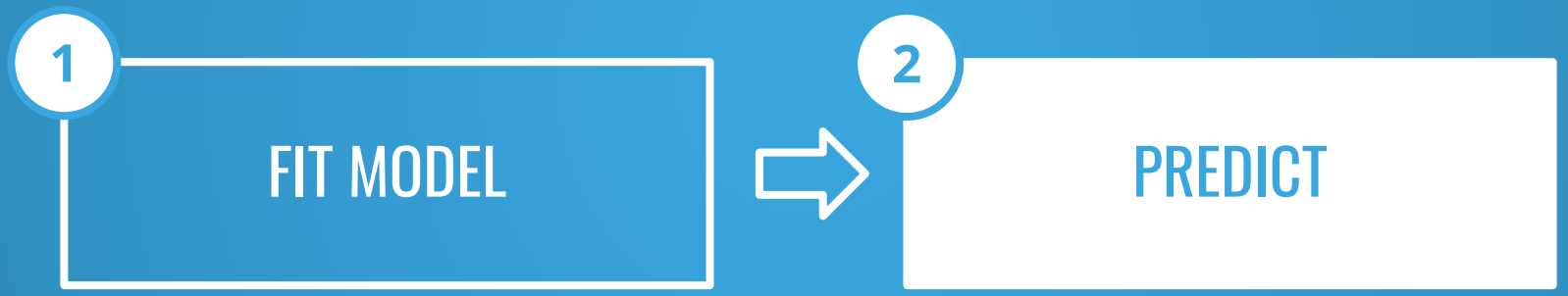
# 4

# TRAIN THE MODEL

# FIT & PREDICT



# FIT & PREDICT



# What can we predict

# CLASSIFICATION

Do we have survived on Titanic ?

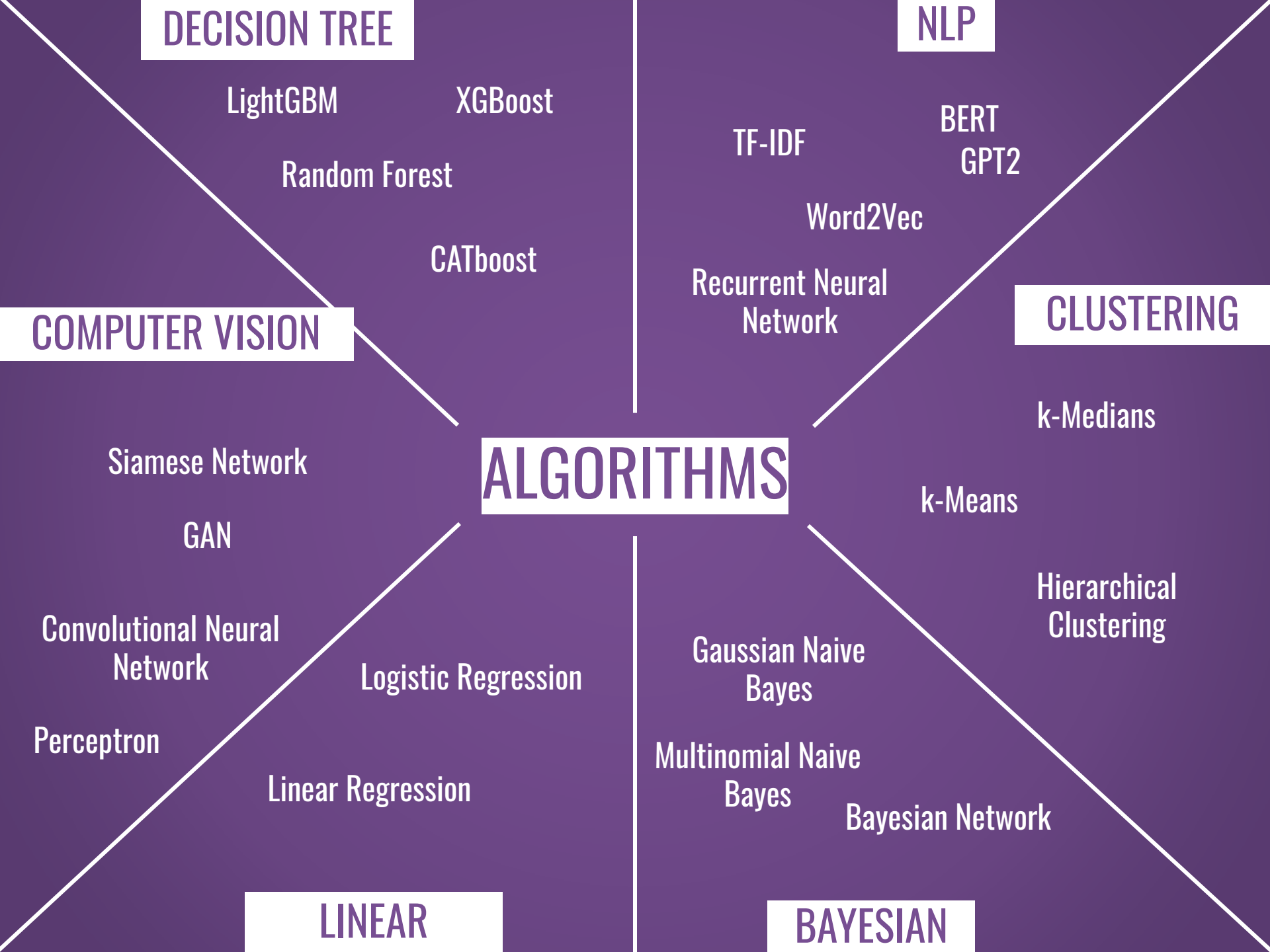
| NAME | AGE | CLASS | DIED ? |
|---|---|---|---|
| John | 23 | 3 | Yes |
| Marry | 31 | 1 | No |
| Henry | 23 | 2 | Yes |
| Nicolas | 41 | 1 | No |
| Anna | 18 | 3 | ? |

# REGRESSION

What is the price of the ticket ?

| NAME | AGE | CLASS | FARE |
|---|---|---|---|
| John | 23 | 3 | 71 |
| Marry | 31 | 1 | 8 |
| Henry | 23 | 2 | 53 |
| Nicolas | 41 | 1 | 7 |
| Anna | 18 | 3 | ? |

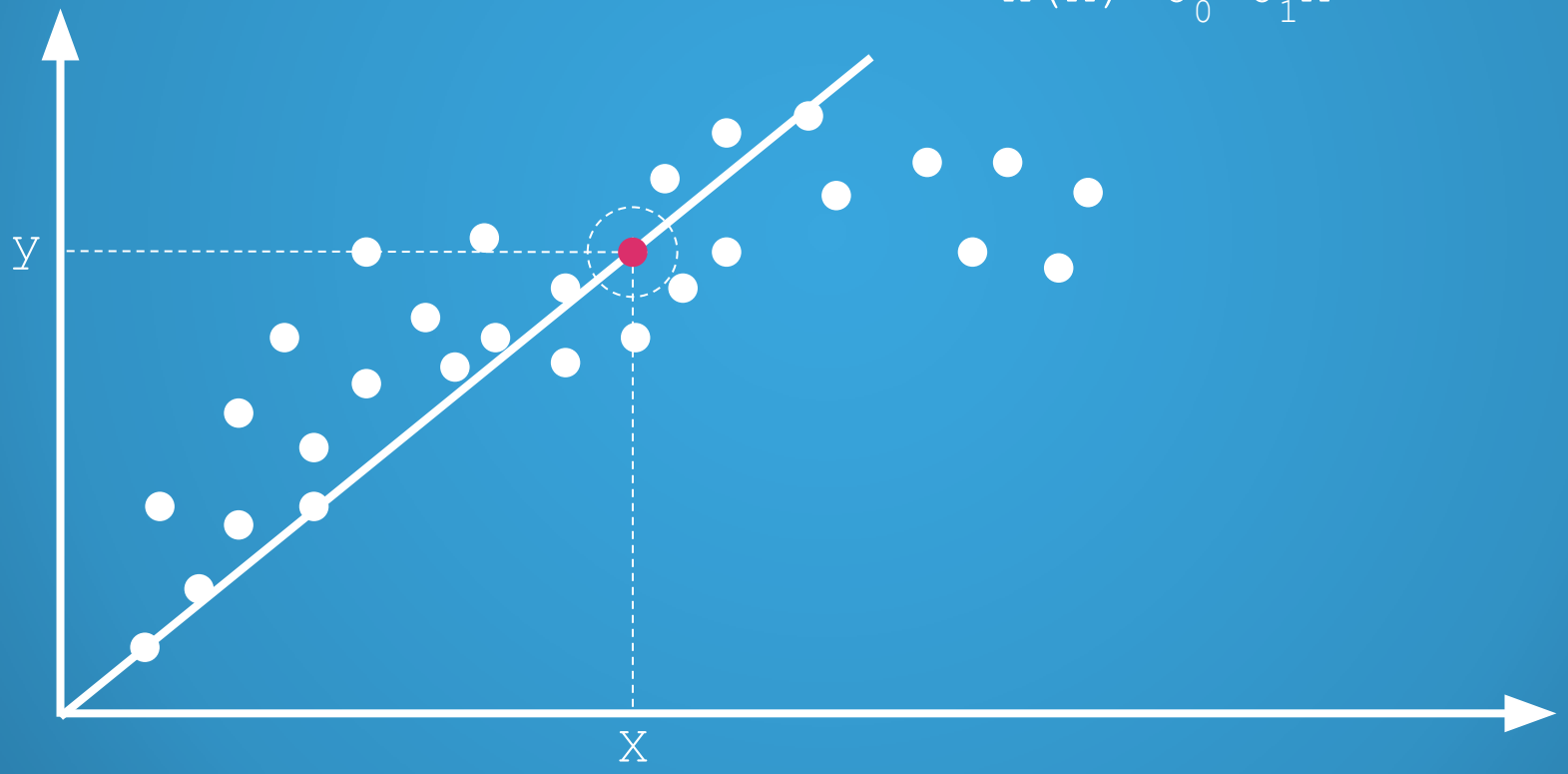
# What algorithm can we choose


DECISION TREE
LightGBM
XGBoost
Random Forest
CATboost
NLP
TF-IDF
BERT
GPT2
Word2Vec
Recurrent Neural Network
COMPUTER VISION
CLUSTERING
ALGORITHMS
Siamese Network
GAN
Convolutional Neural Network
Perceptron
Logistic Regression
Linear Regression
k-Medians
k-Means
Hierarchical Clustering
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network
LINEAR
BAYESIAN

# What is Linear Regression

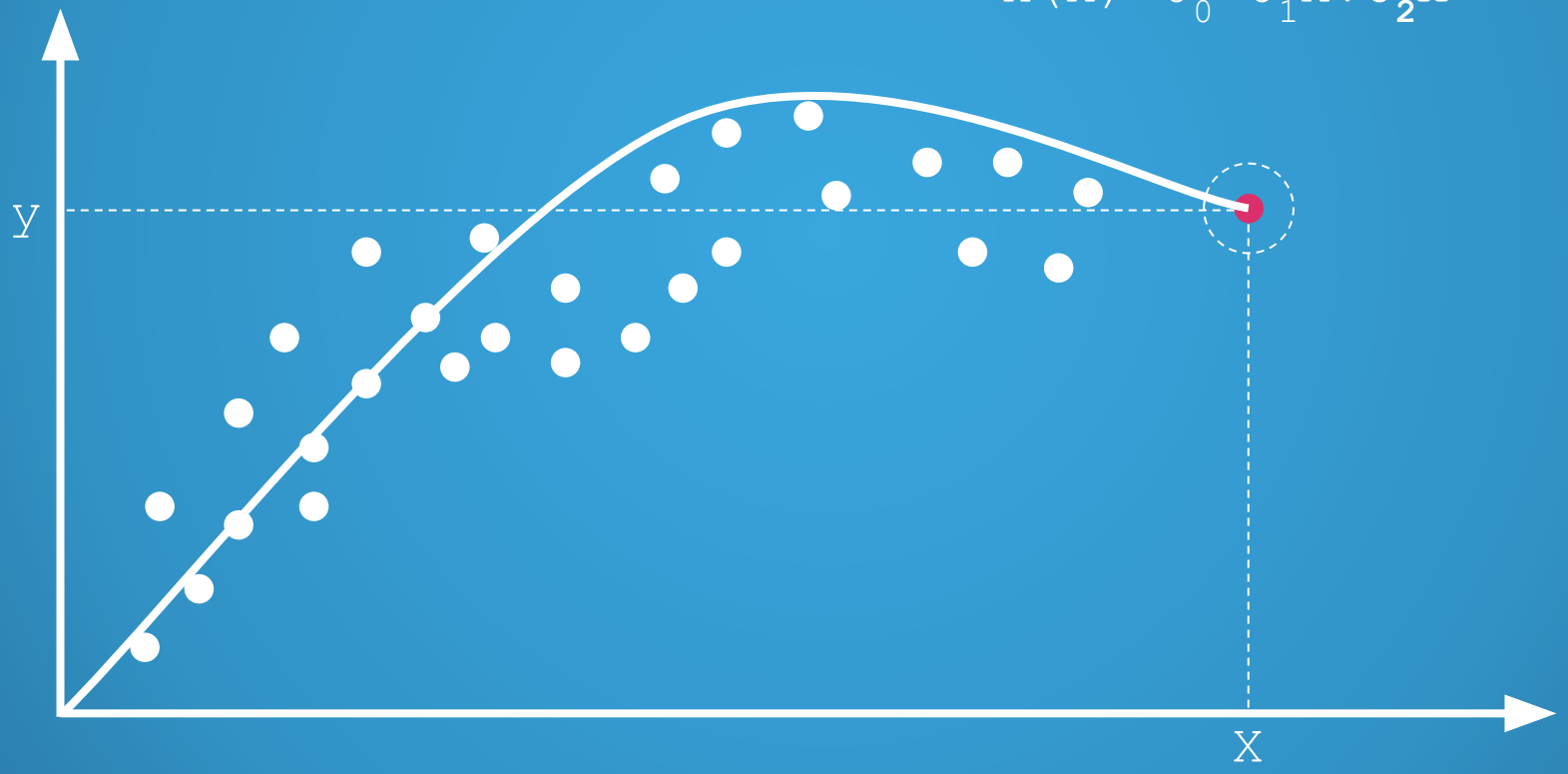
# REGRESSION / LINEAR

$$h(X) = \theta_0 + \theta_1 X$$



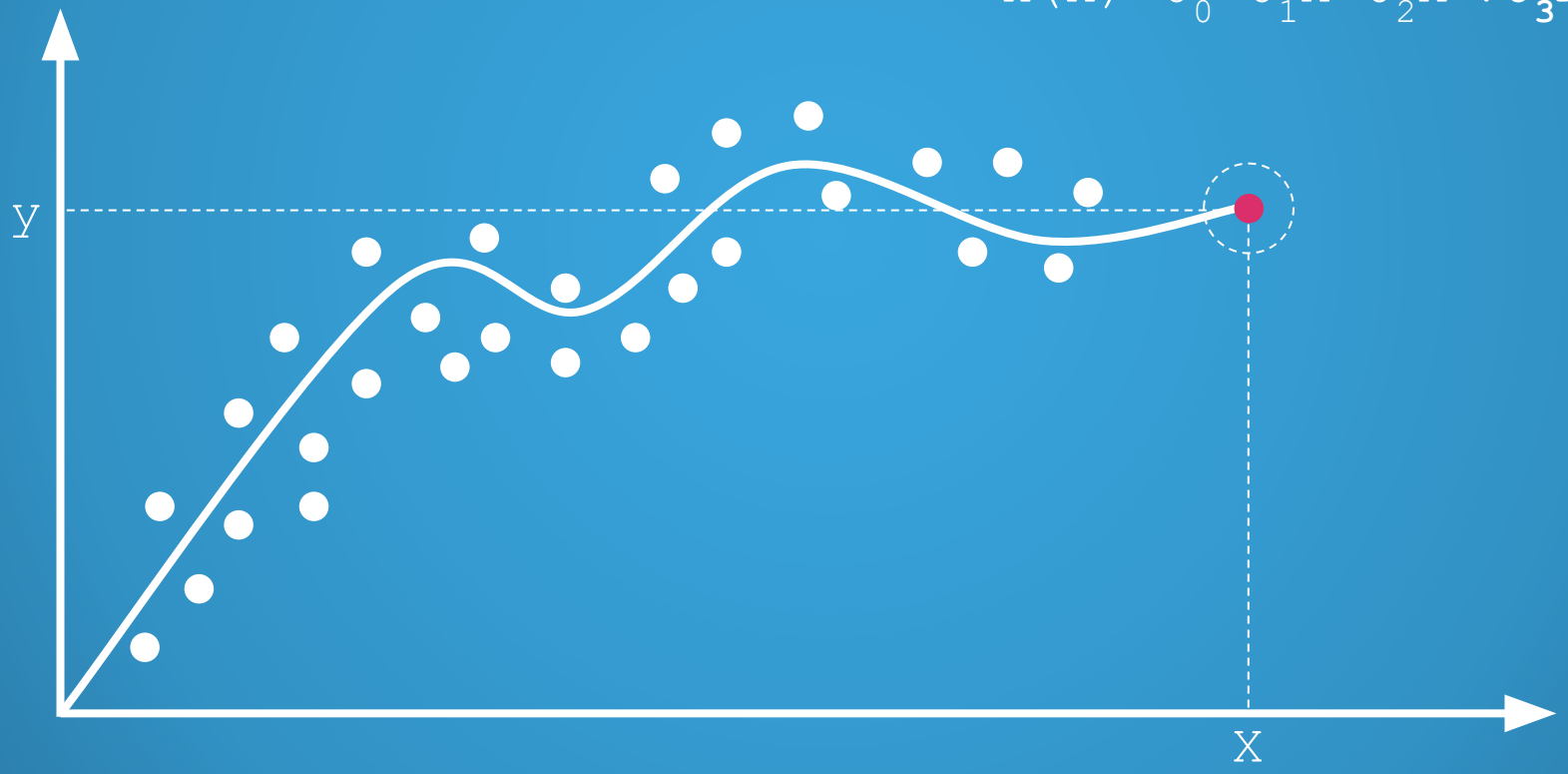
# REGRESSION / POLYNOMIAL

$$h(X) = \theta_0 + \theta_1 X + \theta_2 X^2$$



# REGRESSION / POLYNOMIAL

$$h(X) = \theta_0 + \theta_1 X + \theta_2 X^2 + \theta_3 X^3$$

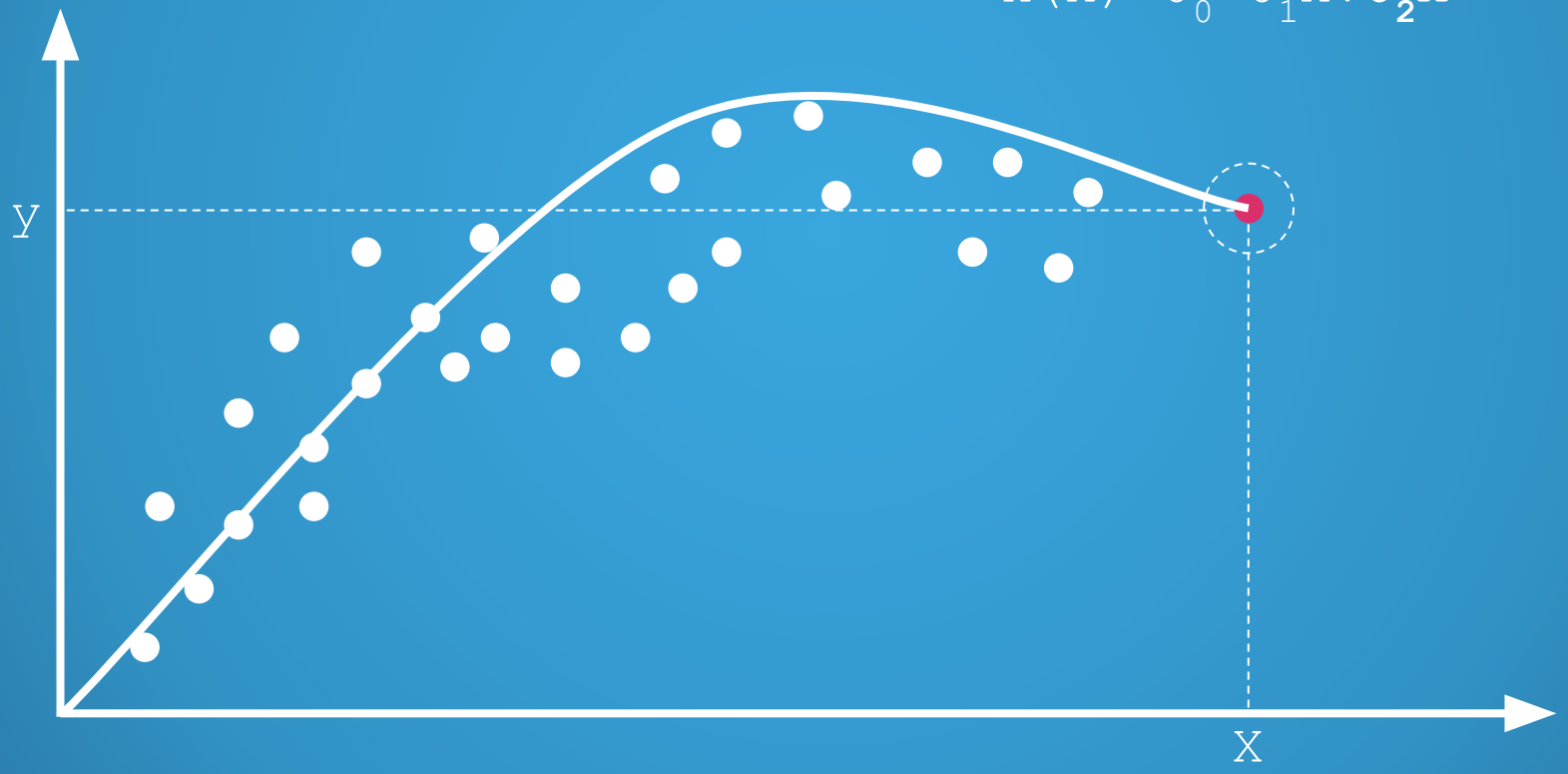


# REGRESSION / POLYNOMIAL

$$h(X) = \theta_0 + \theta_1 X + \theta_2 X^2 + \theta_3 X^3$$

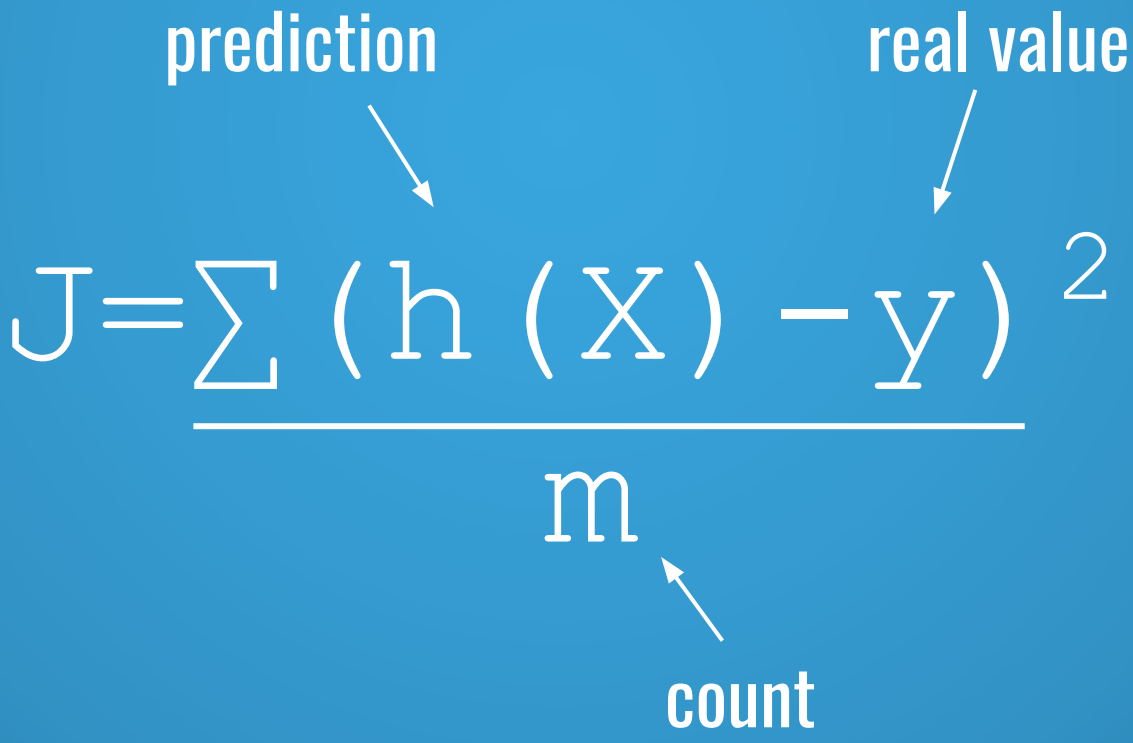
# REGRESSION / POLYNOMIAL

$$h(X) = \theta_0 + \theta_1 X + \theta_2 X^2$$



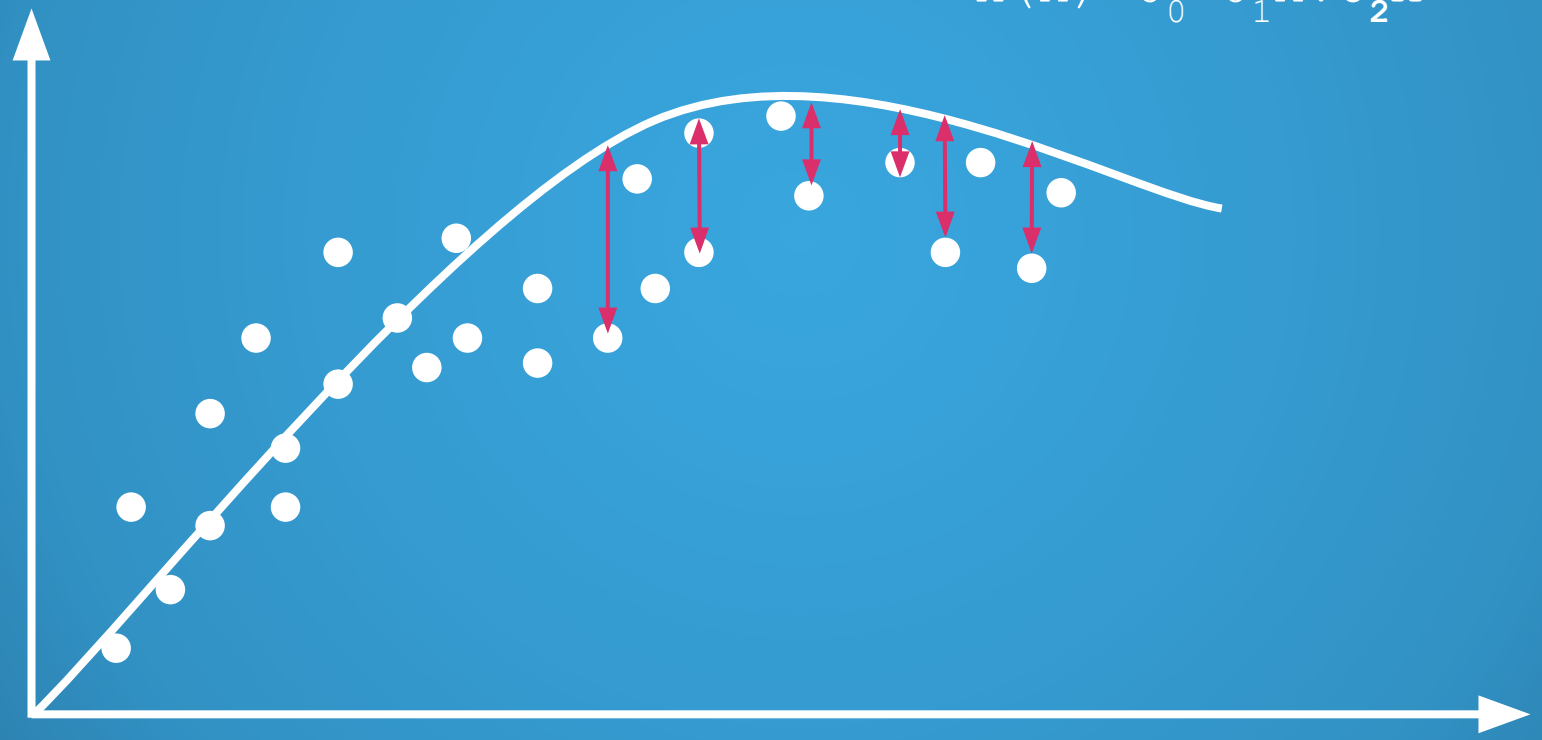
# REGRESSION / COST FUNCTION

Sum of errors :



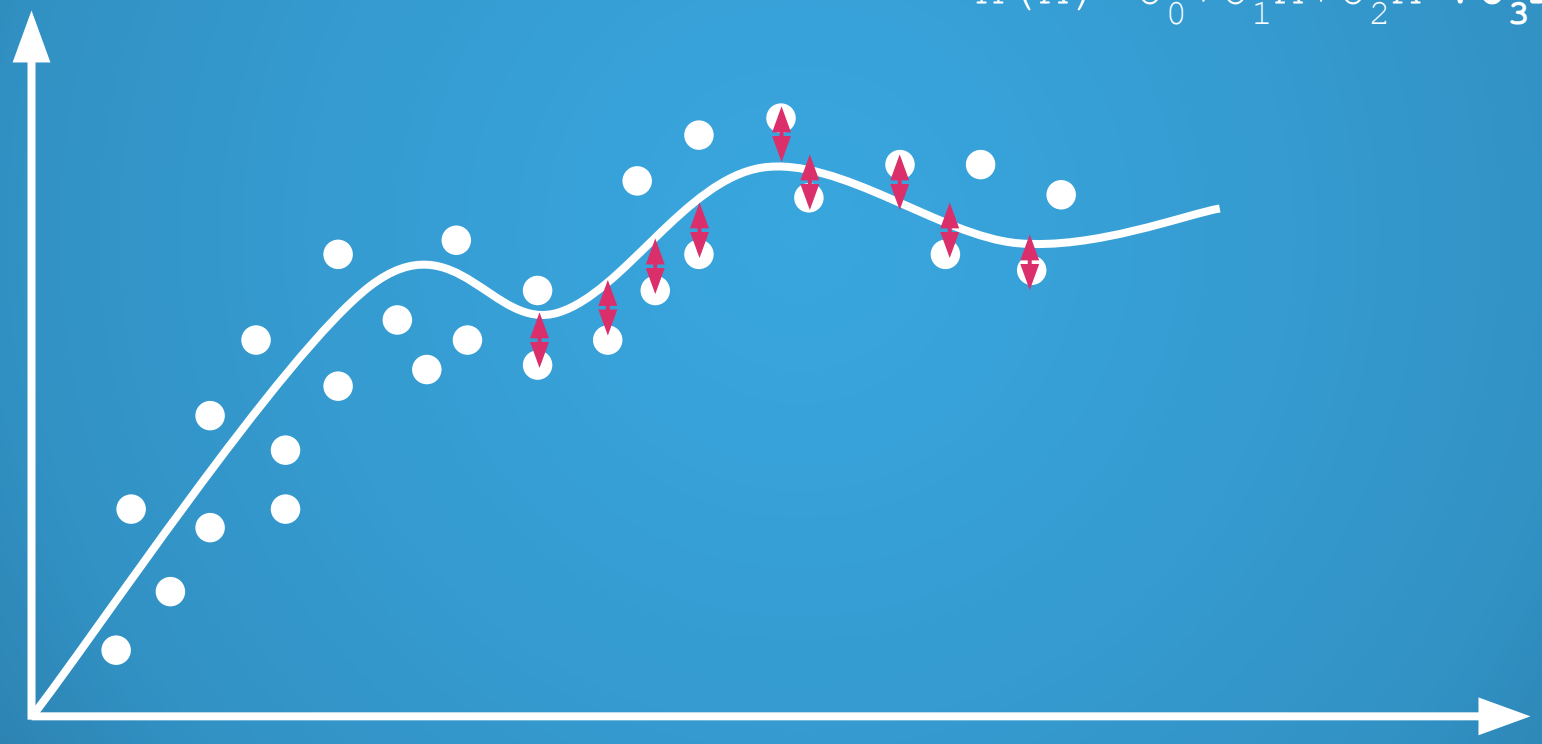
# REGRESSION / COST FUNCTION

$$h(X) = \theta_0 + \theta_1 X + \theta_2 X^2$$



# REGRESSION / COST FUNCTION

$$h(X) = \theta_0 + \theta_1 X + \theta_2 X^2 + \theta_3 X^3$$



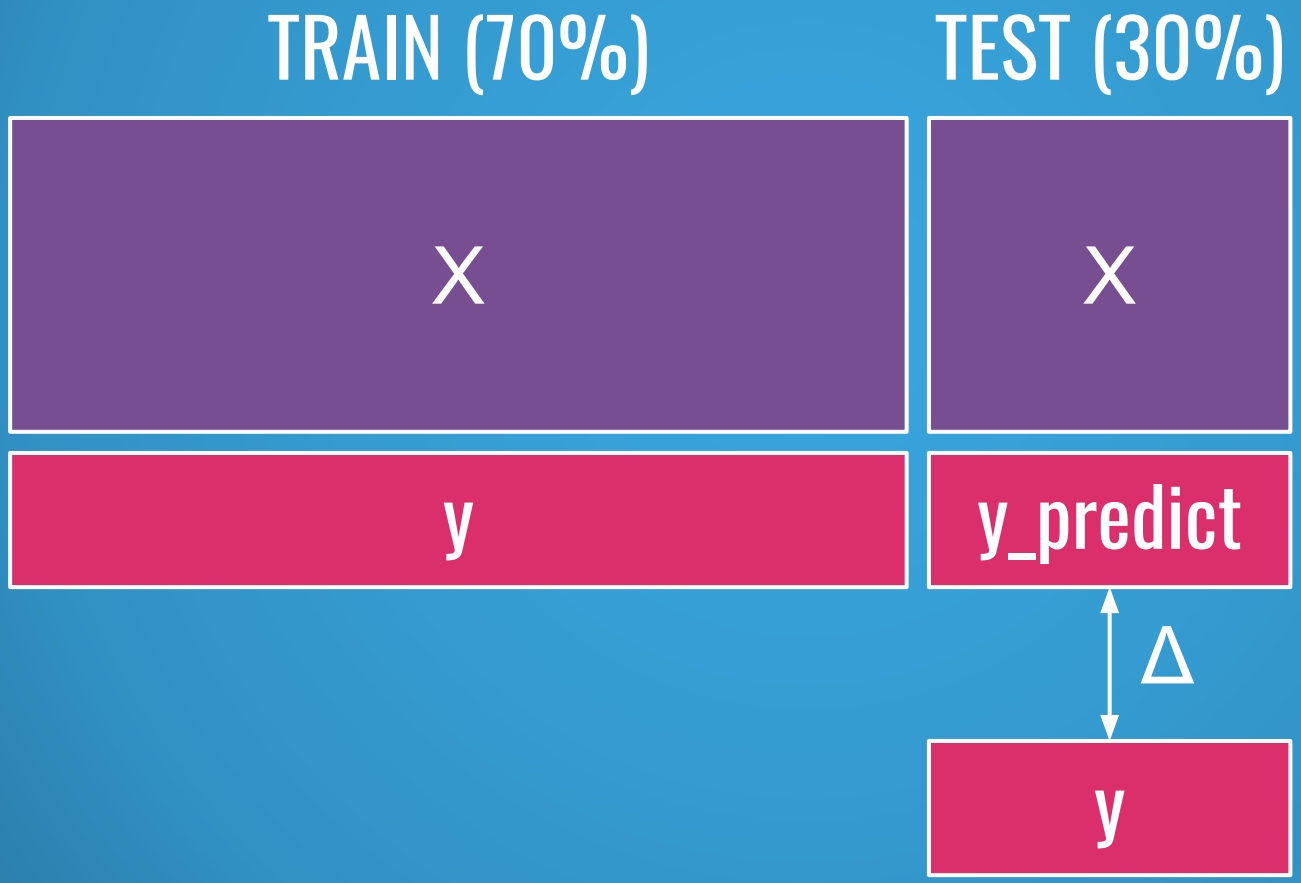
DEMO TIME !

# 5

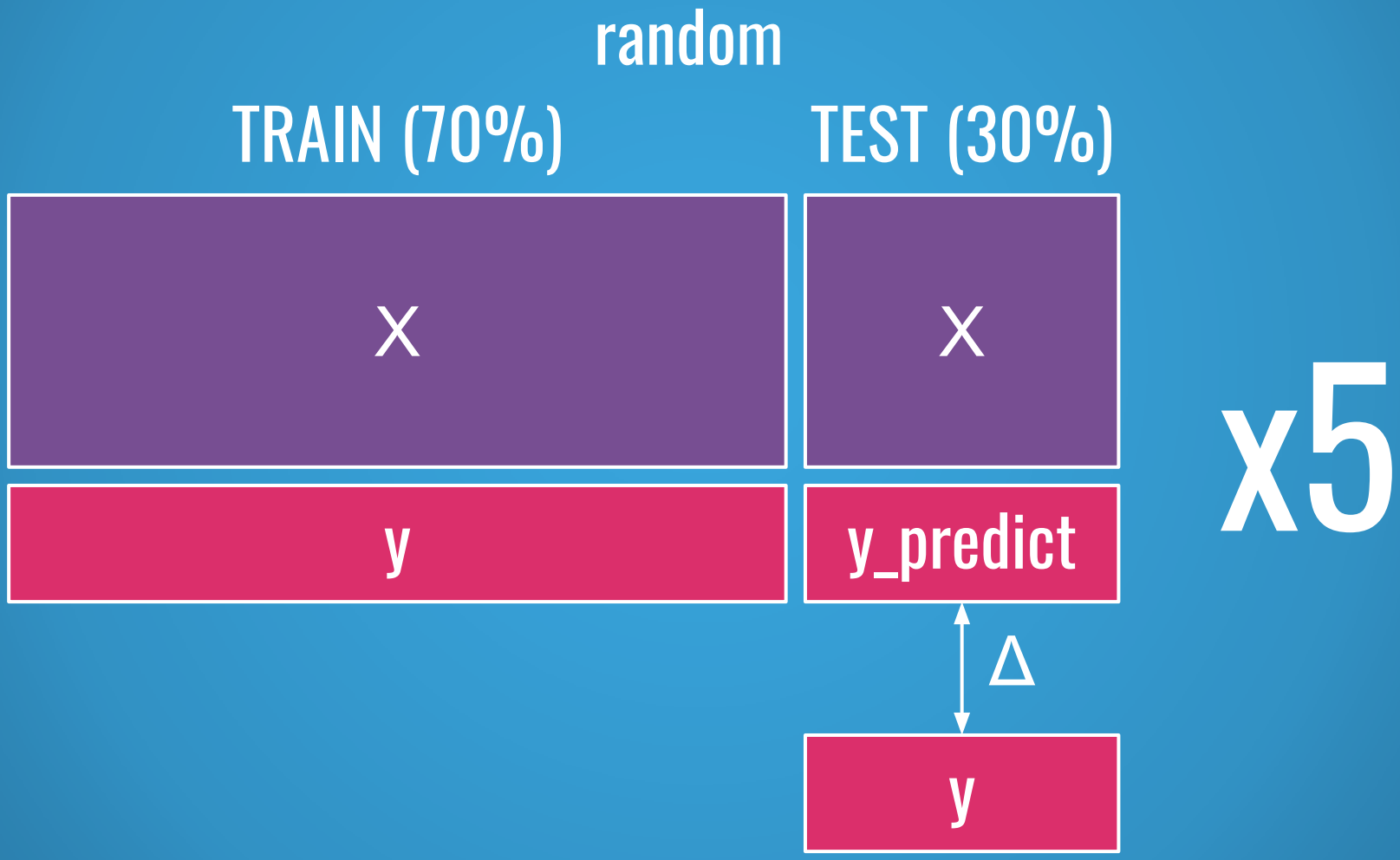
# VALIDATE THE PREDICTION

# How to validate a trained model

# TRAIN & TEST

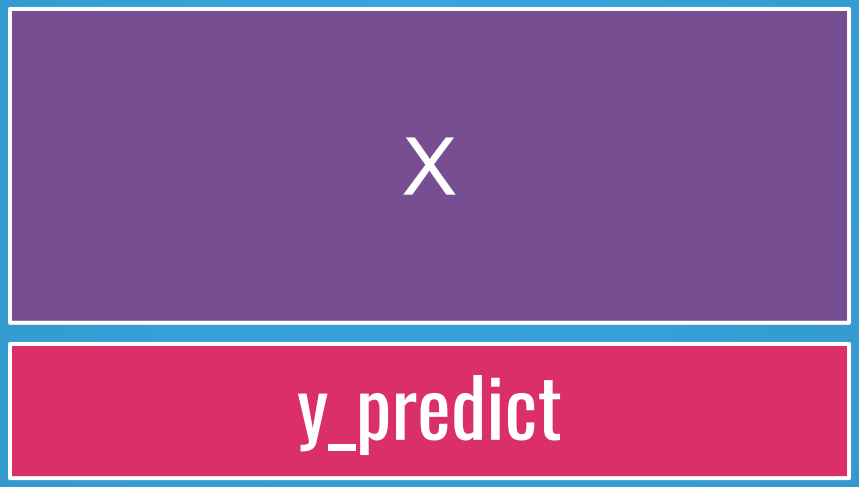


# CROSS VALIDATION



# PREDICT

NEW DATA



# DOES JACK SURVIVE ?

6

# DEMO TIME !

# RESOURCES

- Coursera Machine Learning

  https://www.coursera.org/learn/machine-learning

- Coursera Deep Learning

  https://www.coursera.org/specializations/deep-learning

- Kaggle

  http://www.kaggle.com

# THANKS

## ANY QUESTIONS ?

zelros.com / fabien.vauchelles.@zelros.com / @fabienv

http://bit.ly/ml-aismarttech