ASTRA

OFFICIAL PRESTIGE PARTNER

Jakarta Palembang 2018

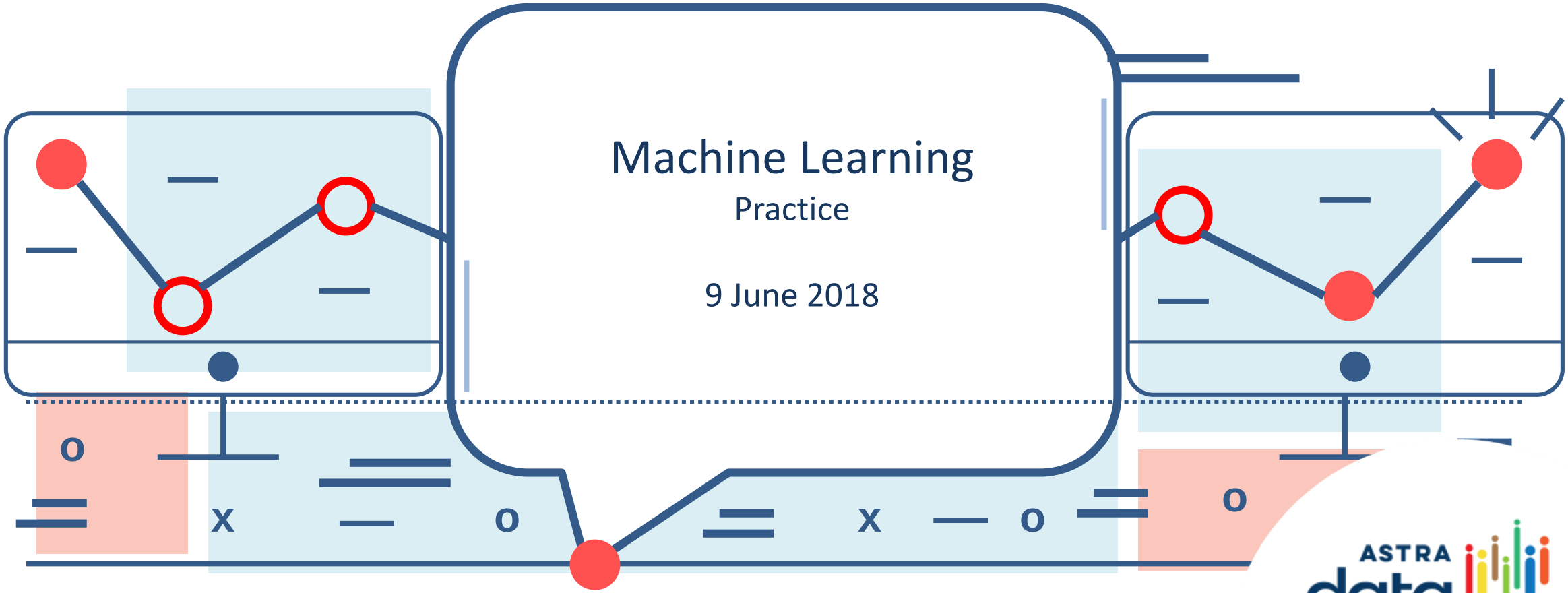
satu Indonesia

Semangat Astra Terpadu Untuk Indonesia

# Machine Learning

## Practice

9 June 2018




iykra

ASTRA data scientist bootcamp

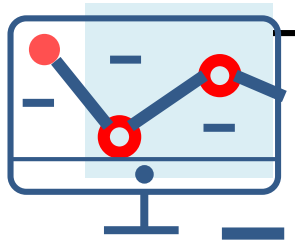
Adi Wijaya

Co-Founder of DataLabs Indonesia. I Want to Bring Together People, Data and Technology to Help Enterprise Accelerate Analytics Transformation. Has more than 5 years experience in data & analytics technology world, helping organizations understand the beauty of Data Analytics

**Poke me, and let's talk about**

Email : adi@datalabs.id  Website : datalabs.id
Linkedin : https://www.linkedin.com/in/adiwidjaja

ASTRA data scientist bootcamp



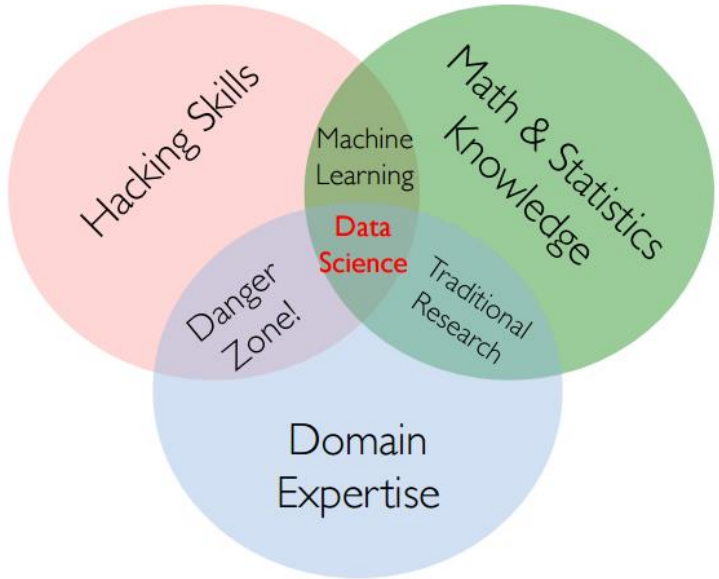
# AGENDA

1. Practice
2. Practice
3. Practice

ASTRA data scientist bootcamp

# What is Data Science?

Data Science aims to derive knowledge from big data, efficiently and intelligently Data Science encompasses the set of activities, tools, and methods that enable data-driven activities in science, business, medicine, and government




ASTRA data scientist bootcamp

# Roles in Data Science

## Data Engineer

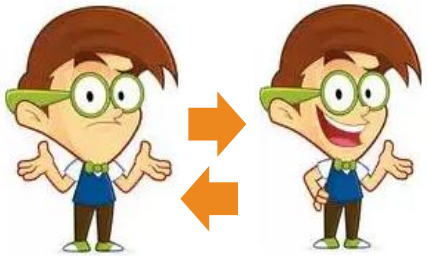
I am able to

- Answer any data related question
- Give information of data availability
- Provide data needed
- Operationalized Analytics Result

## Data Scientist

I am able to

- Answer any question with fact and exact number
- Solve business problem using data and math
- Visualize and explain the findings to grandmother



## Business Analyst

I am able to

- Gather important questions and build story from questions and facts
- Communicate to people outside the shell (team)
- Provide accurate knowledge library

ASTRA data scientist bootcamp

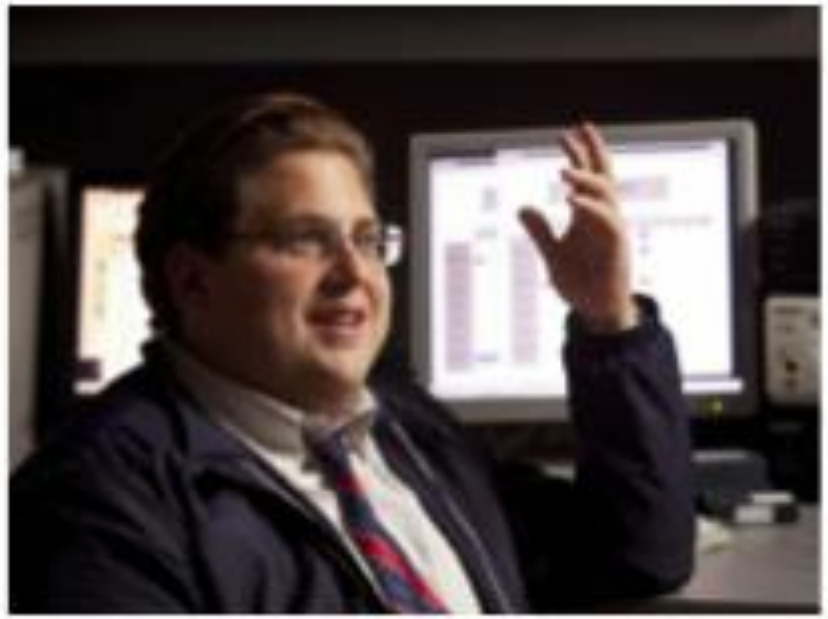
# What is Data Scientist?

## Data Scientist:

### *The Sexiest Job of the 21st Century*

***Meet the people who can coax treasure out of messy, unstructured data.***

*by Thomas H. Davenport and D.J. Patil*



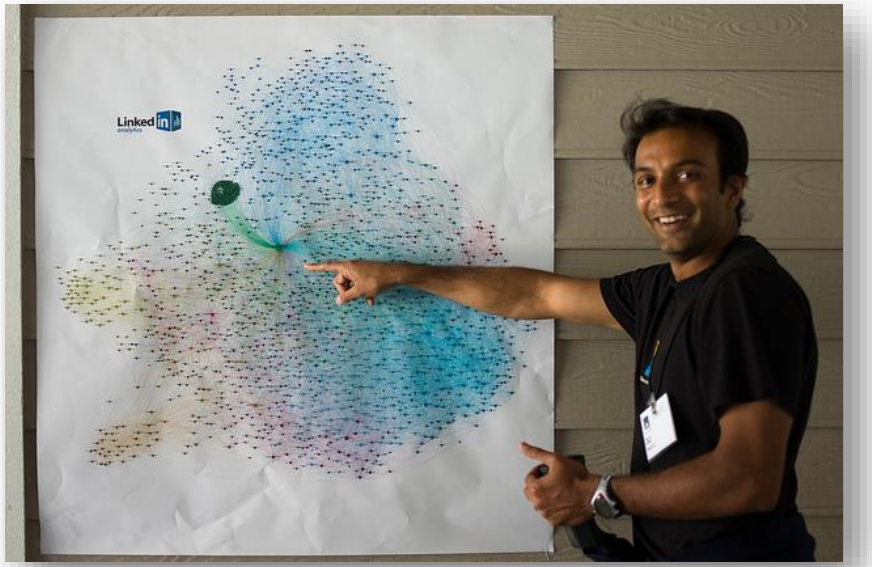
ASTRA data scientist bootcamp

# Top Salary for Math's Related Role by Quora

| RANK | JOB TITLE | MEDIAN PAY | % MATH MAJORS | MOST COMMON DEGREE LEVEL |
|---|---|---|---|---|
| 1. | Data Scientist, IT | $109,700 | 8% | Bachelor's |
| 2. | Quantitative Analyst | $103,300 | 14% | Master's |
| 3. | Actuary | $97,900 | 33% | Bachelor's |
| 4. | Data Modeler | $91,200 | 18% | Master's |
| 5. | Senior Actuarial Analyst | $86,600 | 40% | Bachelor's |
| 6. | Statistician | $82,200 | 11% | Master's |

ASTRA data scientist bootcamp

# DATA SCIENTIST

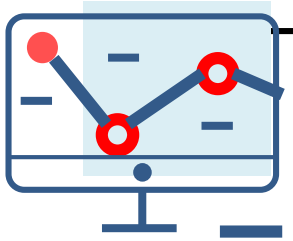



DJ Patil, former Linkedin and White House Data Scientist. Together with Jeff H (former Facebook) invent the term Data Scientist in 2011

"A Data Scientist is that a **unique** blend of skills that can both **unlock** the insights of data and **tell** a fantastic story via the data"

**-- DJ Patil --**

ASTRA data scientist bootcamp



# DATA SCIENTIST

"A Data Scientist is that a **unique** blend of skills that can both **unlock** the insights of data and **tell** a fantastic story via the data"

## Key Factors

Unique Blend of Skills

Unlock Insights

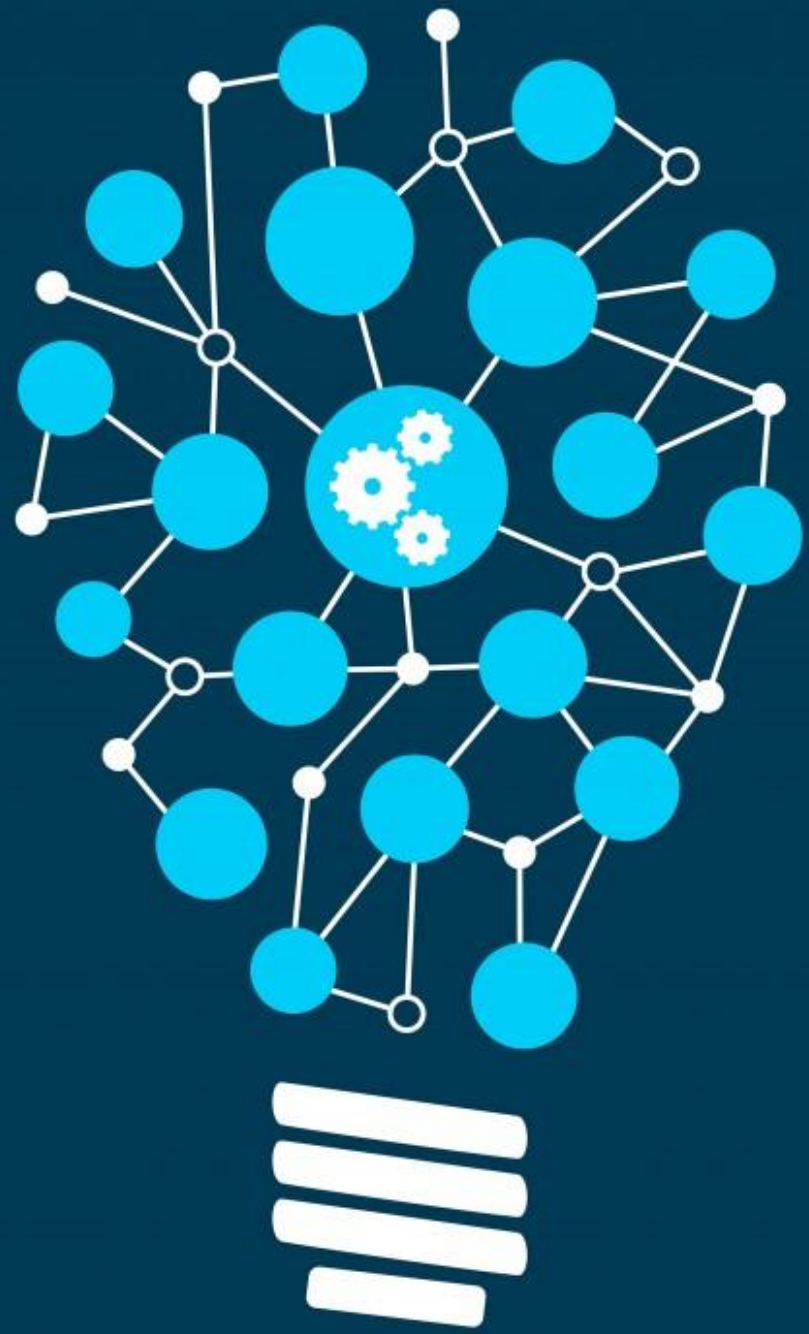
Tell Fantastic Story

## Skills Required

- Curiosity
- Creativity
- Business Intuition
- Data Cleansing
- Programming Language
- Machine Learning
- Visualization
- Communication
- Presentation

ASTRA data scientist bootcamp

# MACHINE LEARNING



# Machine Learning

**Arthur Samuel(1959).**
Machine Learning : Field of study that gives computers the ability to learn without being explicitly programmed

ASTRA data scientist bootcamp

# Types of Machine Learning

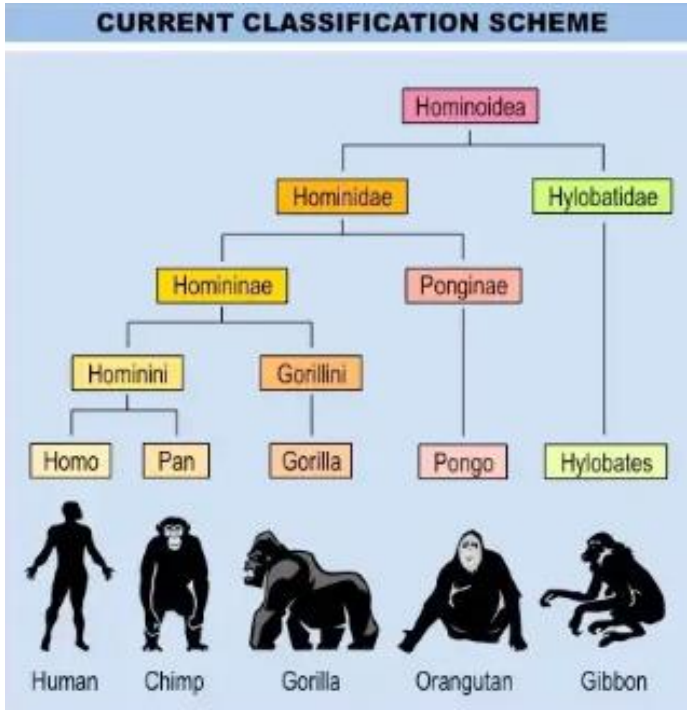
Supervised Learning

Unsupervised Learning

ASTRA data scientist bootcamp

# Supervised Learning

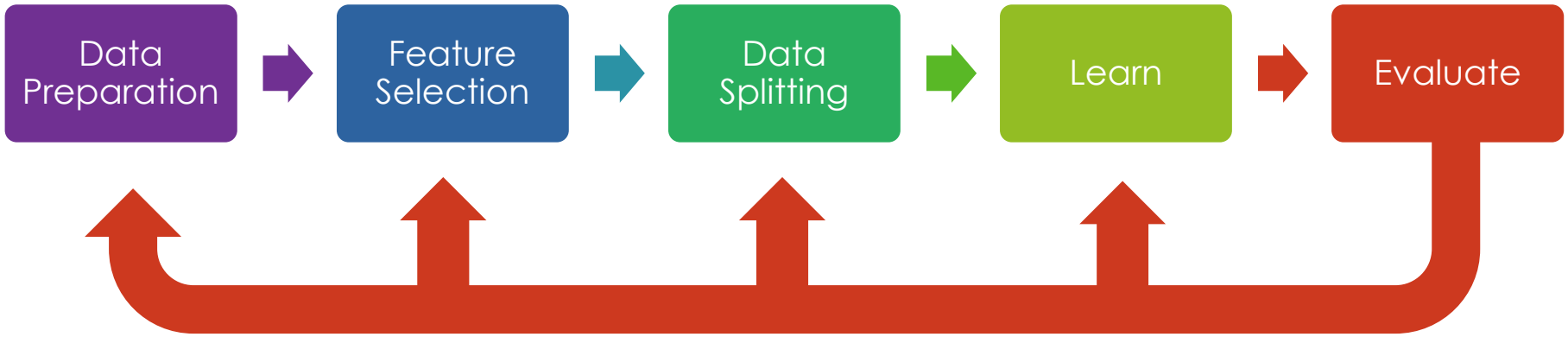
**Regression**
(Amount & Number)




**Classification**
(True False & Class)

ASTRA data scientist bootcamp

# Machine Learning Steps




Machine Learning steps:

**Data preparation** including set data to acceptable format, cleansing data, filtering
**Feature selection** is when you need to decide which variables best for build your model, this play big role on how accurate the model
**Data Splitting** It is required for supervised ML algorithm to evaluate the model, there are no rule of thumb for splitting, but most probably we split data into 70 % training set and 30% test data set
**Learn** is process when you teach computer using selected algorithm and matematical method using training dataset
**Evaluate** Test the model using test data and evaluate the accuracy, there are some measurement available like RSS, RMSE, R2, etc

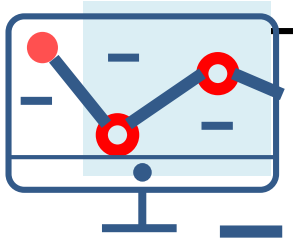
ASTRA data scientist bootcamp

# Machine Learning
# Common Mistakes
## for Beginner

Accuracy is GOD

1. Include Target Column as Features
2. For classification : Only check global accuracy
3. Use old algorithm

ASTRA data scientist bootcamp



# Thank You

ASTRA data scientist bootcamp