Modul 3

ML Introduction

Data Science Program

Purwadhika
Startup and Coding School

Outline

What is Machine Learning ?

ML Application

Why Machine Learning ?

ML Type

Data

ML Business Formulation

Purwadhika
Startup and Coding School

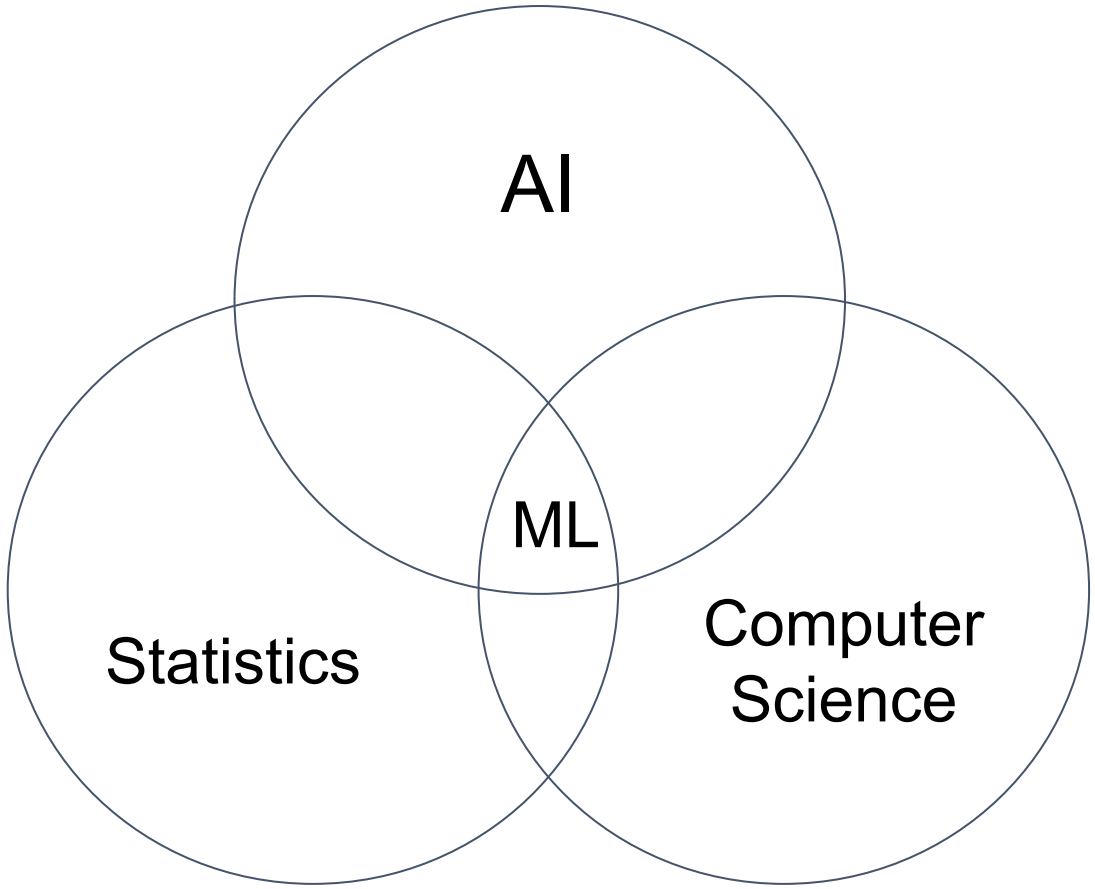
What is Machine Learning (ML) ?

An algorithm that learn from data and then making prediction

Imagine you bring a child to a park and we “teach” them to recognize many things such as chair, table, cat, bike, etc..

Purwadhika
Startup and Coding School

What is Machine Learning (ML) ?



ML also known as:

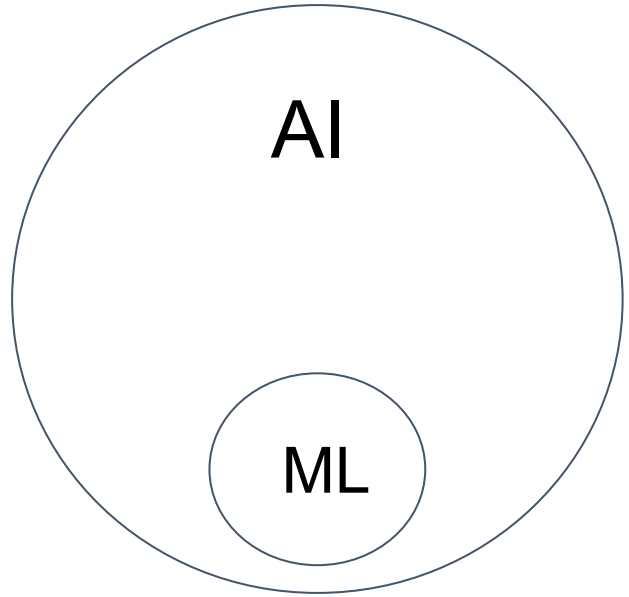
- predictive analytics
- statistical learning

Purwadhika
Startup and Coding School

AI vs ML

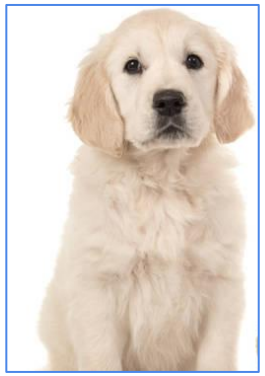
AI : make computer mimic human.

ML : A subset of AI that make computer mimic human without specifically programmed and let it learn from data



Purwadhika
Startup and Coding School

AI but not considered as ML



Input either cat images or dog

Specifically Programmed Model/Rule

- Dog is usually bigger or taller than cat
- dogs have a longer snout than cats
- the nose is black
- etc

“Dog”

Purwadhika
Startup and Coding School

AI considered as ML

Input an image

Input many Cat and Dog Images

Learning Process

Model/Rule

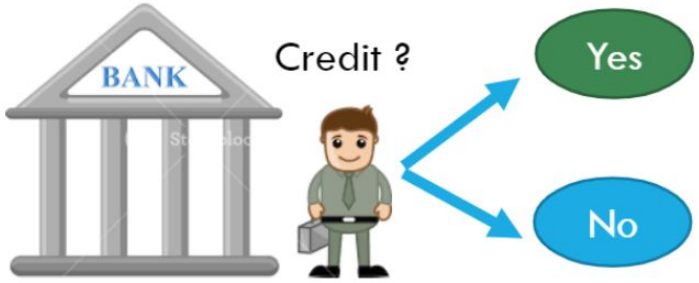
“Cat”

Purwadhika
Startup and Coding School

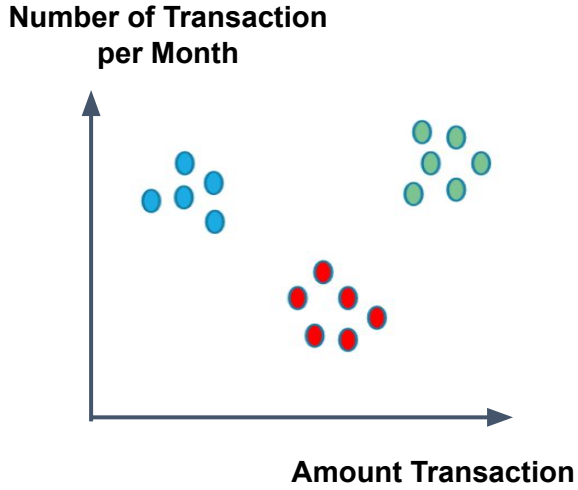
ML Application Various Industry

- Financial Services
- Government
- Health Care
- Retail
- Oil and Gas
- Transportation

ML Application



How Much ?



Purwadhika
Startup and Coding School

ML Application

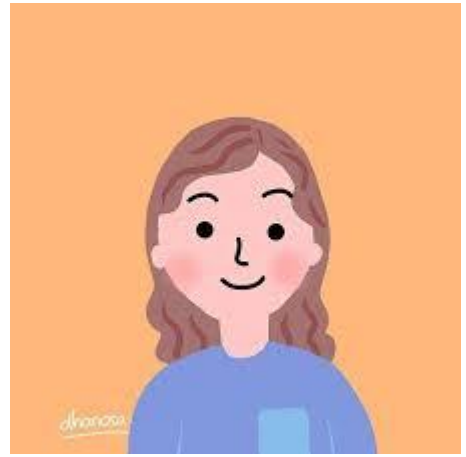
Movie Recommendation

Image Classification

Purwadhika
Startup and Coding School

Why do we need ML ?

- Human is really good at **detecting pattern**. However, we know the concept of **exhausting**.
- **Computer** is not born with our privileges. However, they can **work endlessly** 24/7, 7 days a week, 30 days per month, and 365 days to catch up with us.
- We can **leverage** this computer **advantages** to help us solve complex problems, which we might tired or unable to do.
- ML can **adapt** if exposed to **new data**. Adaptation is very necessary because data volume is growing and data variety increase.



40 years old
woman
rent a home
2 dependents

Loan
Rejected

Purwadhika
Startup and Coding School

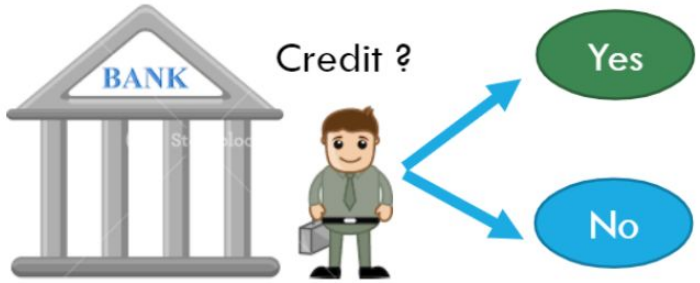
Type of ML

- bring the kid to park and also teach them the name (supervised)
- bring the kid but let them observed by themselves (unsupervised)

Supervised Learning

Unsupervised Learning

Classification



Regression

How Much ?

Telco Customer Segmentation

Purwadhika
Startup and Coding School

Type of ML

	Supervised	Unsupervised
Dataset	Labeled Dataset	Unlabeled Dataset
Task	Predicting or Classifying based on previous labelled data	Find hidden pattern and grouping from unlabelled data
Method	Classification. Regression	Clustering, Dimensional Reduction, Anomaly Detection
Algorithm	Linear Regression, Logistic Regression, KNN, Decision Tree, Random Forest, Boosting	K-Means, DBScan, t-SNE, PCA, Factor Analysis, One-class SVM, Isolation Forest
Use cases	Credit Scoring, Churn Analysis, Propensity Analysis, House Price	Customer segmentation, Market Basket Analysis, Fraud detection

Another Type of ML

- Semi-supervised Learning
- Reinforcement Learning

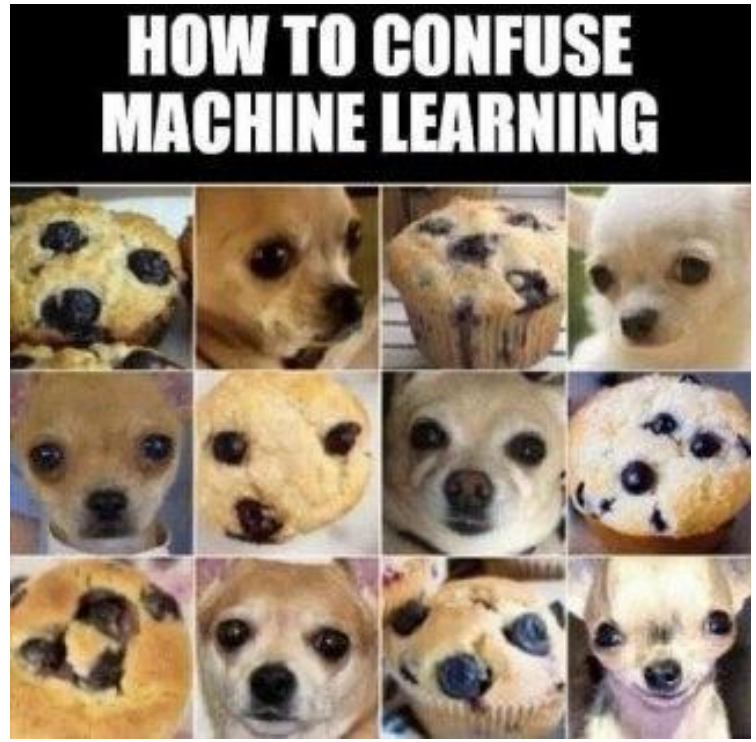
Data

You ML model only as good as your data

For example, you want to predict gender and the only feature you have is their last name. No algorithm will be able to predict their gender.

If you have another feature like their first name, it will be much better as it is often possible to tell the gender by a person's first name.

More features, can make you have more confidence.



source : 9gag

Purwadhika
Startup and Coding School

Data Illustration : Credit Scoring

Features

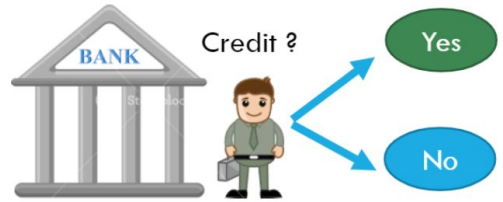
Label

Cust ID	Age	Gender	...	Edu	Balance	...	Income	...	...	Default
										BAD
										BAD
										GOOD
										...
										...
										GOOD

About identity/demographic

About their transaction

etc



Purwadhika
Startup and Coding School

Label and Features

Label :

- The label should come after the feature, because it’s predictive

Features :

- The features should come before the label, because it’s predictive
- The features should be available when prediction needed.
- A feature should be a phenomenon that related with the label

Label and Features

Age	...	Duration	Interested (call by phone)
23	...		Yes
24	...		No
26	...		No
27	...		Yes
...	...	...	...

For example we want to predict whether a customer will interested to take home loan before we call them.

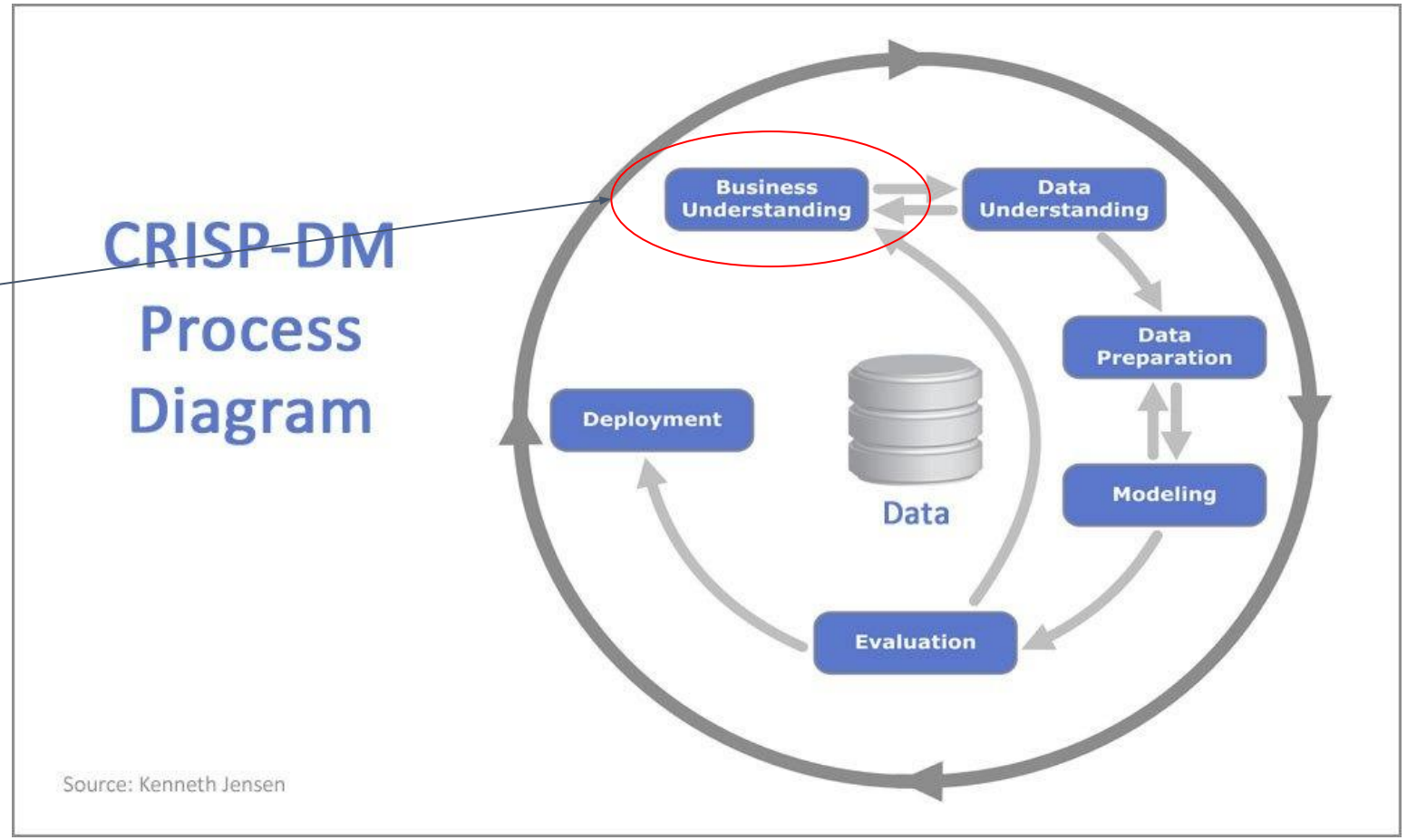
We use the historical data and when build machine learning model duration included as feature.

Even you got good performance in ML but this makes the ML model useless because duration is not available when we need it

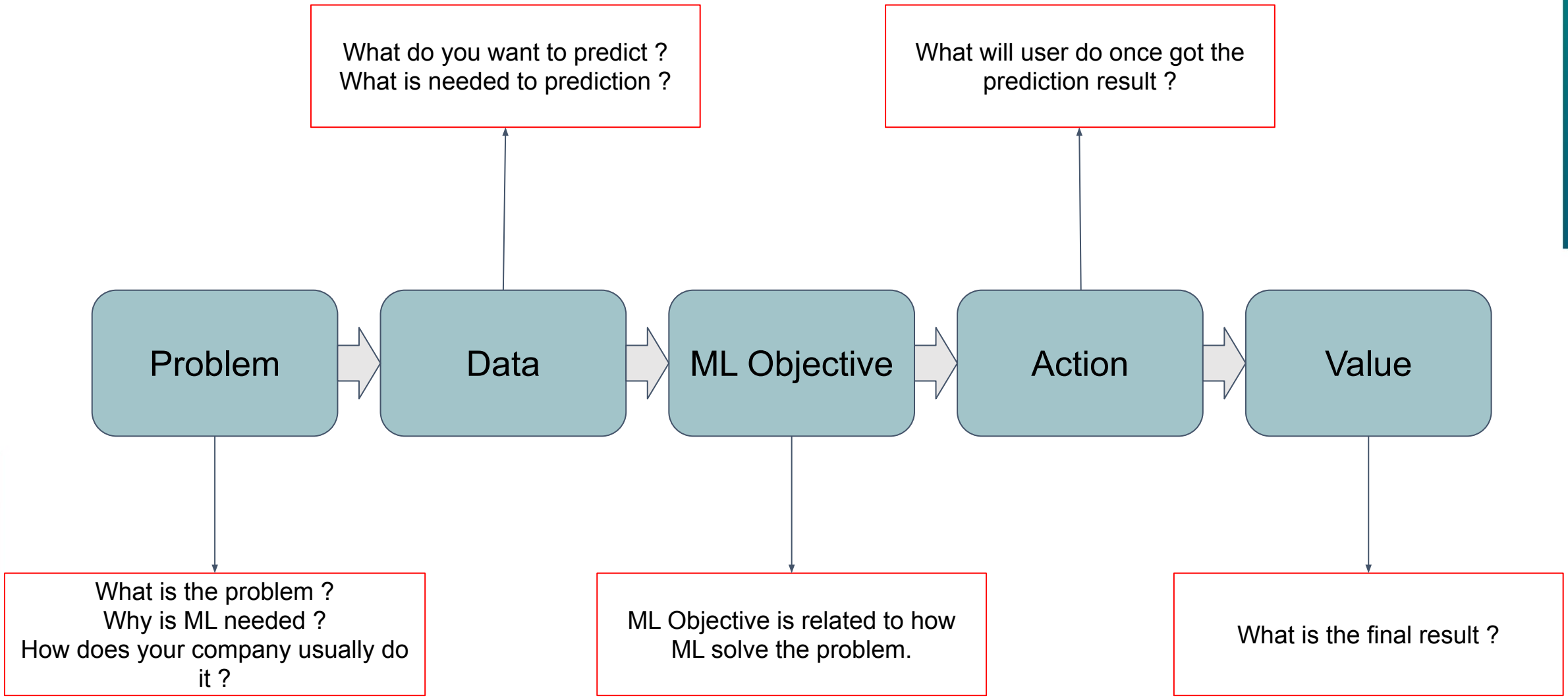
Purwadhika
Startup and Coding School

ML Business Formulation

- The Problem
- Data
- Objective
- Action
- Value



Purwadhika
Startup and Coding School

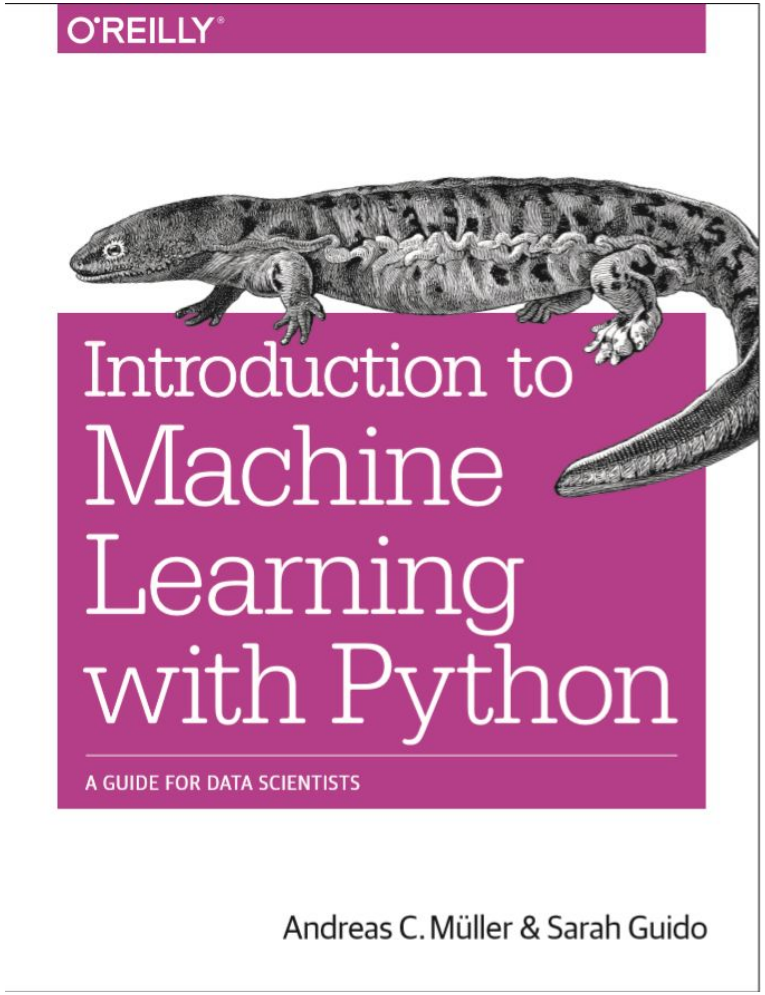


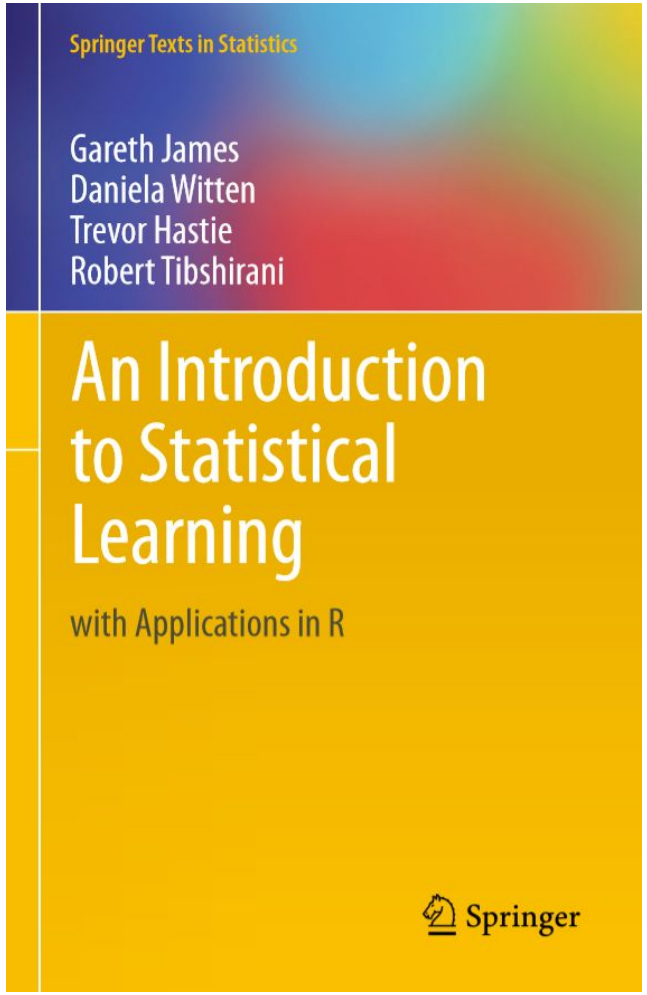
Purwadhika
Startup and Coding School

Problem	How to predict **default risk of the new applicant** so we can **allocate loan efficiently** and **increase profit** from loan ?
Data	• What is being predicted ? default risk of the new applicant • What is needed in prediction ? Demographical, Transaction behaviour, income, ect
ML Objective	Maximize (profit - potential revenue lost)
Action	Do not allocate loan to a customer when the risk is too high
Value	Profit Increase

Purwadhika
Startup and Coding School

References





Purwadhika
Startup and Coding School

References

https://www.the-modeling-agency.com/crisp-dm.pdf

https://scikit-learn.org/stable/

Purwadhika
Startup and Coding School