Introduction to Machine Learning using Tensorflow
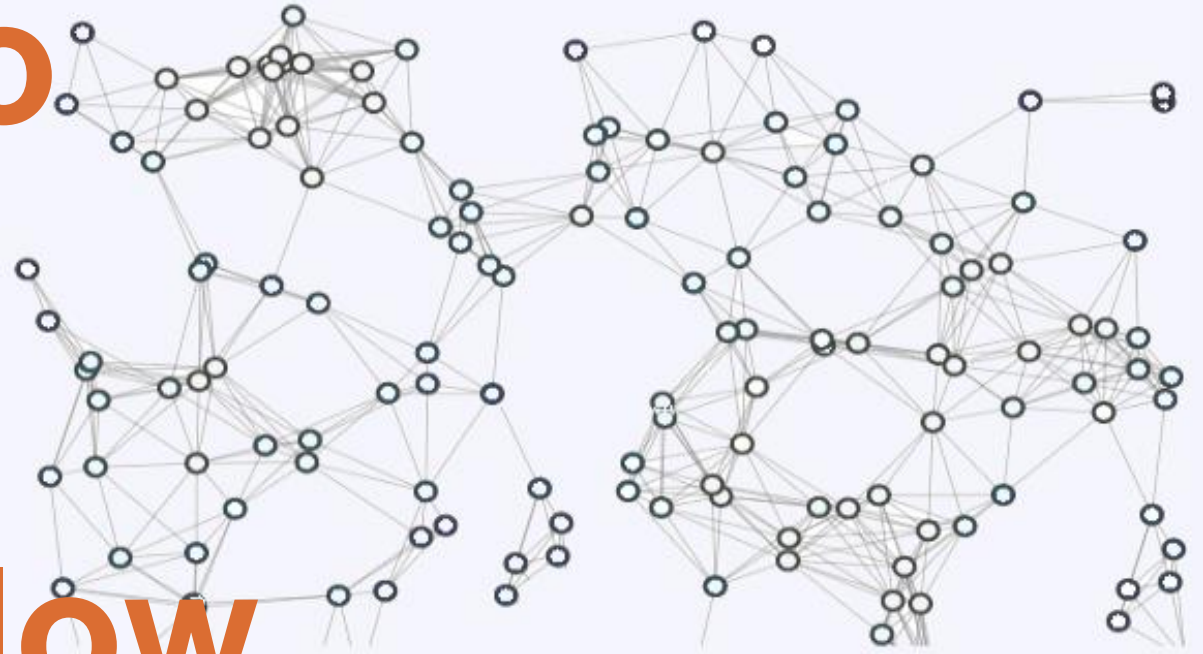
jundy.raga@gmail.com

Rodolfo C. Raga Jr.

September 23-24, 2019

San Beda College

Topic Outline

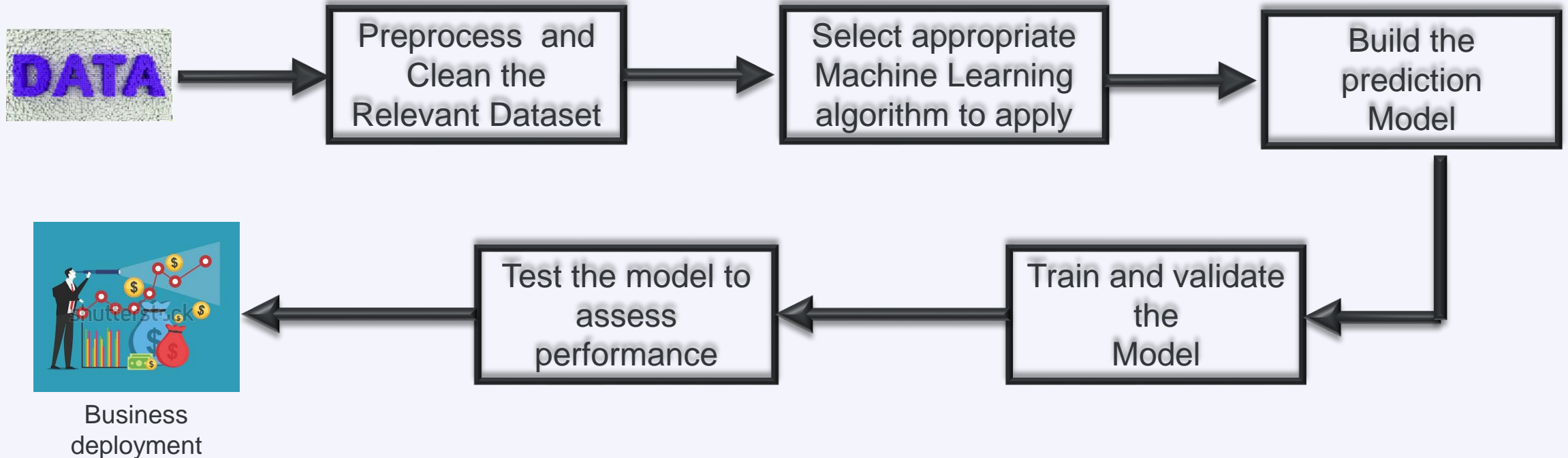| Day 1	Day 2

1) Intro to Machine Learning and Tensorflow
2) Data Preprocessing
3) Learning Models (Regression)
   3.1. NN Simple Linear
   3.2. NN Multiple Linear
4) Model Training
	1) Learning Models (Classification)
	   1.1. NN Logistic Regression
	   1.2. CNN Deep Learning
	2) Model Testing
	3) Model Evaluation / Validation Performance
	4) Data Visualization

Topic Outline

Day 1	Day 2
1) Intro to Machine Learning and Tensorflow	1) Learning Models (Classification)
2) Data Preprocessing	1.1. NN Logistic Regression
3) Learning Models (Regression)	1.2. CNN Deep Learning
3.1. NN Simple Linear	2) Model Testing
3.2. NN Multiple Linear	3) Model Evaluation / Validation Performance
4) Model Training	4) Data Visualization

Introduction

- **Machine learning** is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that **systems can learn from data**, identify patterns and make decisions with minimal human intervention. Source: https://www.sas.com/en_us/insights/analytics/machine-learning.html

Machine Learning Pipeline

DATA → Preprocess and Clean the Relevant Dataset → Select appropriate Machine Learning algorithm to apply → Build the prediction Model → Train and validate the Model → Test the model to assess performance → Business deployment

Machine learning tasks

- Supervised learning
  - **Input**: training data + desired outputs (labels)
  - regression: predict numerical values
  - classification: predict categorical values, i.e., labels
- Unsupervised learning
  - **Input**: training data (without desired outputs)
  - clustering: group data according to "distance"
  - association: find frequent co-occurrences
  - link prediction: discover relationships in data
  - data reduction: project features to fewer features

Supervised Learning

- The aim of **supervised learning** is to build a model that is 'good at' predicting the target variable, given the predictor variables.

- If the target is a continuously varying variable (e.g. price of a house), it is a **regression** task.

- Alternatively, if the target variable consists of categories (e.g. 'click' or 'not', 'malignant' or 'benign' tumor), we call the learning task **classification**.

Regression Algorithms

- Simple Linear Regression
- Multiple Linear Regression
- Support Vector Machines
- Perceptron

Classification

- Logistic Regression
- Support Vector Machines
- Deep Neural Networks

Supervised Learning pipeline

1. Load the dataset
2. Perform data preprocessing
3. Select type of model, e.g., regression, (deep) neural network, ...
4. Train model, i.e., determine parameters
   1. Data: input + output
      1. training data $\rightarrow$ determine model parameters
      2. validation data $\rightarrow$ yardstick for training performance
5. Test model (Evaluate model accuracy)
   1. Data: input + output
      1. testing data $\rightarrow$ final scoring of the model
6. Production (Predict results using the model)
   1. Data: input $\rightarrow$ predict output

OVERFITTING   OPTIMUM   UNDERFITTING
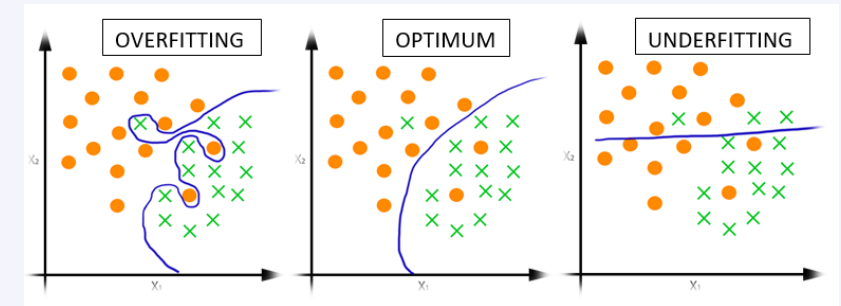
Development Tools

- Programming languages
  - Python
  - R
  - C++
  - ...
- Many libraries
  - Numpy
  - Pandas
  - Scikit-learn
  - PyTorch
  - TensorFlow
  - Keras
  - …

classic machine learning

deep learning frameworks

What is TensorFlow?

- TensorFlow was originally created by Google in 2011 as an internal machine learning tool.

- In November 2015, an implementation of it was open sourced under the Apache 2.0 License.

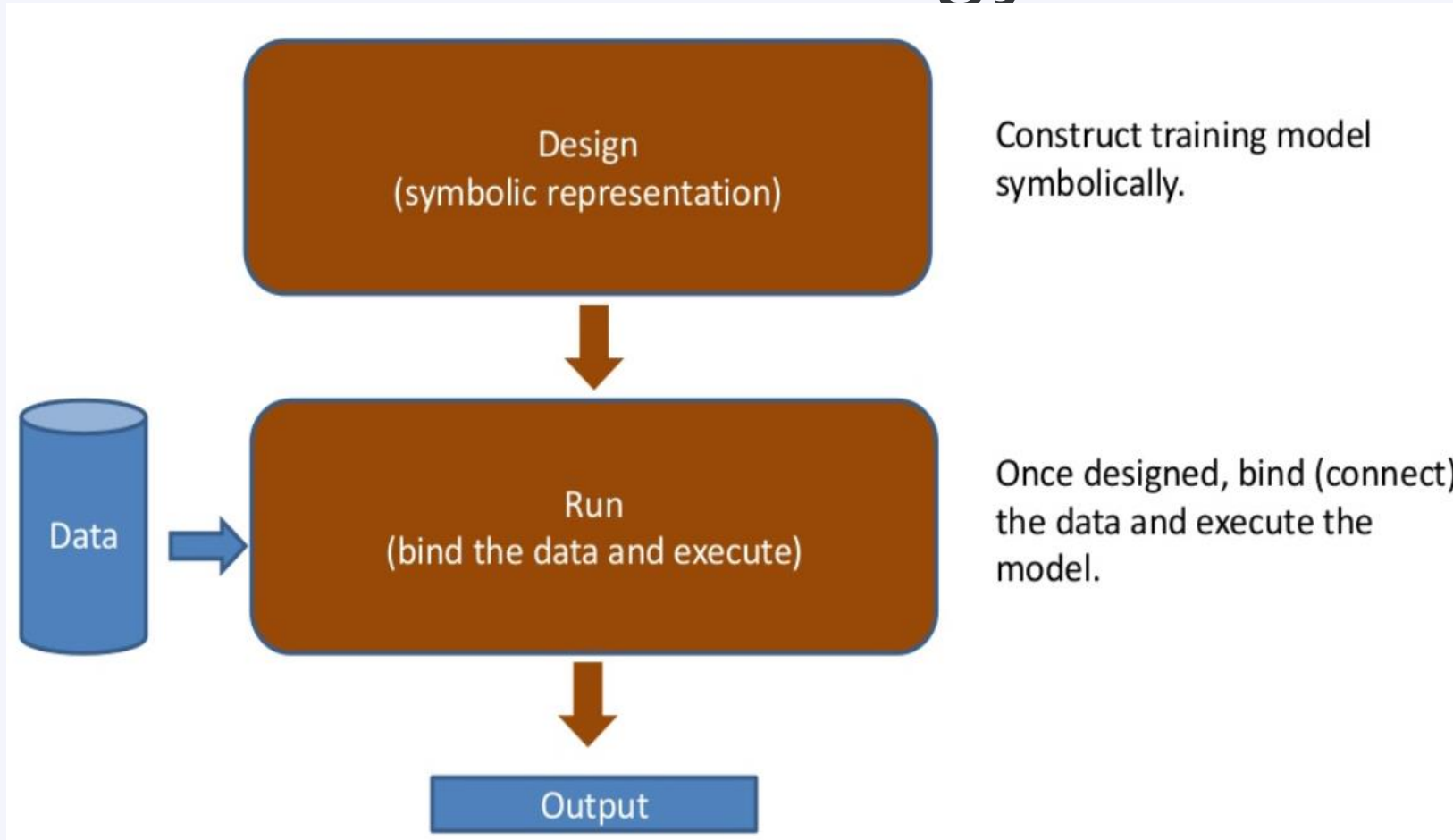- However, Google still maintains its own internal version.

What is TensorFlow?

- "TensorFlow is an interface for expressing machine learning algorithms, and an implementation for executing such algorithms".

- It is a computational framework for building machine learning and deep learning models.

TensorFlow employs a Design & Run Methodology

Design (symbolic representation) — Construct training model symbolically.

Data → Run (bind the data and execute) — Once designed, bind (connect) the data and execute the model.

Output

Companies using Tensorflow

- Google
- OpenAI
- DeepMind
- Snapchat
- Uber
- Airbus
- eBay
- Dropbox
- A bunch of startups