Welcome!

My Introduction

- Co-founder - DPhi/Skillspace AI
- Founding team - Complidata
- Entrepreneur - Problem-Solver - Data Scientist - Marketer.
- Love building and growing communities/teams.
- Led lean teams across various startups.

DPhi Vision:

Make Data Science education accessible to everyone!

Learning Objectives

ABC of Machine Learning & its use-cases

Types of Machine Learning

ML Keywords

Problem Solving

DPhi

The ABC of Machine Learning

DPhi

What is Machine Learning?

- Machine learning is the study of computer algorithms **that improve automatically through experience.**
- It allow computers to discover **hidden and useful insights**
- We feed examples to the system and the Machine Learning algorithms provide us with required results
- **In nutshell, Machine Learning is a new way of communicating your wishes to a computer.**

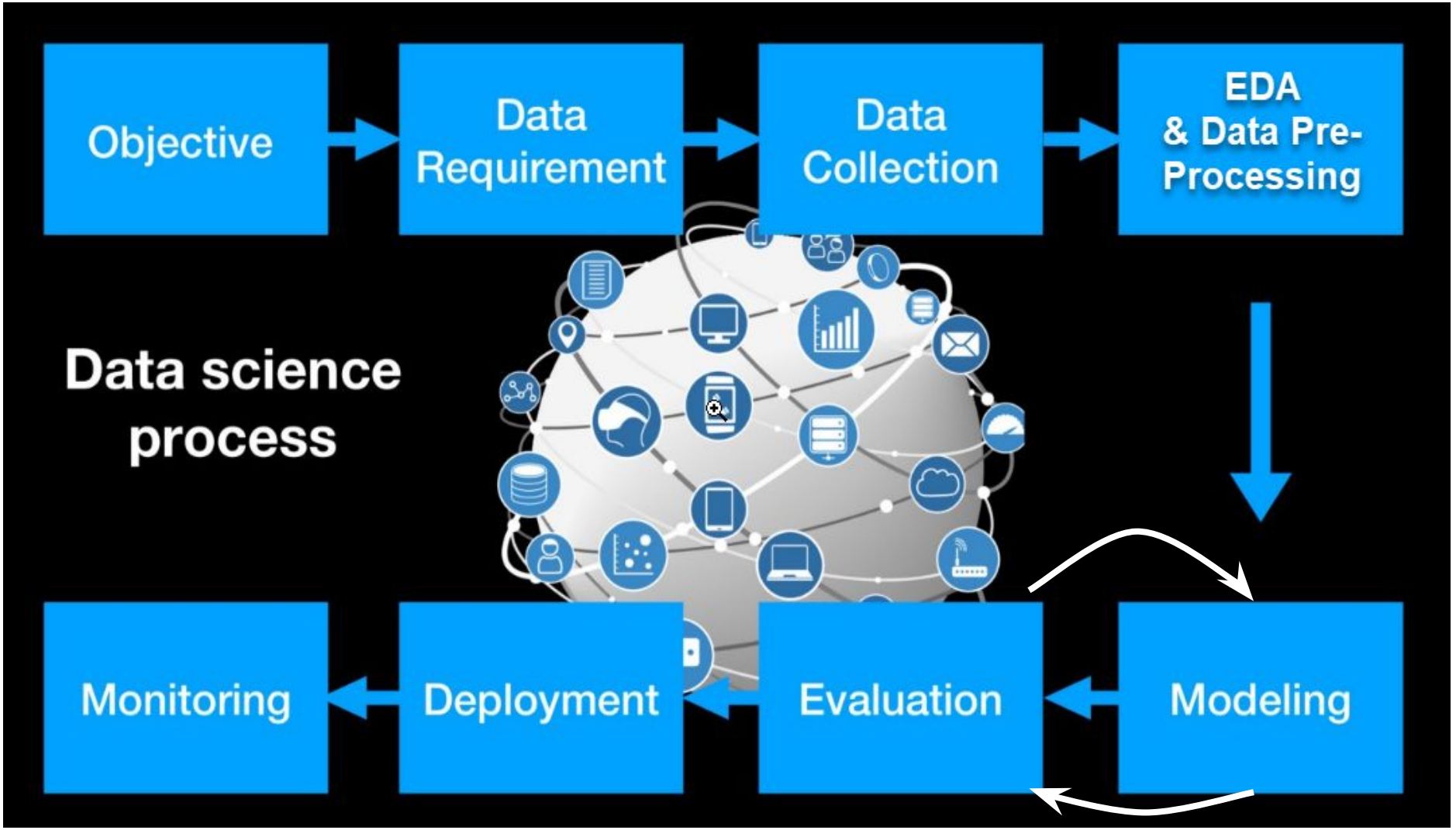
What is a Machine Learning (ML) model?

- For now, let's consider it is a Magical box that help us to predict what we want. In the below case we want to predict whether an incoming email should land in our inbox or spam box. We will discuss more about ML models soon.

In other terms this is nothing but **data**. This data will have variables such as: sender email id, subject of email, email body etc

Once the incoming emails go through the Machine Learning Model it categorizes and predicts whether a mail should go in your inbox or spam box

How? The Process



Credits: https://towardsdatascience.com/data-science-modeling-process-fa6e8e45bf02

DPhi

Machine Learning is used in..

- **Fraud detection** - **Eg:** Credit card fraud detection. It will help us to detect whether a transaction is fraud or not.

- **Email spam filtering** - **Eg:** Helps in categorising whether a particular email should go in inbox or spam box.

- **Product Recommendations** - **Eg:** E-commerce platforms like Amazon can recommend you a similar product based on your previously browsed list of products

- and many more!!!

Machine Learning is used in..

- How Google is using Machine Learning
- https://algorithmia.com/blog/machine-learning-use-cases
- How Facebook is using Machine Learning

Machine Learning is used in..

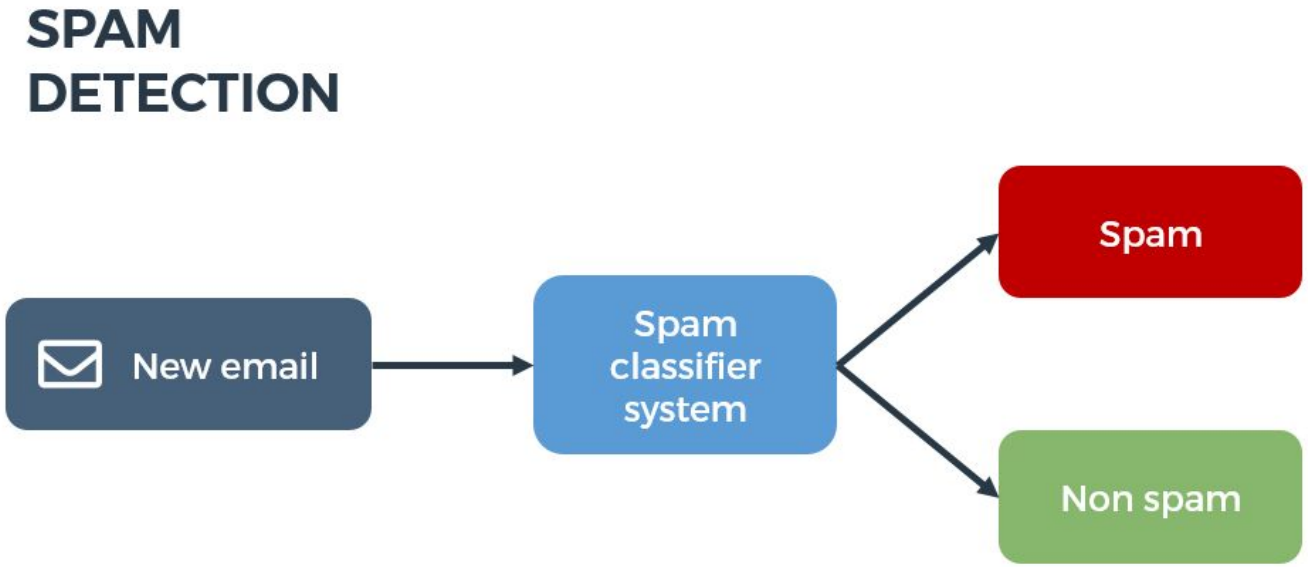
Any other use-cases?

DPhi

Let’s understand some keywords in ML!

Label/Target Variable

- The target variable or label of a dataset is the variable of a dataset about which you want to gain a deeper understanding.
- It is the variable that is, or should be the output.
- In the example of detecting spam emails, the label will be the category the email belongs to, i.e it will be either ‘spam’ or ‘not spam’.



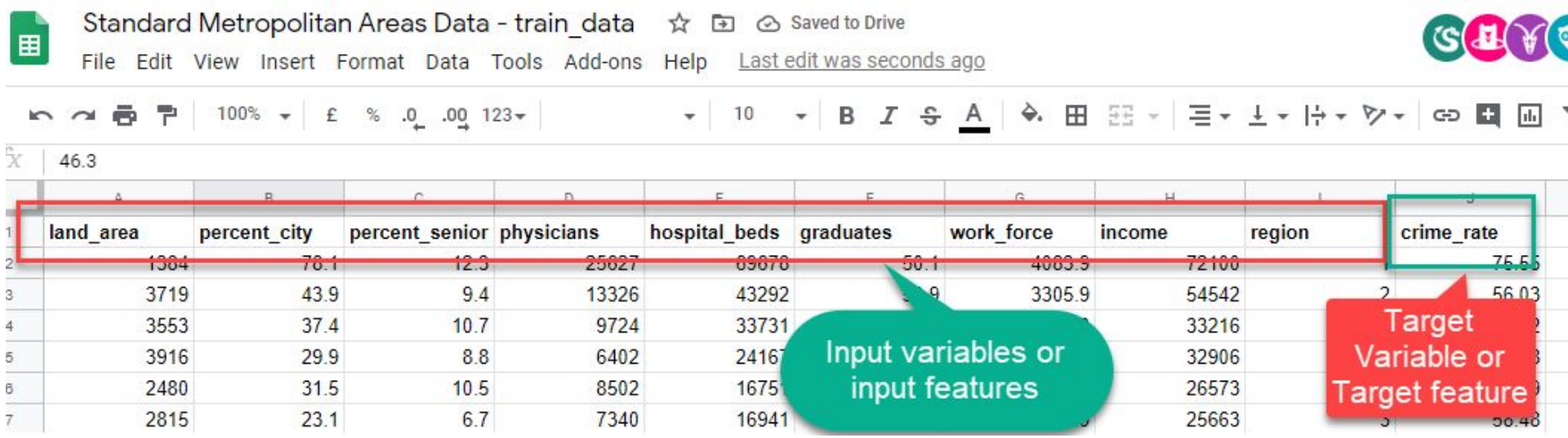
Predictor/Input Variables/Features

- One or more variables that are used to determine (or predict) the 'Target Variable' are known as Input Variables. They are sometimes called Predictor Variable as well.

- In the spam detector example, the features could include the following:
  - words in the email text
  - sender's address
  - time of day the email was sent
  - email contains the phrase "congrats you won $1 billion - share your bank details."

Another Example

- Standard Metropolitan Areas Data: In this dataset **we might be curious to predict "crime_rate" in future**, so that becomes our target and rest of the variables become input variables or features for building a machine learning model.



DPhi

Another Example

- Consider the data about 99 standard metropolitan areas in the US. The data set provides information on 11 variables for each area for the period 1976-1977. The areas have been divided into 4 geographic regions: 1=North-East, 2=North-Central, 3=South, 4=West. The variables provided are listed in the table below:

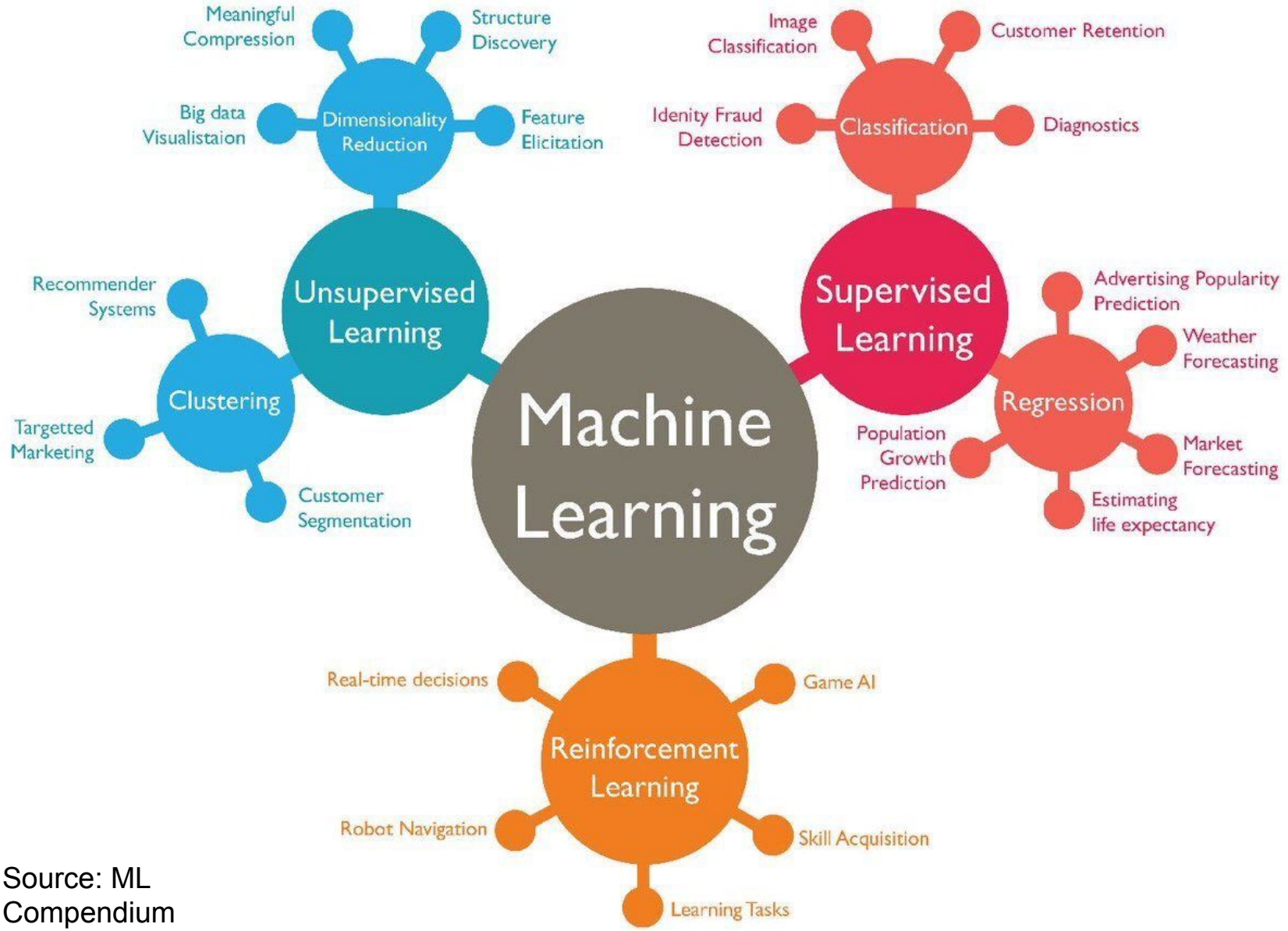
Variable name	Description
land_area	size in square miles
percent_city	percent of population in central city/cities
percent_senior	percent of population ≤ 65 years
physicians	number of professionally active physicians
hospital_beds	total number of hospital beds
graduates	percent of adults that finished high school
work_force	number of persons in work force in thousands
income	total income in 1976 in millions of dollars
crime_rate	Ratio of number of serious crimes by total population
region	geographic region according to US Census

Objective: Create a predictive model for crime_rate.

Types of Machine Learning

- **Supervised Learning**
  - **Classification**
  - **Regression**
- **Unsupervised Learning**
- **Reinforcement Learning**

Types of Machine Learning



Source: ML Compendium

DPhi

Supervised Learning

- Makes machine learn explicitly
- Data with clear defined output is given
- Direct feedback is given
- Predicts outcome/ future
- Resolve classification and regression problem

DPhi

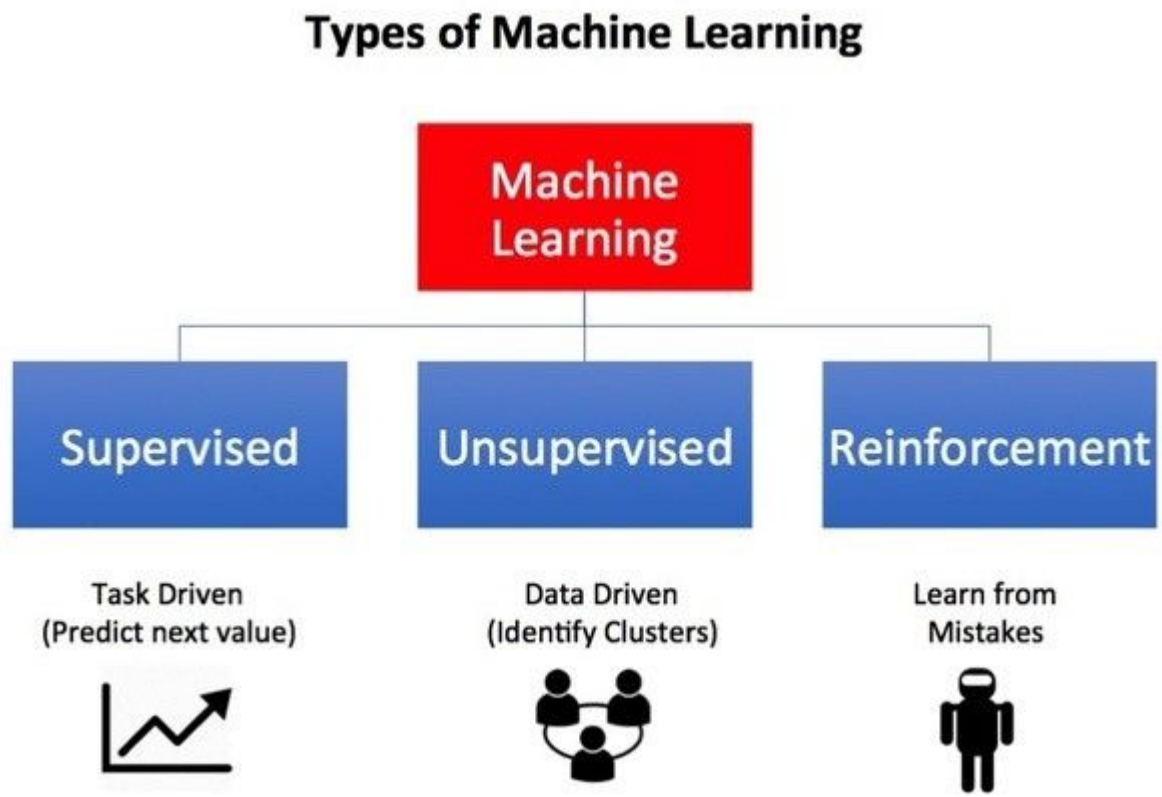
Unsupervised Learning

- Machine understand the data (identifies patterns/ structures)
- Evaluation is qualitative or indirect
- Does not predict or find anything specific

Reinforcement Learning

- An approach to AI
- Reward based learning
- Learn from positive and negative reinforcement
- Machine learns how to act in certain environment
- To maximize reward or minimize punishment

Summary



DPhi

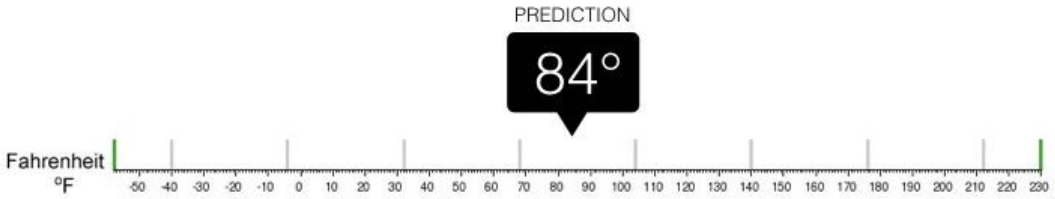
Supervised Machine Learning

Supervised Learning

Types of Problems

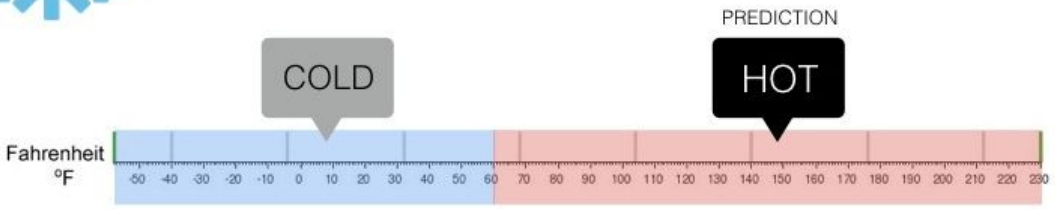
Regression

What is the temperature going to be tomorrow?



Classification

Will it be Cold or Hot tomorrow?



Source: Pinterest

DPhi

Supervised Learning

Classification:

- Classify the outcome
- **Examples:**
  - Predict whether a transaction is fraud or not fraud
  - Predict whether to give loan or not
  - Predict whether to give college admission or not
  - Predict the grade (Grade A, B, C, D)
  - Note: Classification can be more than two

Regression

- Regression is the problem of predicting a continuous outcome (a numeric outcome)
- **Examples:**
  - Predict house price
  - Predict crime rate

CLASSIFICATION VS REGRESSION

Student Profile

CLASSIFICATION

Predicting Student

Pass Or Fail

Student Profile

REGRESSION

Predicting Student Marks

Percentage

@ dataaspirant.com

Workflow

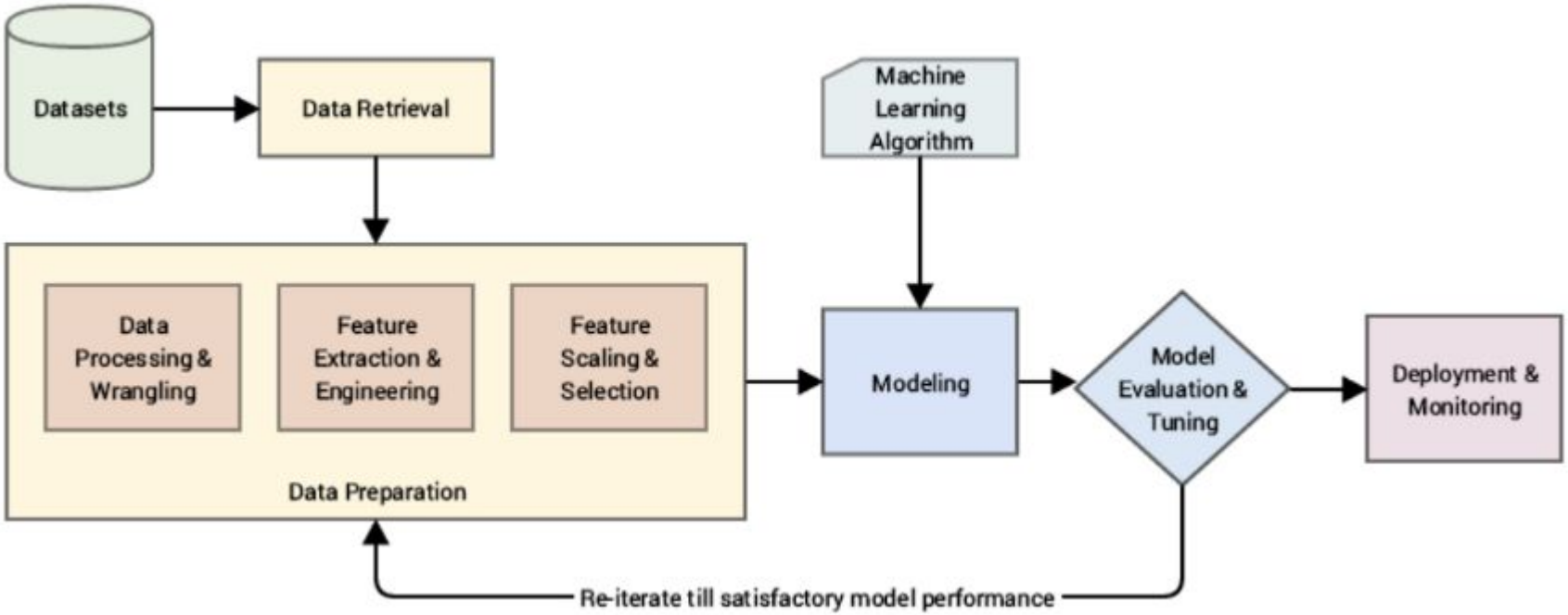


image credits: practical machine learning

DPhi

Problem Solving

- **Define Objective or understand the problem statement**
- Data Requirements
- Data Collection
- **Exploratory Data Analysis**
- **Data Pre-processing**
- **Build a model**
    - Understand whether it is a regression or classification problem
- **Evaluate**
- **Optimise**
- **Production**
- **Monitor**
- **You keep Optimising it every now and then**

Objective/Problem Statement

The goal of the model is to **predict whether a patient has liver disease or not**, given a set of data points.

DPhi

Data Requirements & Collection

We have the data!

DPhi

Understanding the Data

- Age of the patient
- Gender of the patient
- Total Bilirubin
- Direct Bilirubin
- Alkaline Phosphotase
- Alamine Aminotransferase
- Aspartate Aminotransferase
- Total Protiens
- Albumin
- Albumin and Globulin Ratio
- liver_disease: 1 --> Have liver disease, 0 --> No liver disease

Slide Download Link

You can download these slides from the below link:

https://docs.google.com/presentation/d/1l5ZqKDYVCPSlgCg57pNMhF4GxsBd_KXhCM0o61Hu_O8/edit?usp=sharing

DPhi

That’s it for the day. Thank you!

We can stay connected:

Linkedin: https://www.linkedin.com/in/chanukyapatnaik/

Twitter: @chanukya_p

Medium: @chanukya

DPhi

In today's hands on session

Fundamentals of Python for Data Science/ML

DPhi

How to procure tomatoes?

Farming

Buy from grocery

DPhi

Libraries Analogy

Farming is like write code from scratch

Buy from grocery store is like using libraries

DPhi

Explore the data

Let’s get to the notebook

Notebook Link:
https://github.com/dphi-official/Machine_Learning_Bootcamp/blob/master/Python_Fundamentals_for_Data_Science.ipynb

DPhi

What are we achieving out of it?

- Technically an efficient model can **empower doctors who are treating Liver Patients**
- **Avoid Errors:** Feeding in the patient records could help assess doctors to understand the significance of the disease.
- Practical Advanced Use-Case: Parkinson Disease using Kinect
- What is Parkinson Disease? Disorder of the central nervous system that affects movement. Patients often visit doctors in a few months and their concerns are usually influenced by recency factors.

ML/Data Science can empower people to arrive at a meaningful decision using Data.