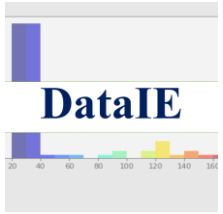
Introduction to Machine Learning

Presented by: TANMOY DAS



Introduction to machine learning

LIVE

Date: 21st and 22nd Octover

Time: 9:00PM -10:00PM

TANMOY DAS

Industrial Engineer & Data Scientist,
Ph.D. Research in Training, Dalhousie University, Canada

Machine Learning
Day 1

Machine Learning: Day 1

What is ML & what is NOT ML?

Technical Details

Q&A

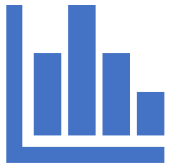
Disclaimer

- None of today's content is ABSOLUTE truth! It's all about perspective!!
  - R programmers are data scientist!!
  - Python vs Excel?

Machine Learning

Intro

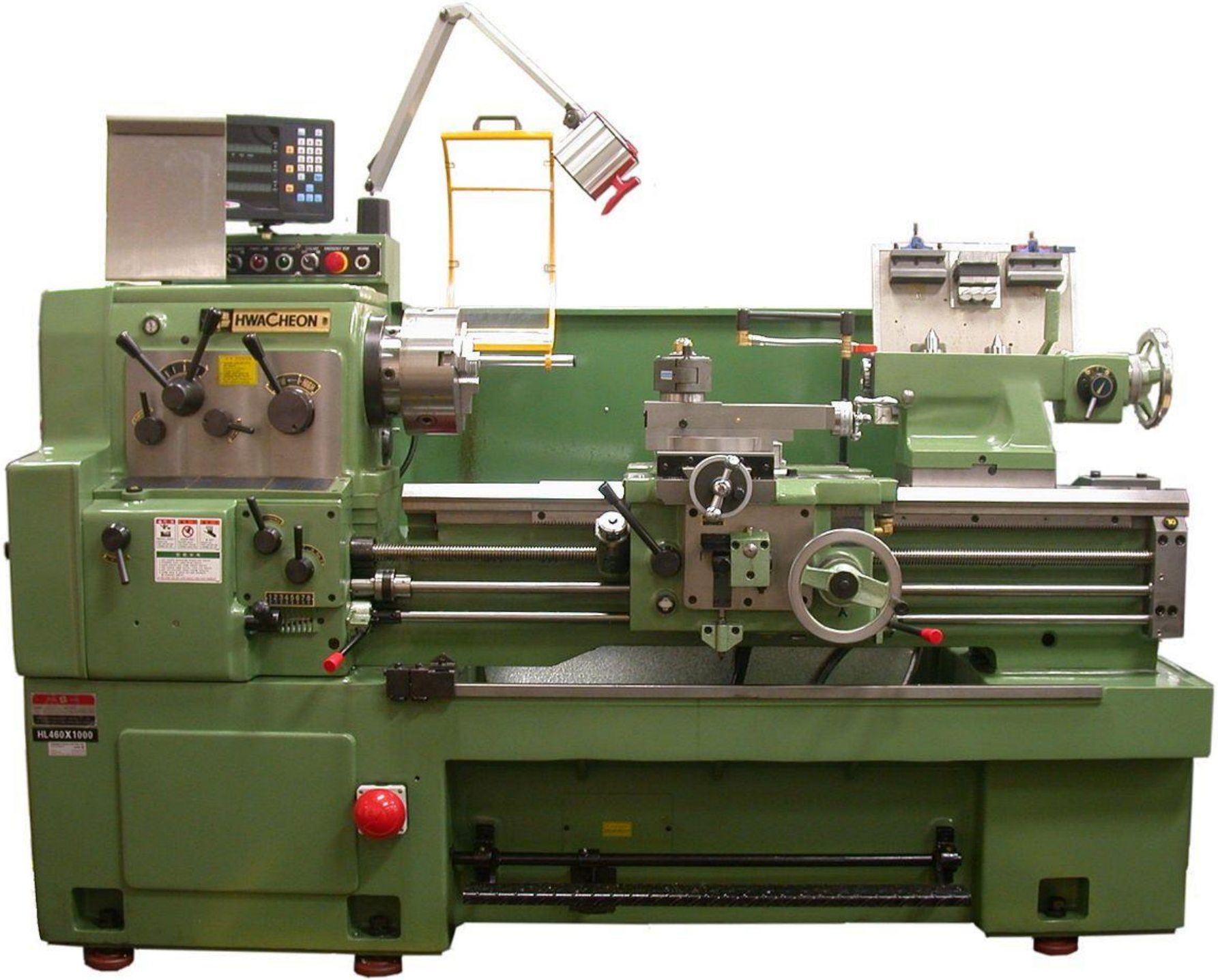
What is Machine Learning



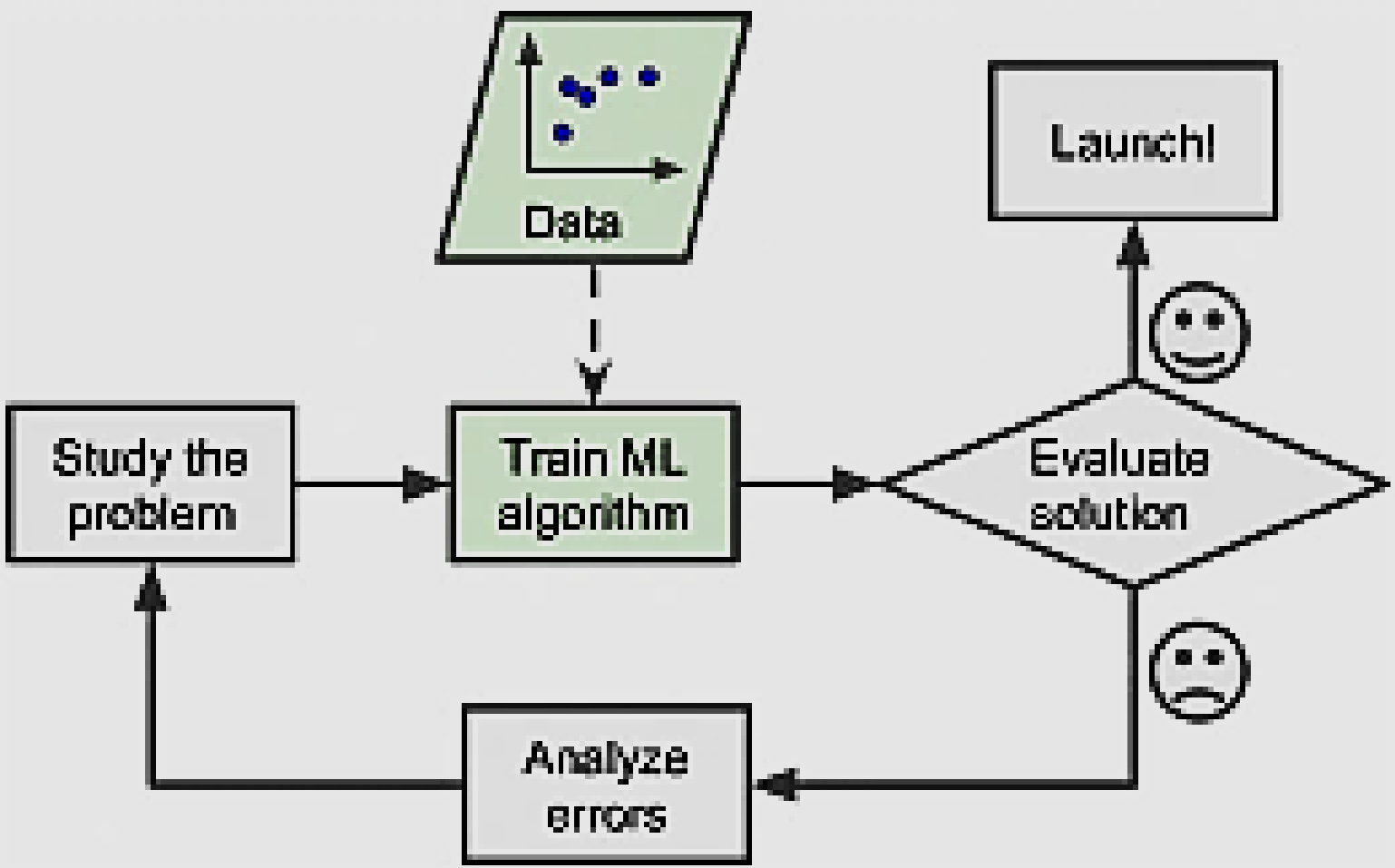
Learning from Data

What is Machine??

- **Machine learning** (ML) is the study of computer algorithms that improve automatically through experience.
  - Source: Wiki

- 
HWACHEON
HL460X1000

Machine Learning Approach



Source: Hands-On-Machine-Learning-with-Scikit-Learn-Keras-and-Tensorflow_-Concepts-Tools-and-Techniques

What is Machine Learning?

Learn from experience

Learn from ~~experience~~ data

Follow instructions

0:45 / 30:52

Example of Machine Learning

- Virtual Personal Assistants. ...
- Predictions while Commuting.
- Videos Surveillance. ...
- Social Media Services. ...
- Email Spam and Malware Filtering...
- Online Customer Support. ...
- Search Engine Result Refining
  - Source

Why we talk about Machine Learning in 2020?

- Computational power
  - Google Scholar
  - Amazon AWS
  - Supercomputer rent

Machine Learning vs Statistics

	Statistics	Machine Learning
Approach	Data Generating Process	Algorithmic Model
Driver	Math, Theory	Fitting Data
Focus	Hypothesis Testing, Interpretability	Predictive Accuracy
Data Size	Any Reasonable Set	Big Data
Dimensions	Used Mostly for Low Dimensions	High Dimensional Data
Inference	Parameter Estimation, Predictions, Estimating Error Bars	Prediction
Model Choice	Parameter Significance, In-sample Goodness of Fit	Cross-validation of Predictive Accuracy on Partitions of Data
Popular Tools	R	Python
Interpretability	High	Low

Article worth reading:
https://www.nature.com/articles/nmeth.4642

Source: https://towardsdatascience.com/how-is-machine-learning-different-from-statistics-and-why-it-matters-5a8ed539976

Eric W
I learn
6d •

5 decisions I've

No regret:
1. Trying to mas
2. Designing ex
3. Knowing whe
4. Asking "why"
5. Investing in p

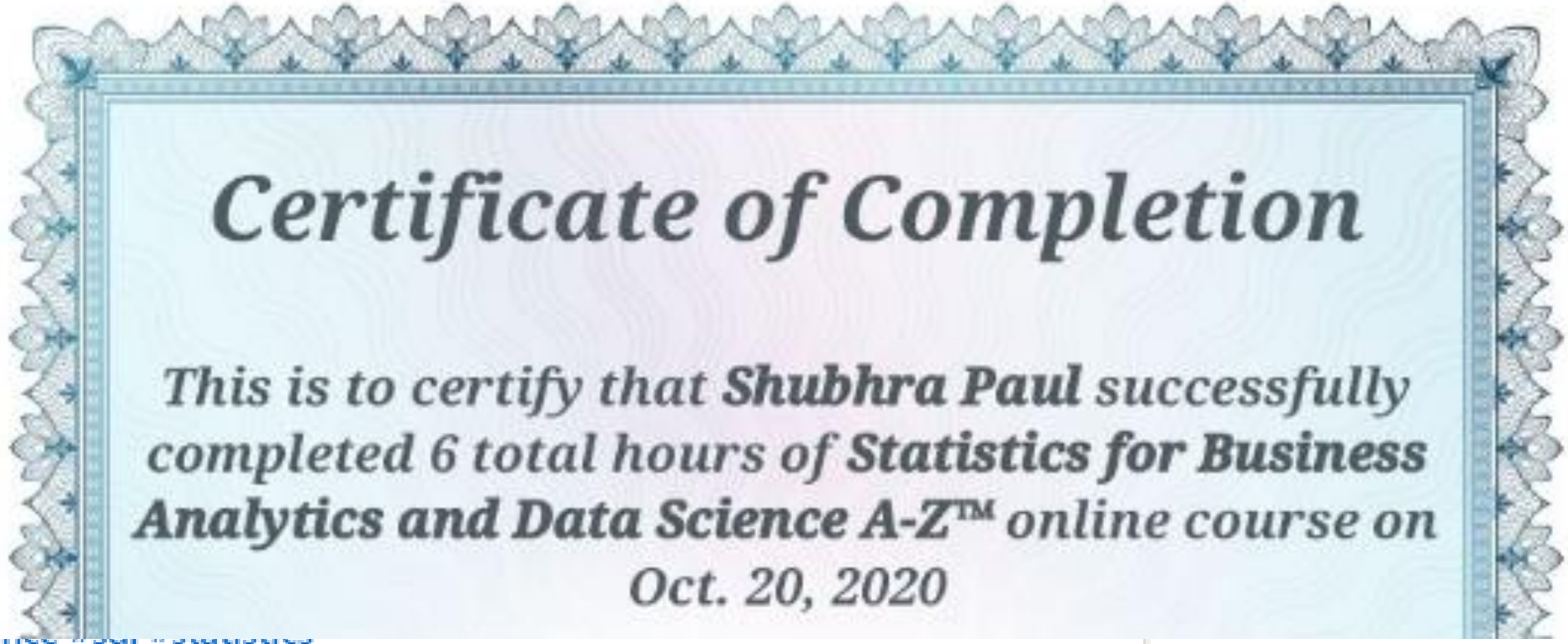
Regretted:
1. Thinking I ne
2. Prepping for
3. Trying to emu
4. Focusing on
5. Learning all a

What decisions

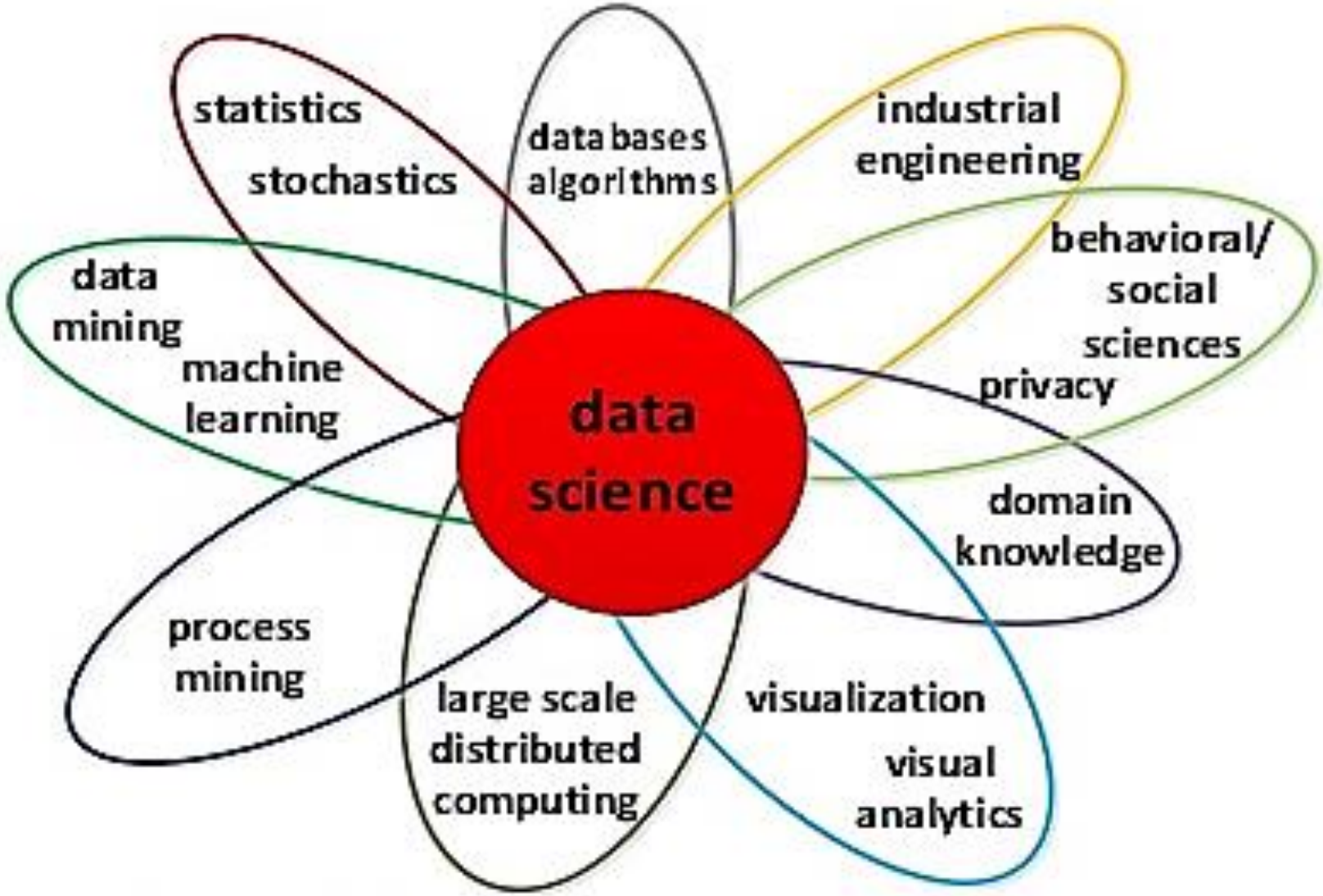
#data #datascience #sql #statistics

Shubhra Paul • 1st
Ph.D. Candidate | Data Science | Operation Research | Supply Chain | Optimization | ...
22h • Edited •

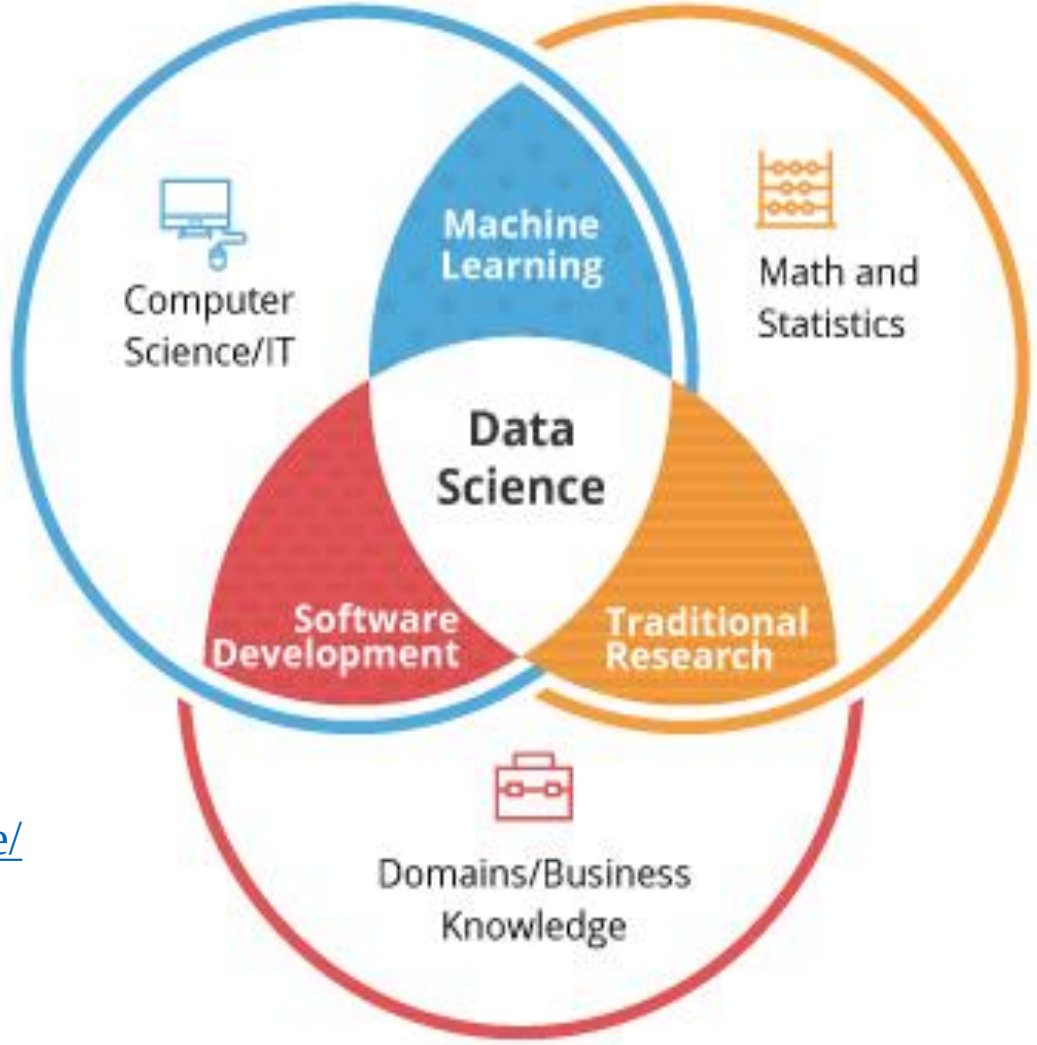
Statistics is one of the key elements for Data Analytics, Data Science, and Machine Learning. I just revised my statistics skills with this course.
#dataanalytics #machinelearning #datascience #interships #statistics #statisticalanalysis #businessanalytics #elearning #udemy



Data Science



Data Science vs Machine Learning



Source: https://www.zeolearn.com/magazine/data-science-vs-machine-learning-artificial-intelligence (Show from this link)

Data Scientist vs Machine Learning Engineer

Writer vs Professor


YOU NEED
YOU NEED
MATHS
MATHS
TO DO MACHINE LEARNING
TO DO MACHINE LEARNING
YOU NEED MATHS TO DO MACHINE LEARNING
YOU DONT NEED TO WORRY ABOUT IT

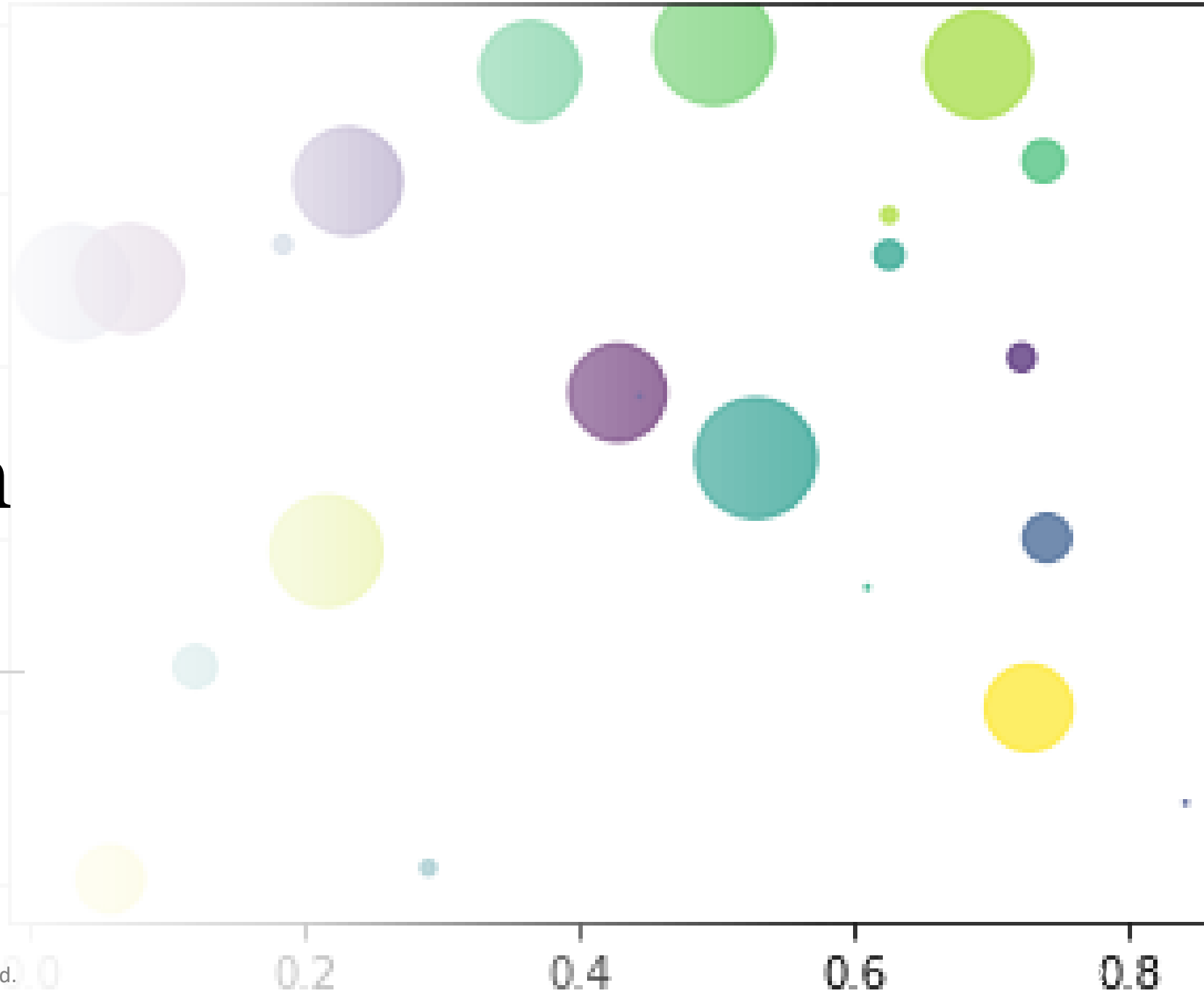
Model Accuracy or Performance??

Fraud detection (1 fraud, 99 no fraud)

Scatter Plot in Python

Documentation: plt.scatter

Source code to run



Kaggle

- How to learn Data Science using Kaggle?

Profile of Tanmoy Das

LinkedIn

Kaggle

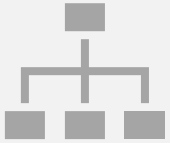

Q&A

Technical Details

Machine Learning

Types of problems in ML

Regression



Classification

Clustering

Explanatory and response variables

X = input variable, feature

y = output variable, target

0, 1, 5, 1, 0,
0, 0, 0, 2, 0,
0, 3, 0, 1

Rainfall

Which question is MOST important?

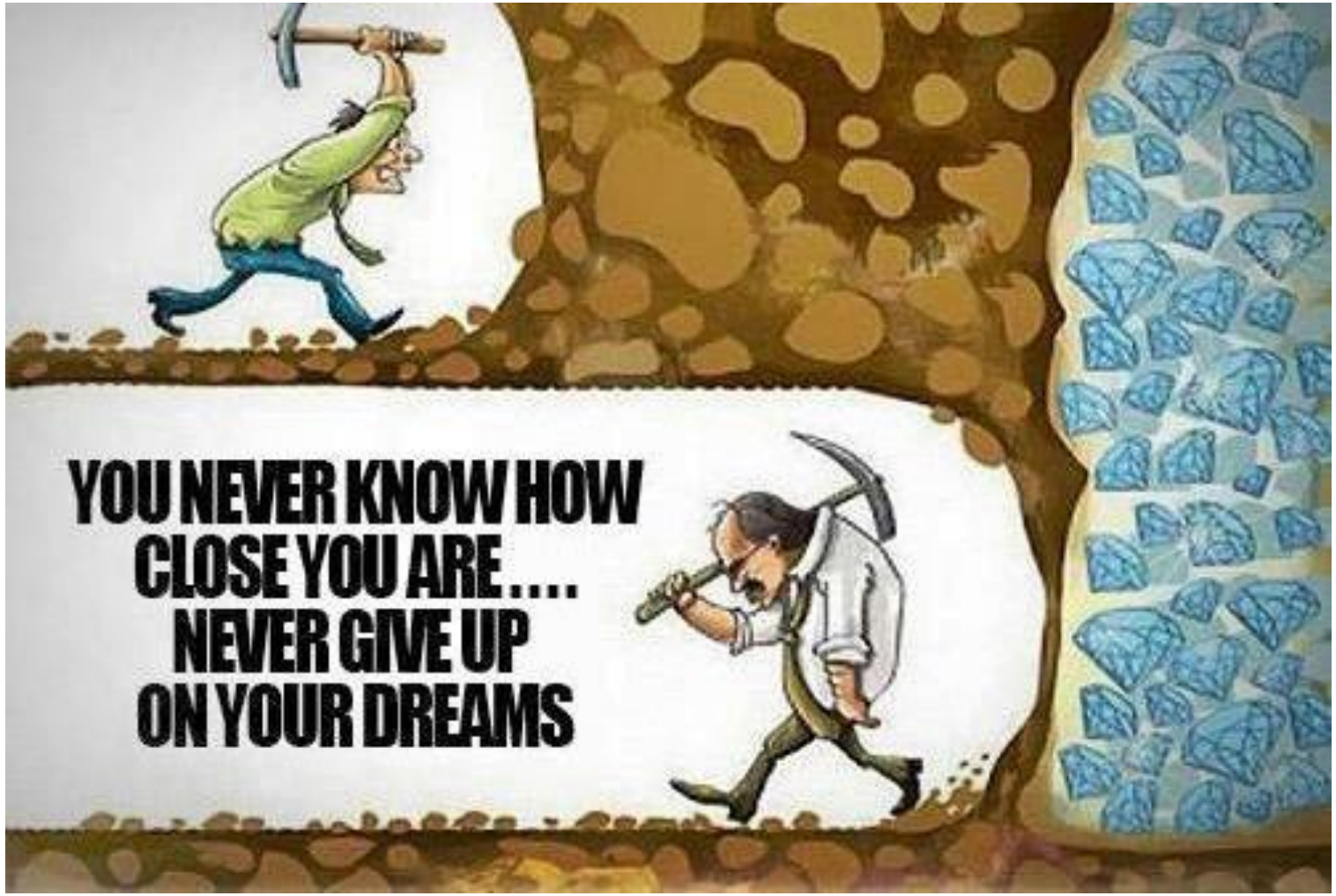
Binary qualitative discrete response variable

- **What'd be the amount of rainfall tomorrow?**
- **What's the mean rainfall?**
- **Will it rain tomorrow?**

Zero-inflated data over-dispersion

Regression vs Classification

- Predicting the amount of rain:
  - continuous value?
- Predicting whether there will be rain or not:
  - Discrete value?
    - What the heck is discrete and continuous variable??
- Source:
  - Business statistics by Linde (Show google Search)
  - MachineLearningMastery
  - Medium



Never Give up!

What about your foundation on math & stat?

Types of ML algorithm

Supervised Learning

Semi-supervised Learning

Unsupervised Learning

Reinforcement Learning

Name of ML algorithm

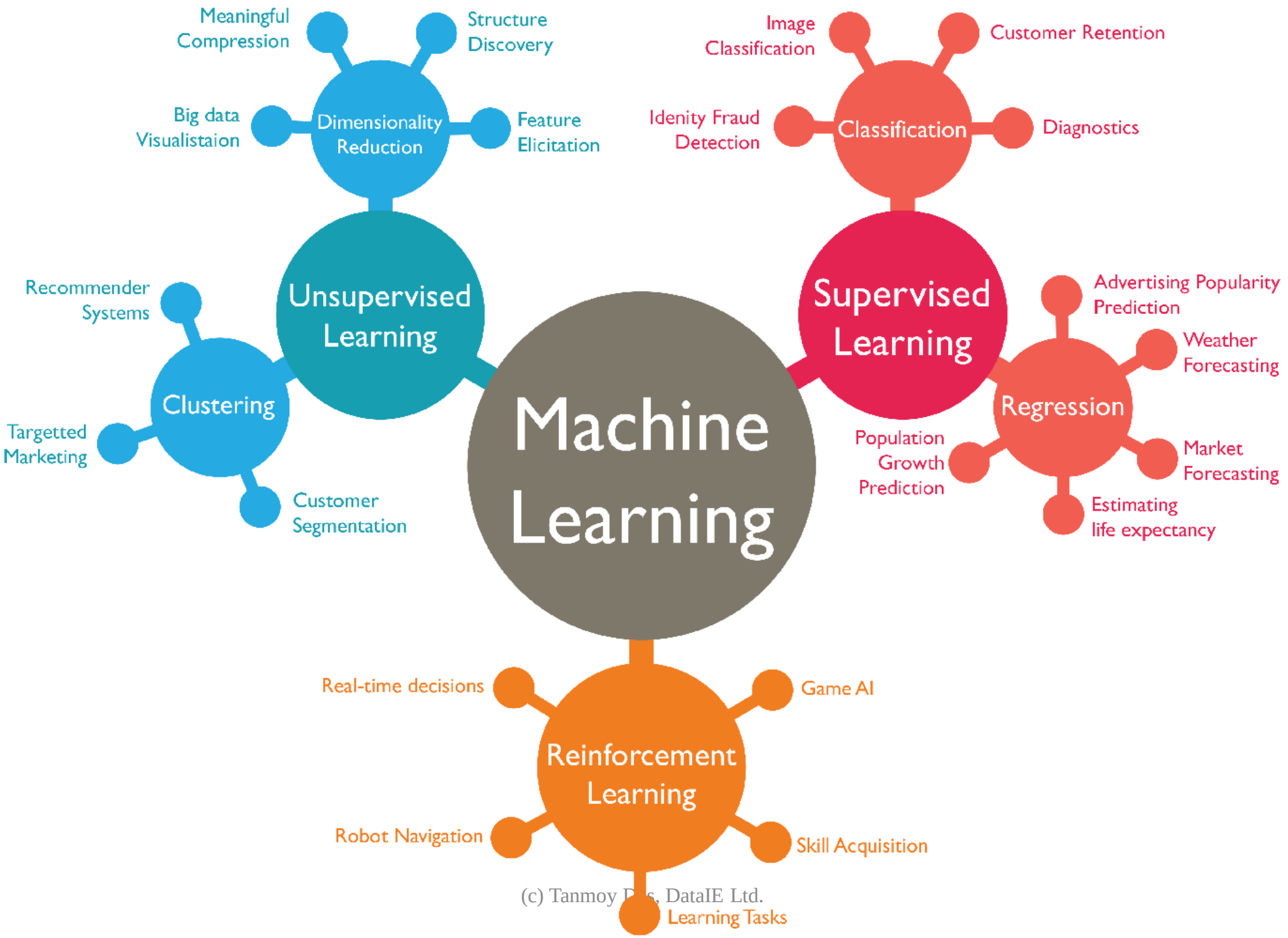
Regression

- Linear Regression
- Support Vector Regression

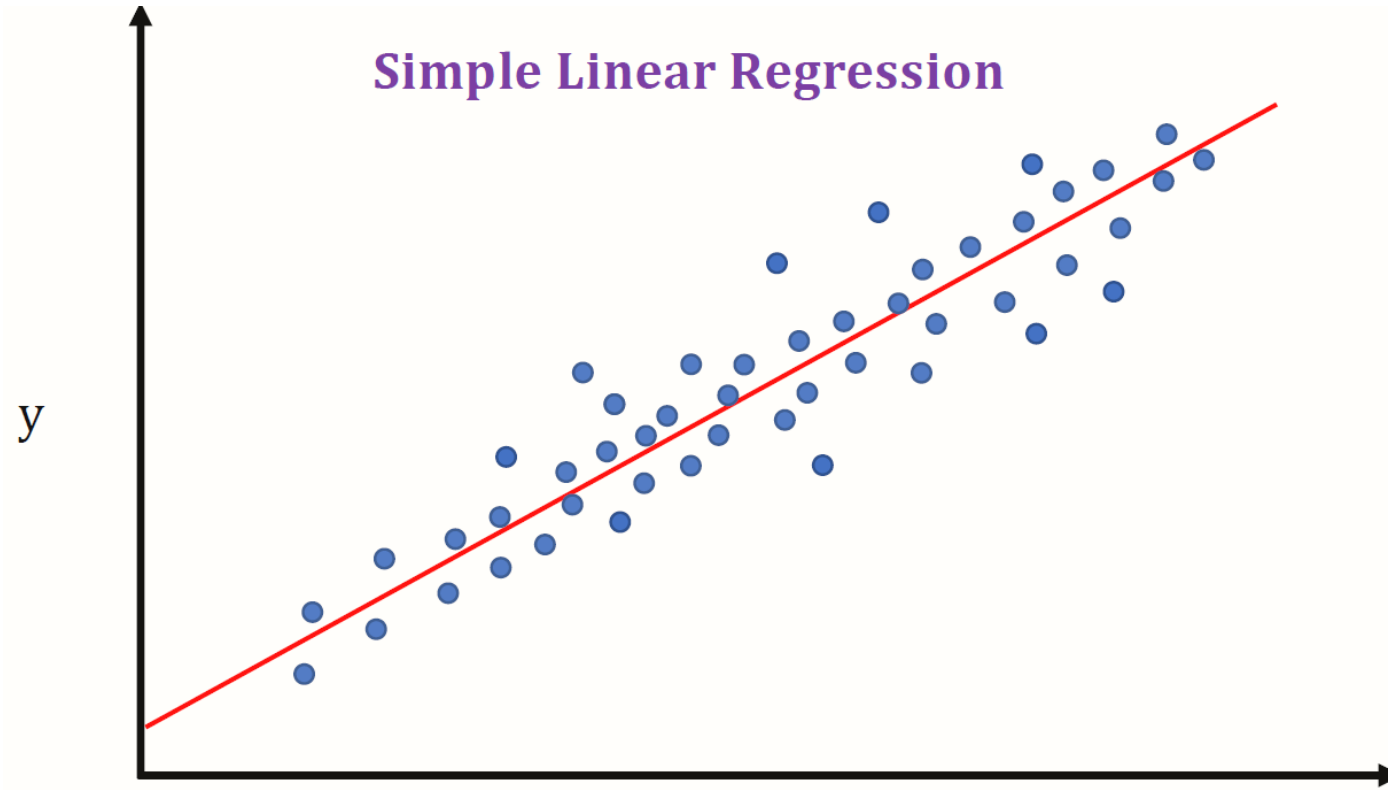
Classification

- Logistic Regression
- Kmean

Different ML algorithms



Linear Regression and Linear Classification



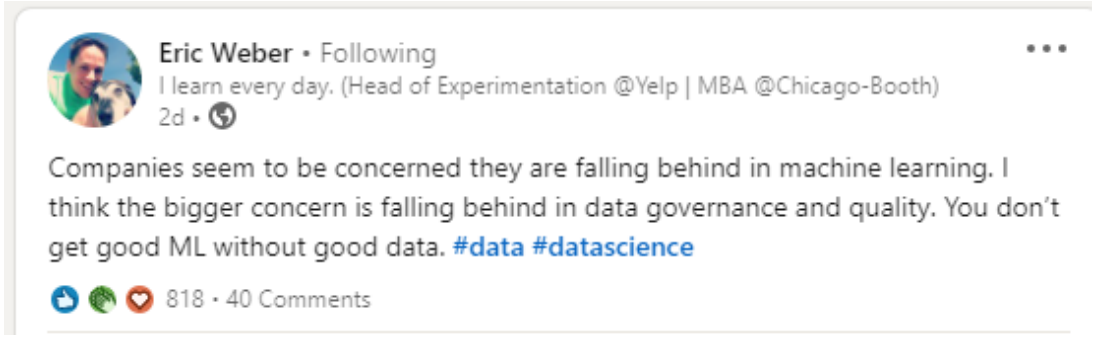
- Linear Regression
- Linear Classifier
  - SVM

You may ignore Deep Learning, but Linear Regression!

Linear Regression vs Deep Learning

Follow Data Science influencers

- Eric Weber



Prior knowledge of statistics, probability theory, calculus and linear algebra is strongly recommended..

Linkedin/ Github/ Kaggle: Connections/ Recommendations

- https://github.com/tanmoyie
- Kaggle scoring
  - IE:- Where to start Data Science as an Industrial Engineer

Python or R?

Sources to learn Machine Learning

Book

- Practical Statistics for Data Scientist
- Data Science for Dummies
- ISLR/ ESL
- Machine Learning for Dummies

Online Courses

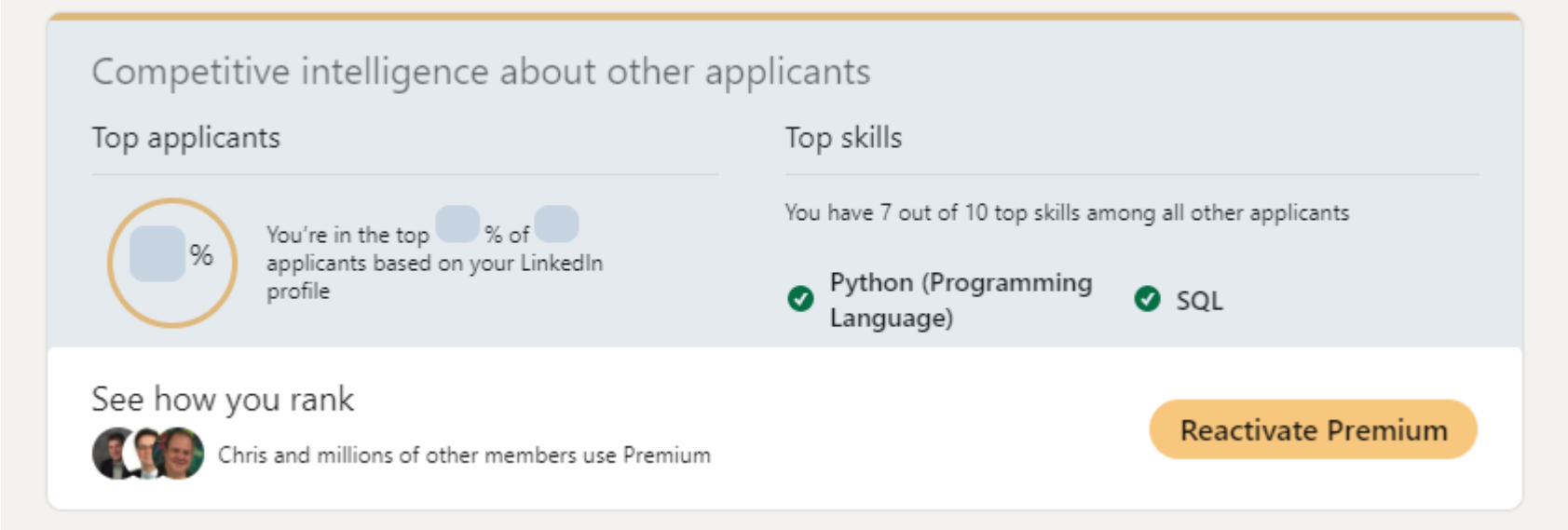
- Machine Learning A-Z on Udemy
- Machine Learning for Everyone on DataCamp
- Yes, I DO NOT recommend courses by Dr. Andrew NG for beginners in ML. Don't get me wrong. If you have a solid foundation of statistics, his courses will be super **helpful** for you. But, for beginner, NO.

Popular blogs

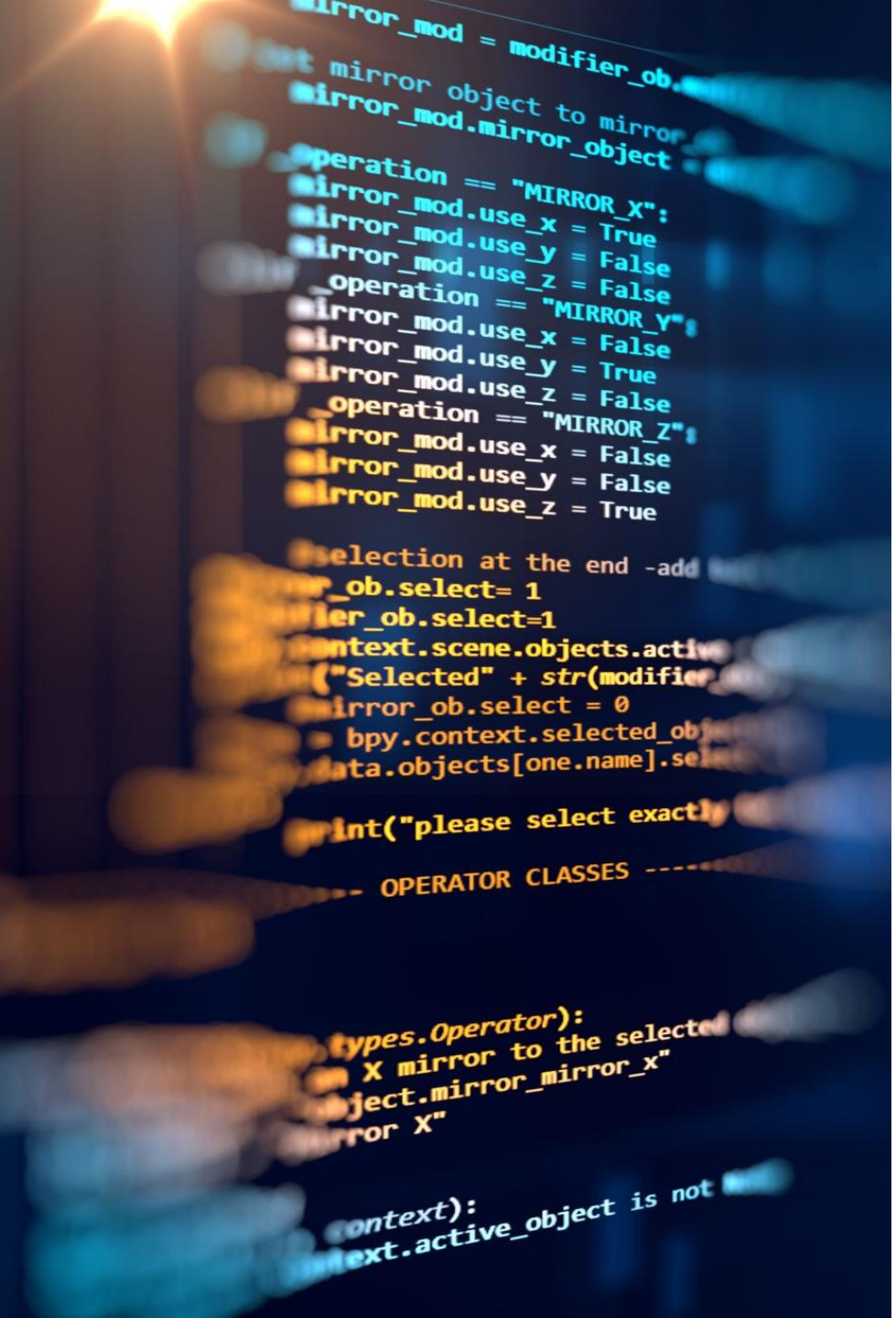
- https://towardsdatascience.com/
- machinelearningmastery.com

Interview questions/ Job circular

- 51 ML interview questions
- Job circular, 7/10

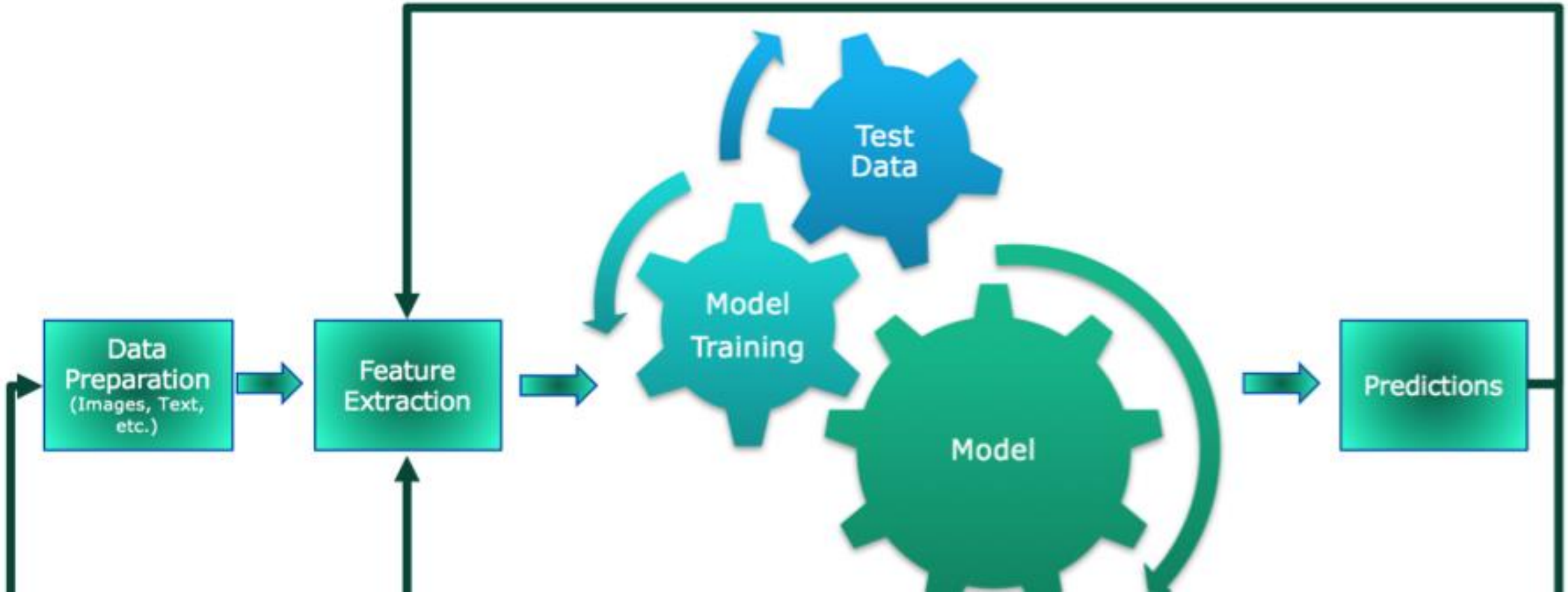


Linear Regression project



Linear Regression implementation in Python

- https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html#sphx-glr-auto-examples-linear-model-plot-ols-py



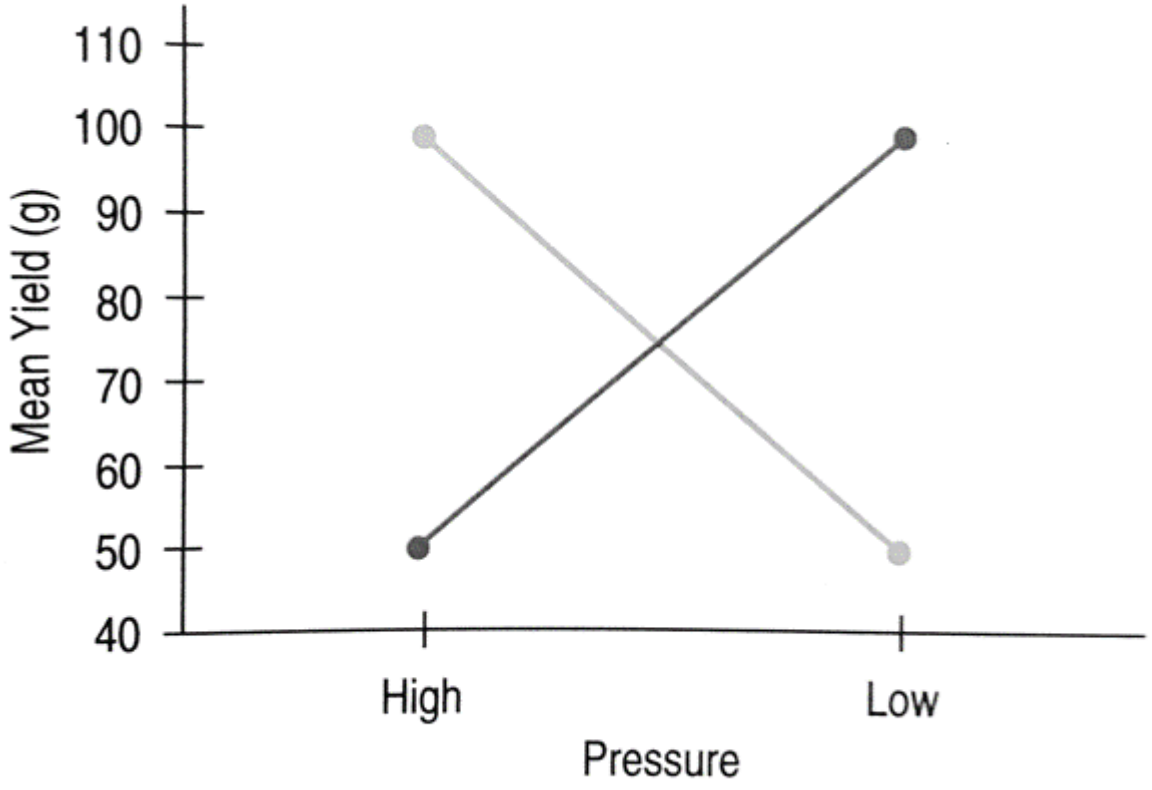
A Machine Learning pipeline

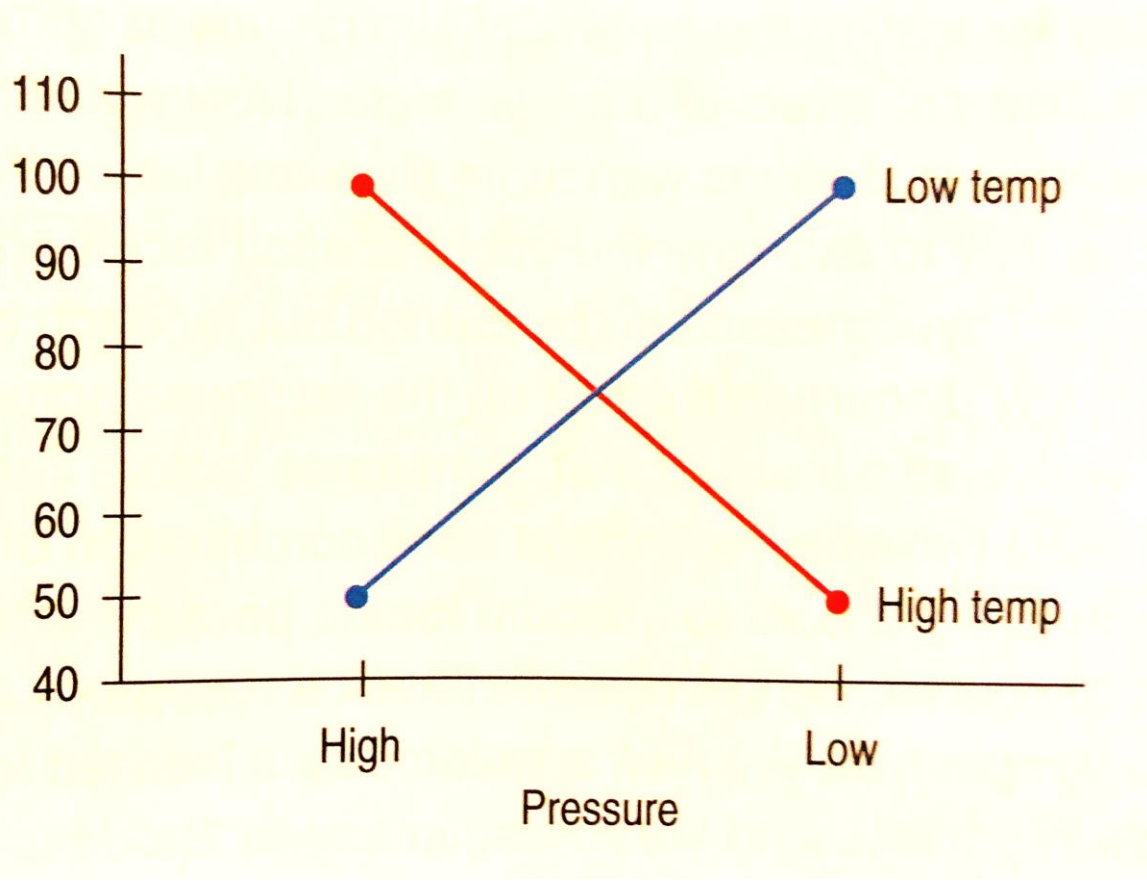
Google Search | Image Source

Kmean vs KNN

- **k-Means** Clustering is an unsupervised learning algorithm that is used for clustering
- **KNN** is a supervised learning algorithm used for classification

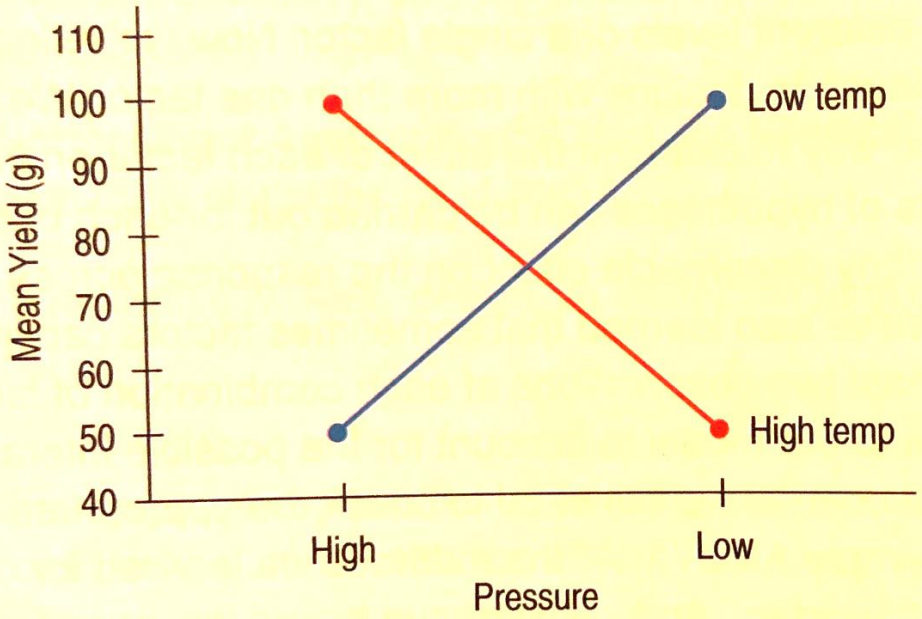
Why we need statistics in Machine Learning: an example





Source: Stats: Data and Models by Richard D. De Veaux

- ***When the interaction effect is significant, don't interpret the main effects.*** Main effects can be very misleading in the presence of interaction terms. Look at this interaction plot.



An interaction plot of *Yield* by *Temperature* and *Pressure*. The main effects are misleading. There is no (main) effect of *Pressure* because the average *Yield* at the two pressures is the same. That doesn't mean that *Pressure* has no effect on the *Yield*. In the presence of an interaction effect, be careful when interpreting the main effects.

Figure 29.12

The experiment was run at two temperatures and two pressure levels. High amounts of material were produced at high pressure with high temperature and at low pressure with low temperature. What's the effect of *Temperature*? Of *Pressure*? Both main effects are 0, but it would be silly (and wrong) to say that neither *Temperature* nor *Pressure* was important. The real story is in the interaction.

Any last question?