Introduction to ML

Premkumar Chandra Shegaran

About the Speaker

Professional Experience

Schlumberger

avanade

accenture

WorleyParsons
resources & energy

HERIOT WATT UNIVERSITY

THE TIMES THE SUNDAY TIMES
GOOD UNIVERSITY GUIDE 2018
INTERNATIONAL UNIVERSITY OF THE YEAR

UCSI University

The Center of Applied Data Science

UDACITY

OXYGEN

Education


UNIVERSITI
TENAGA
NASIONAL


HERIOT
WATT
UNIVERSITY


THE TIMES
THE SUNDAY TIMES
GOOD
UNIVERSITY
GUIDE
2018
INTERNATIONAL
UNIVERSITY
OF THE YEAR


Georgia
Tech
College of
Computing


UDACITY


Microsoft


Schlumberger


HALLIBURTON

Seminars/Talks

Jabatan Perangkaan Malaysia

BANK NEGARA MALAYSIA
CENTRAL BANK OF MALAYSIA

MALAYSIA INSTITUTE OF STATISTICS

DATA✖ innovation enterprise

DEPARTMENT OF STATISTICS MALAYSIA, OFFICIAL PORTAL
The Source of Malaysia's Official Statistics

UTP
UNIVERSITI TEKNOLOGI PETRONAS

A P U
ASIA PACIFIC UNIVERSITY OF TECHNOLOGY & INNOVATION

UCSI University

Mentorship

Startup Station Singapore from facebook

In partnership with:

INFOCOMM MEDIA DEVELOPMENT AUTHORITY

Data Science Society
Data.Platform

UDACITY

Welcome to ALT.labs Series 12

About

- Monthly community event driven by passionate people to thrive in industry 4.0
- Addresses the on-demand skillsets and technologies needed to survive in the current revolution: IOT, Cloud Computing, Robotics, AI&Big (Extreme) Data, and Cyber Security
- To empower Malaysia to become a tech-driven state
- To share and exchange knowledge, solve real world problems and most importantly, create an AWESOME community and to DISRUPT

What you need to do..

- Support us and the community by participating in the monthly co-labs
- Join our LinkedIn site: https://bit.ly/2PX5qeR

Introduction to Supervised and (very basic) Unsupervised Learning

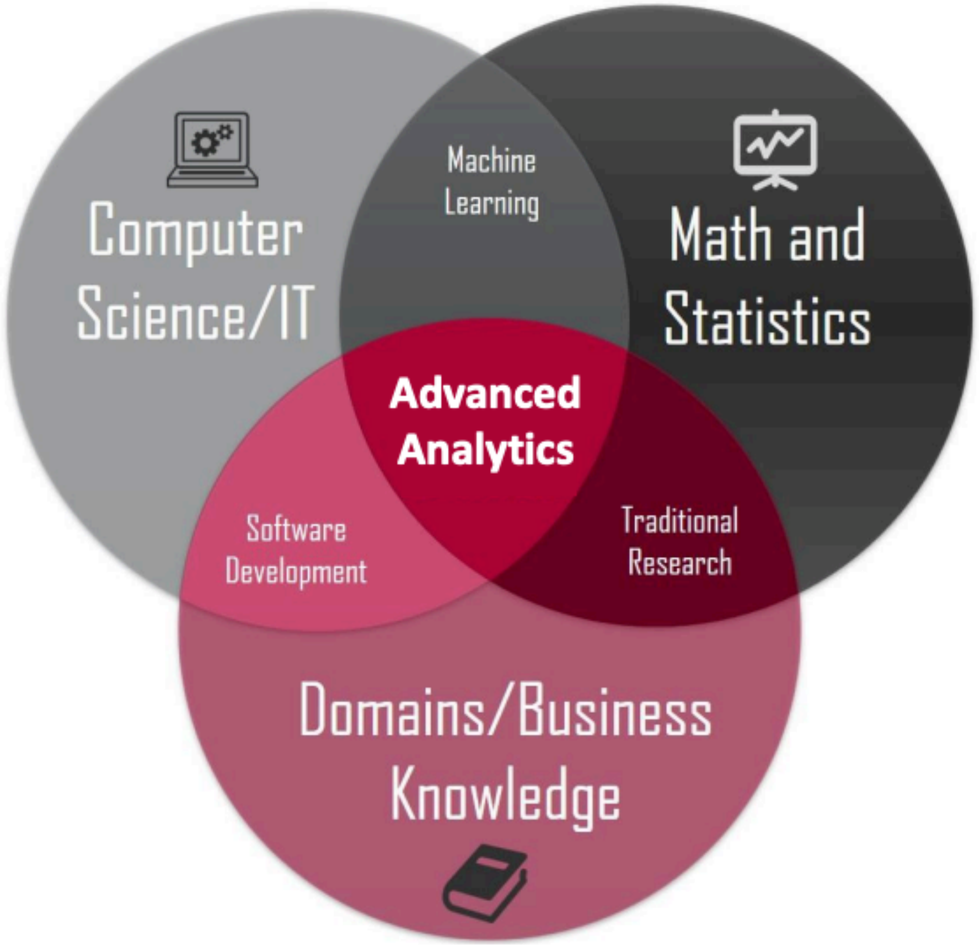
Download Course Materials

- Github page: https://bit.ly/2JyDTz5

Introduction to Machine Learning

“[Machine Learning is the] field of study that gives computers the ability to learn without explicitly being programmed.”

–Arthur Samuel, 1959



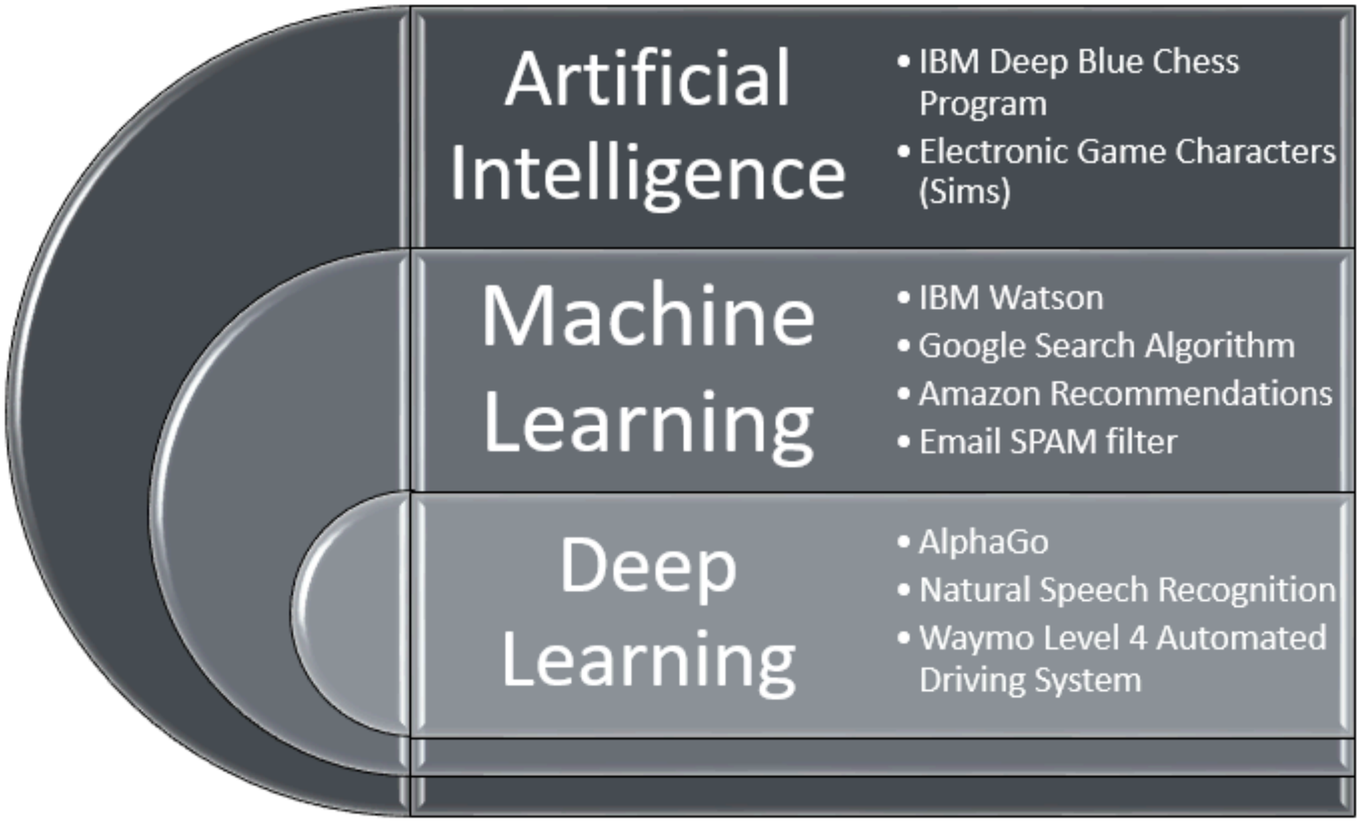
Drew Conway's perspective on data science:

- Computer Science
- Maths and Statistics
- Domain Expertise

Other perspectives:

- Simulation
- Experimentation
- Replication

AI, Deep Learning, Machine Learning?



Credit Card Fraud detection

How would you do it?

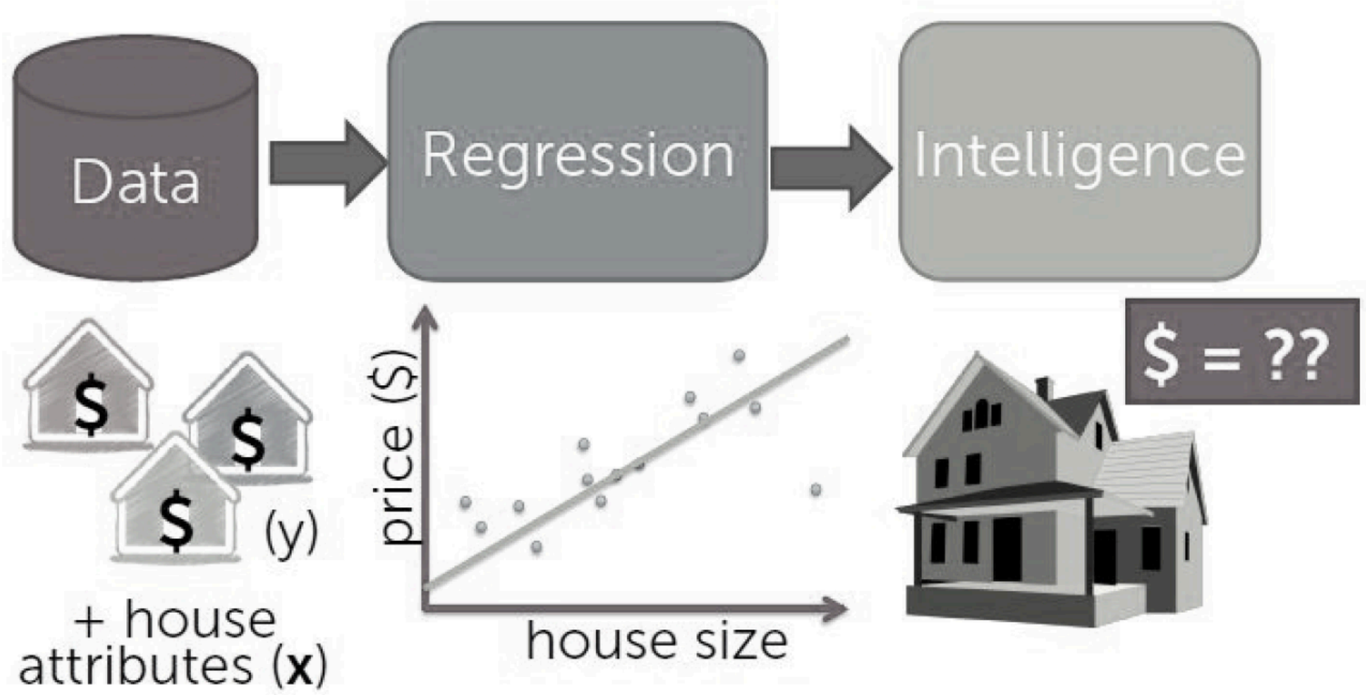
CREDIT CARD FRAUD

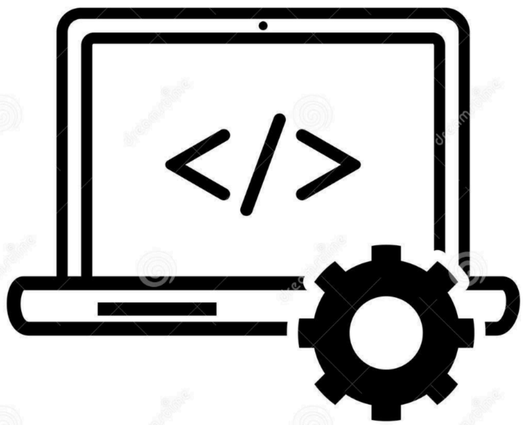
284,335

472

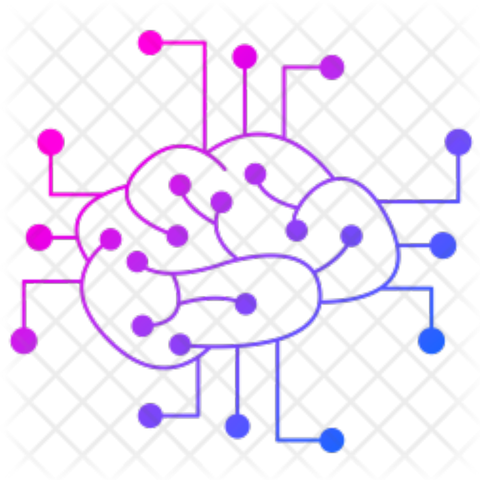
Housing price prediction

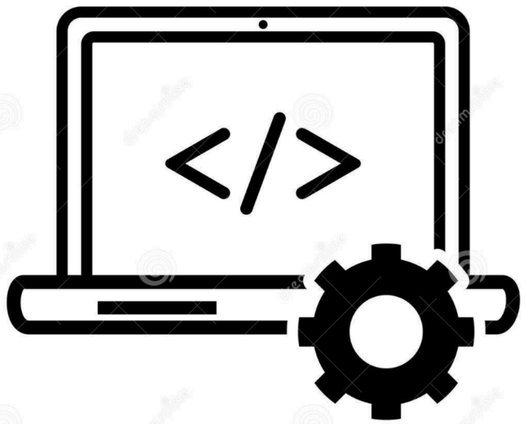
How would you do it?


Data
Regression
Intelligence
$
$
$
(y)
+ house attributes (x)
price ($)
house size
$ = ??



VS

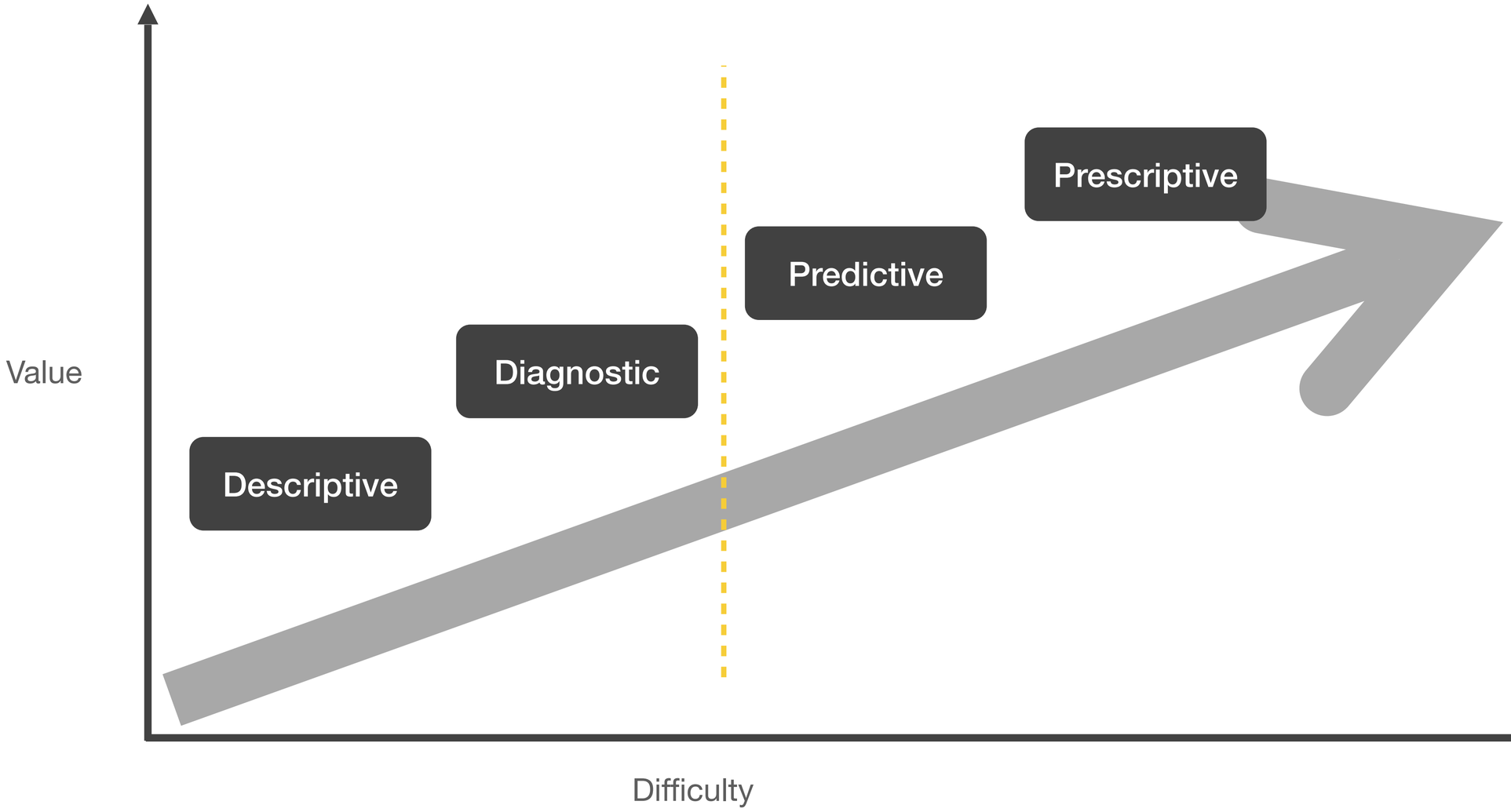




VS

Real Estate		**Chatbots, Investor analytics, Pending loan defaults, Deal matching, Property management**
Telco		**Monitor signal density, Churn prediction, Fraud detection, Improve service applications**
Marketing		**Customer segmentation, Sentiment analysis, Lead generation**
Medical		**Disease detection, Drug discovery, Medical bill estimation, Epidemic outbreak prediction**
Manufacturing		**Yield prediction, Smart manufacturing, Predicting supply chain**
Energy		**Upstream exploration, energy demand forecasting, Storage, Preventive maintenance**

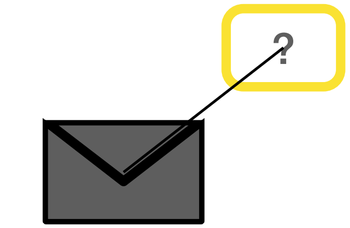
Analytic Value Escalator



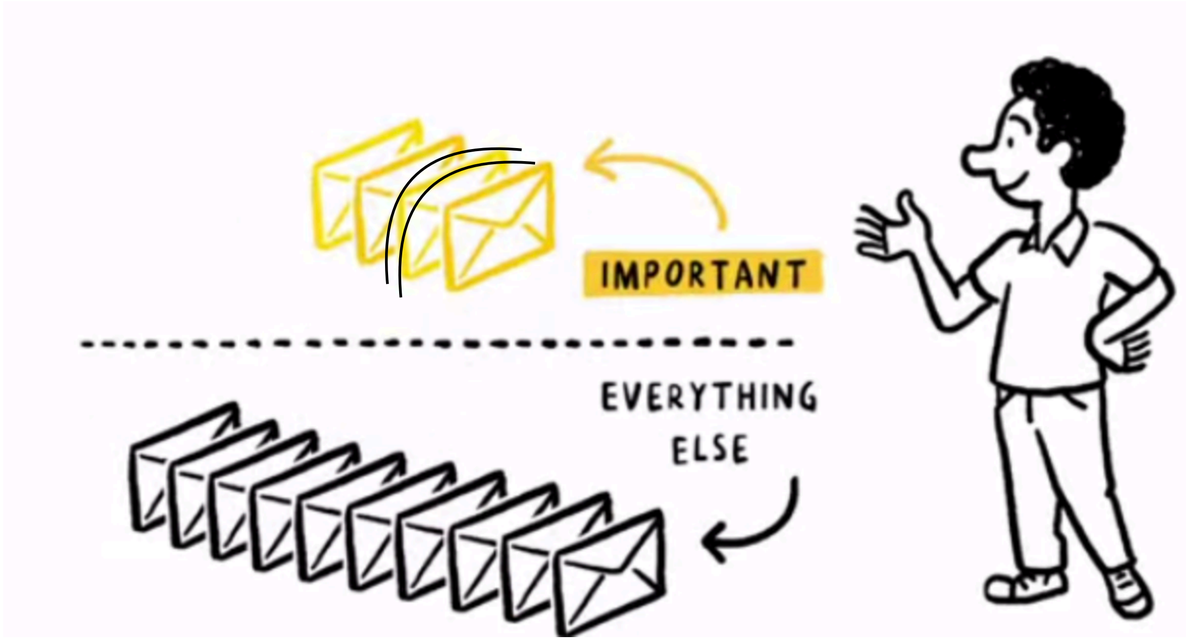
Types of Machine Learning

- **Trained with human intervention or not**: Supervised, Unsupervised,
- **Static or Dynamic systems**: Batch vs Online learners
- **Learn from patterns or compare with known data points**: Instance-based vs model-based

Supervised Learning

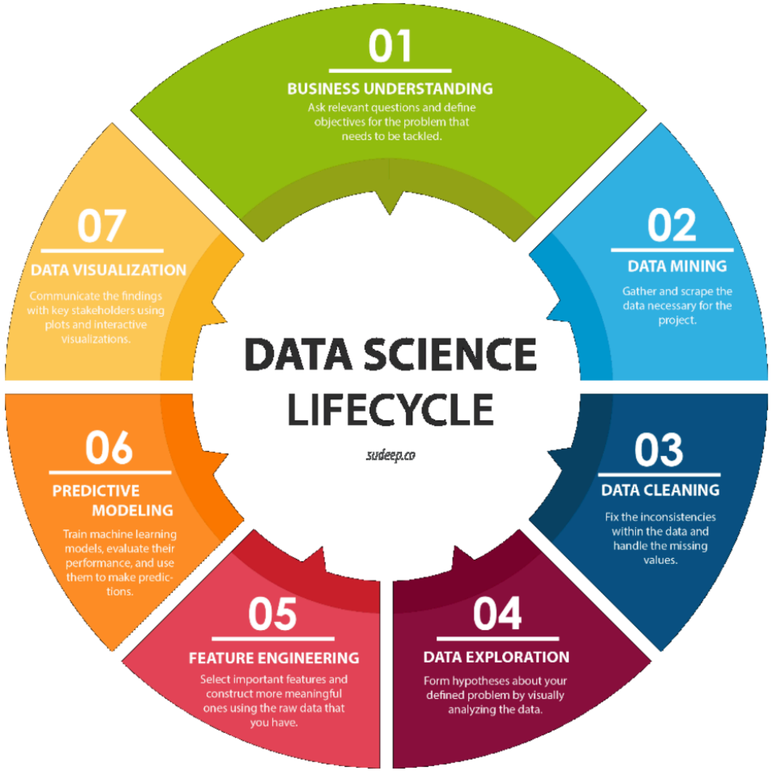


New Instance



Machine Learning Flow

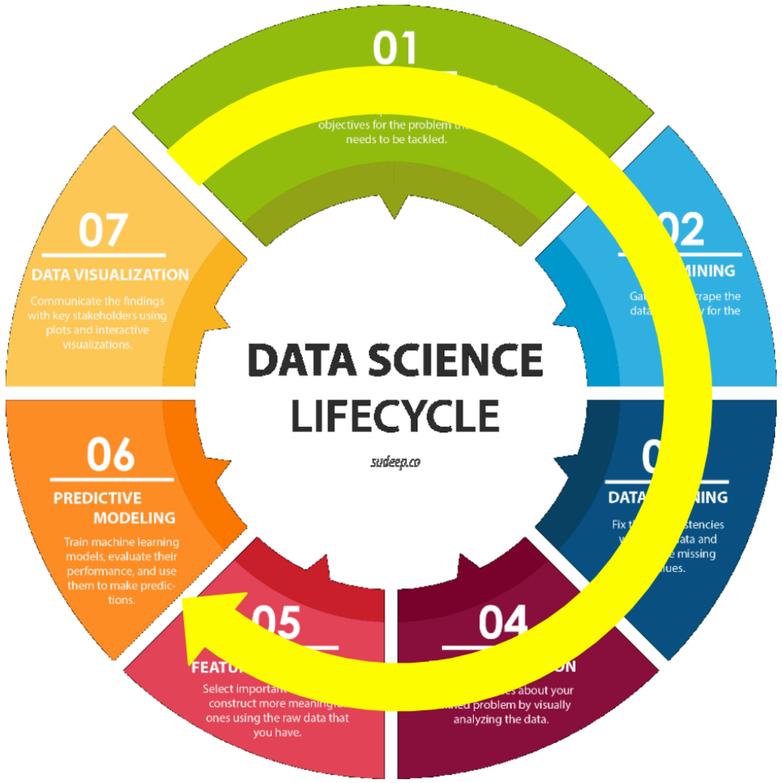
Typical Data Science Workflow



A typical workflow for a data scientist is:

1. Gather the requirements for the business problem
2. Identify data useful for the case
   - Ingest data
3. Cleanse data into a useful format
4. Analyze data
5. Prepare input for your algorithms
6. Execute data science algorithms (ML, AI, etc.)
   - Iterate 2-6 until valuable insights are produced
7. Visualize and share

Typical Data Science Workflow



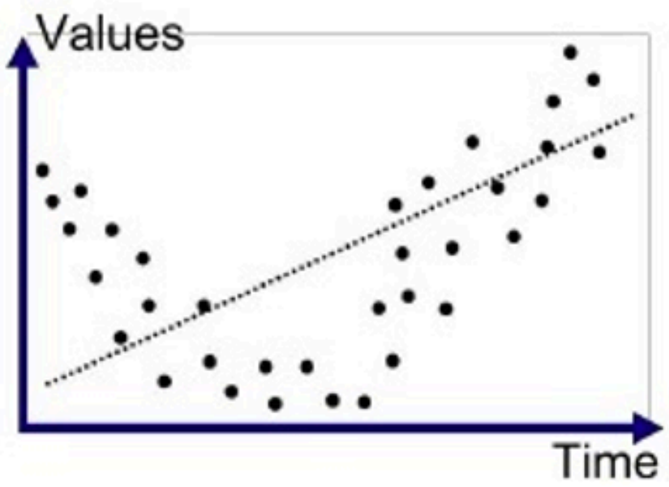
80% of time – Finding and preparing the data

10% of time – Analysis

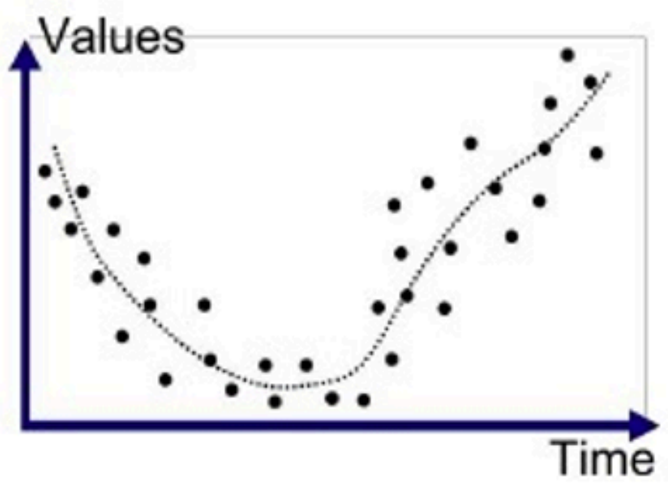
10% of time – Visualizing data

Data Scientist Workflow Steps

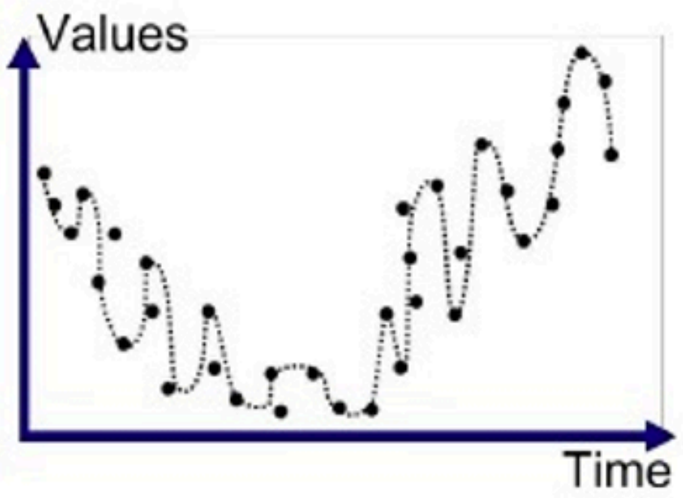
Beware of Overfitting



Underfitted

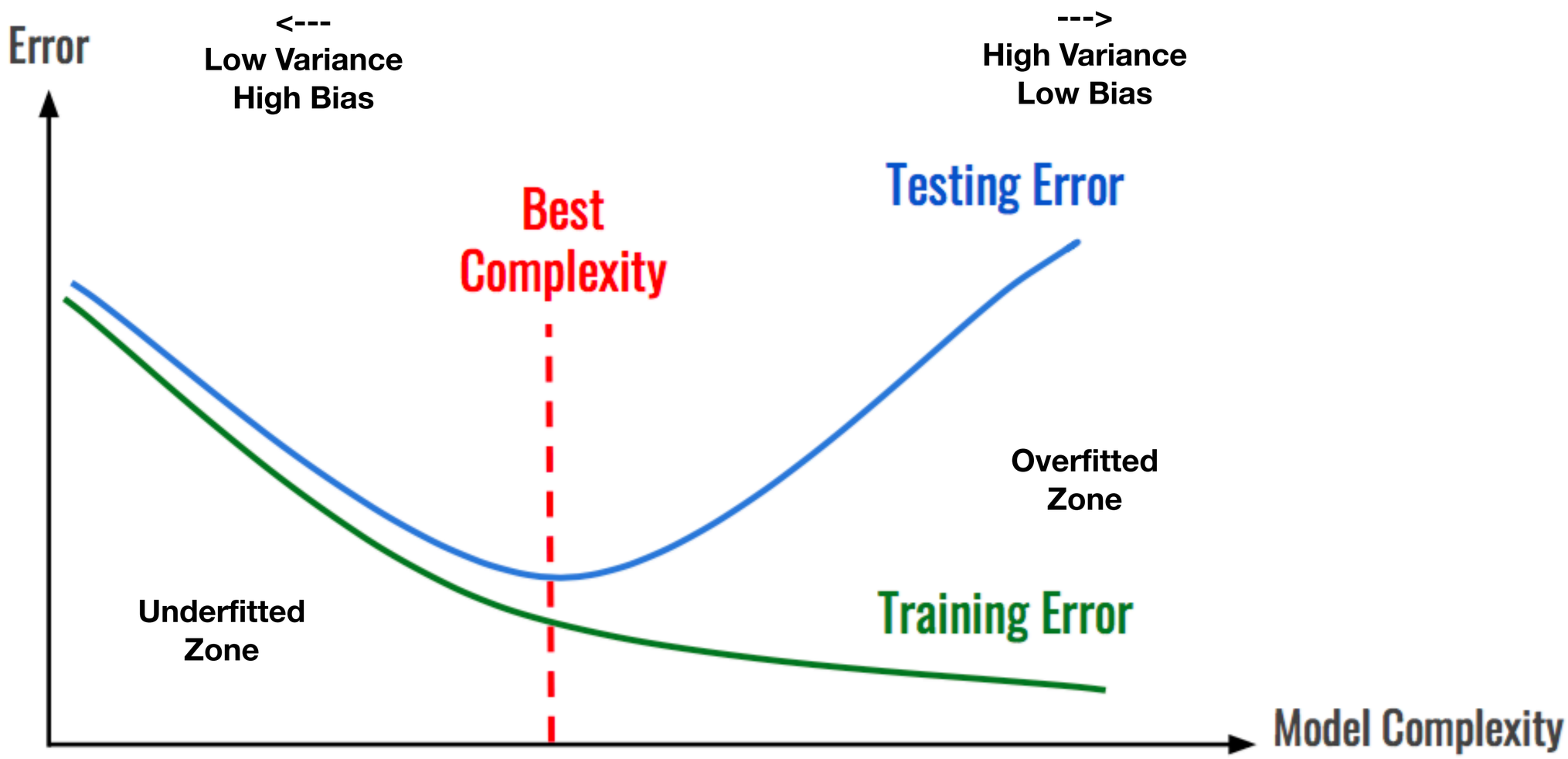


Good Fit/Robust



Overfitted

Bias vs Variance Trade Off



Next Topics in ALT.Labs

Tentative Topics

- Introduction to Unsupervised Learning
- Introduction to Data Science
- Python Programming
- Scratch Programming
- Python for Data Analysis
- IOT Sensors
- Time Series Analysis
- NLP