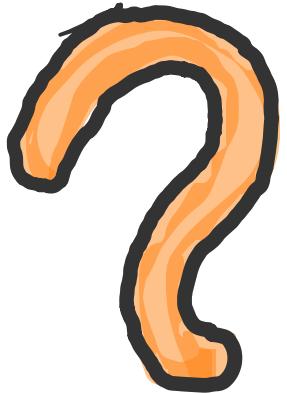


Introduction



The BUZZ words



Data Science

Machine Learning

Artificial Intelligence

Deep Learning

Data Science

- Interdisciplinary field concerning Data
- Almost anything to do with Data:
(e.g.) Data Pipelining, Data Munging, Data Analysis, EDA...
- Includes Soft or Hard science, Small or Big data
- On the Job, Data Scientists do...

What does a Data Scientist do?

Data Scientists are responsible for the collection, cleaning and munging of data to meet the company's purpose. Duties vary according to the industry and may include experimental frameworks for product development and machine learning with the aim to lay a strong data foundation for robust analytics to be performed.

source: indeed.com

Machine Learning

- Subfield of Artificial Intelligence
- About various statistical models and learning algorithms and training from Data
- Supervised/Unsupervised Learning
- Modern version of Statistical Learning
- On the Job, ML engineers do..
Develop and test ML models, Design ML experiments, build ML system....

Artificial Intelligence

- One of the oldest and core subjects in CS
- About Problem-solving with Intelligence
- Theoretical and Practical
- On the Job, AI engineers/experts need skills..

Math and programming skills, Problem-solving skills, Machine learning....

and work on...

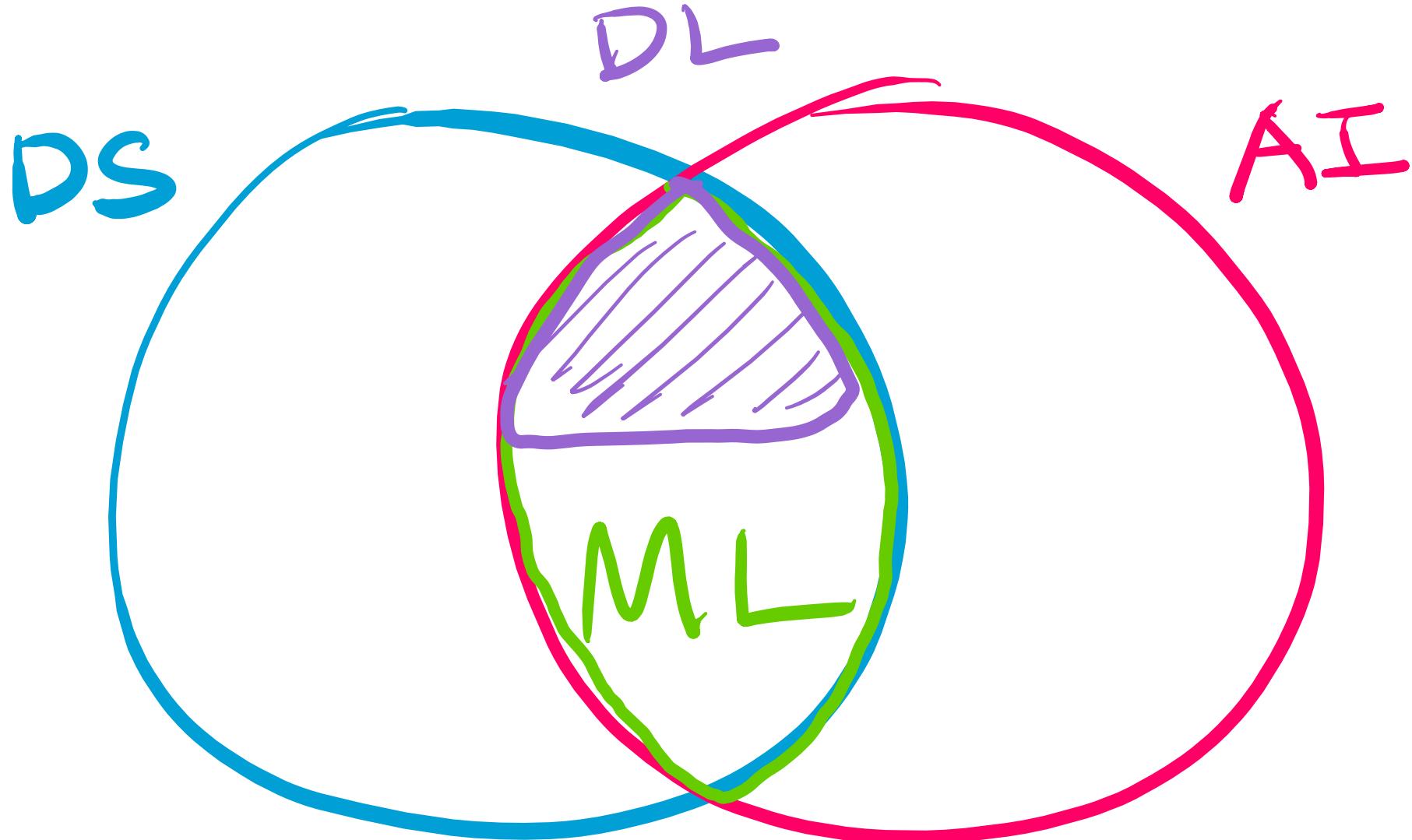
Build and deploy an AI system, build ML models, NLP, Robotics, Computer Vision...

Deep Learning

- About **Neural Network** models, model training from **Data**, and all kinds of training techniques
- Subfield of **Artificial Intelligence**
- Subfield of **Machine Learning**
- On the Job, DL engineers do..

Solving complex technical challenges in various areas of deep learning such as object detection, segmentation, video understanding, sequence prediction, adaptive computing, memory networks, reduced precision training and inference, graph compilers, reinforcement learning, search distributed and federated training, and more

source: NVIDIA



Good Time to Learn Machine Learning!

Machine learning
Field of study

Software engineering
Field of study

+ Add comparison

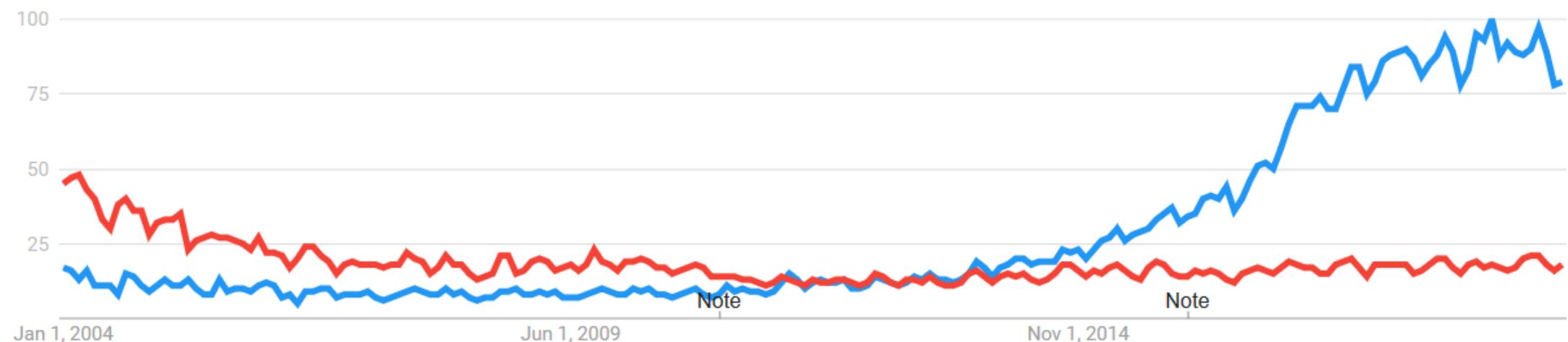
United States ▾

2004 - present ▾

All categories ▾

Web Search ▾

Interest over time ?



Average

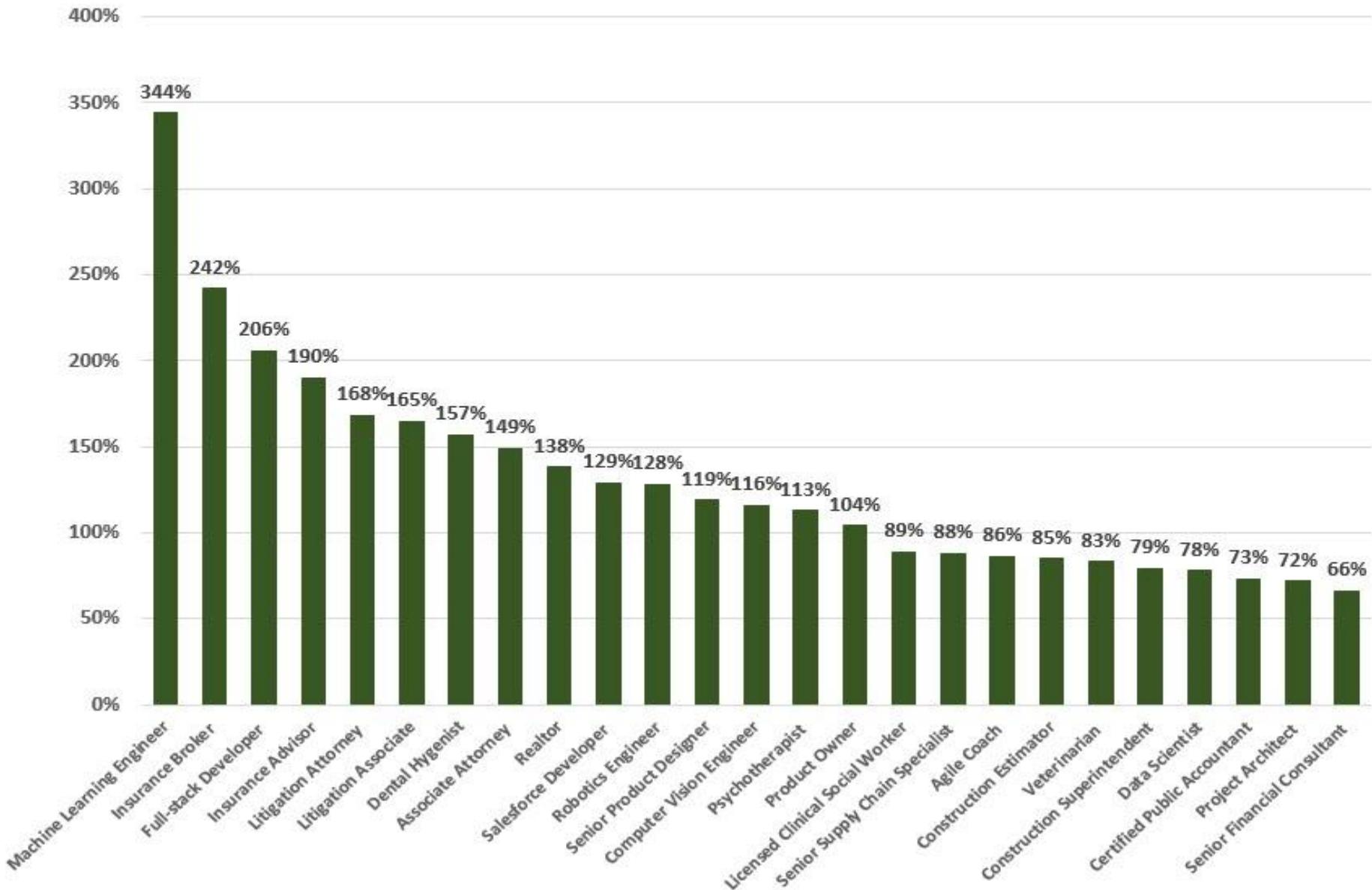
Jan 1, 2004

Jun 1, 2009

Nov 1, 2014

Indeed's Best Jobs In The U.S.
% Growth in # of postings, 2015 - 2018

March 14, 2019



Top 10 jobs involving AI skills

Top jobs seeking machine learning or artificial intelligence skills

Rank	Job title	% of postings containing AI or machine learning	Rank	Job title	% of postings containing AI or machine learning
1.	Machine learning engineer	75.0%	6.	Algorithm developer	46.9%
2.	Deep learning engineer	60.9%	7.	Junior data scientist	45.7%
3.	Senior data scientist	58.1%	8.	Developer consultant	44.5%
4.	Computer vision engineer	55.2%	9.	Director of data science	41.5%
5.	Data scientist	52.1%	10.	Lead data scientist	32.7%

Source: Indeed



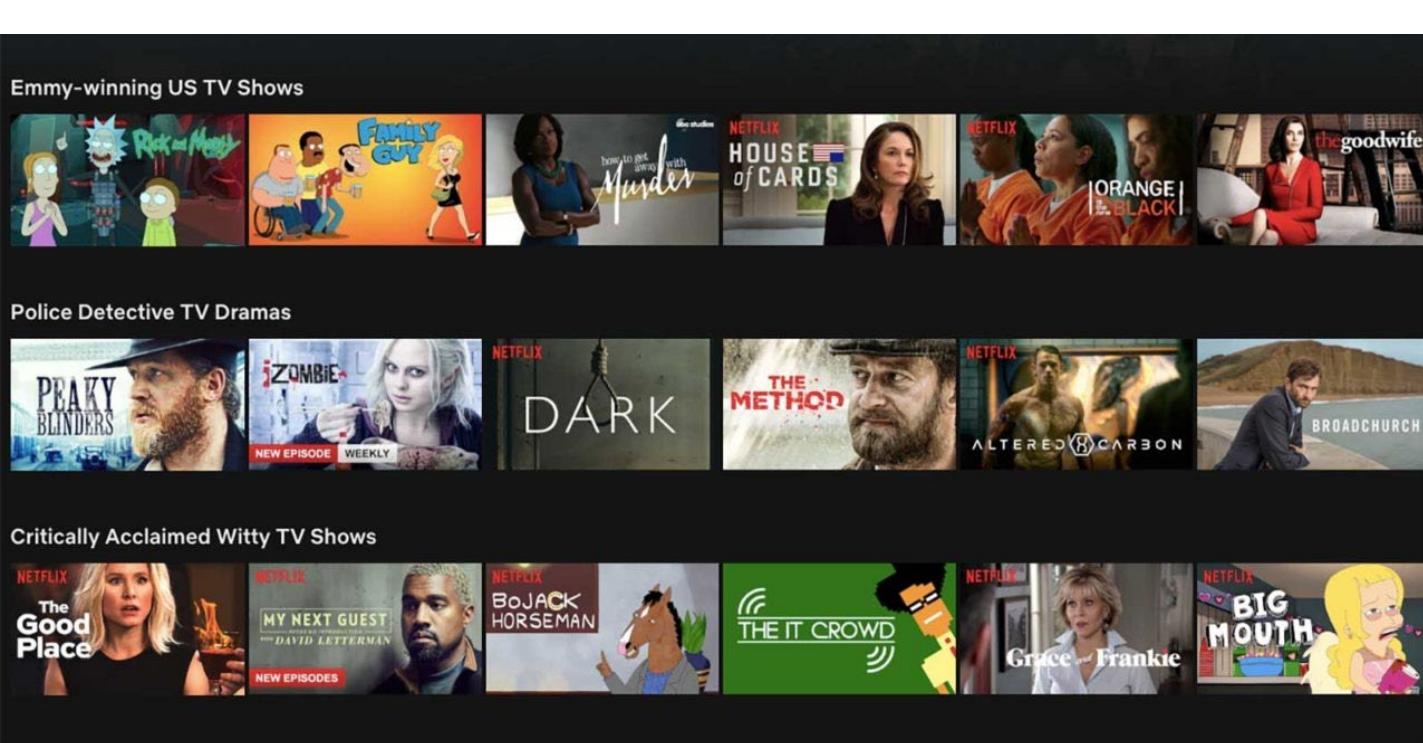
Ok, ML sounds Cool !

What Can I do with ML ?

Machine Learning is Everywhere

Inspired by your browsing history

Product Recommendations



Apple Magic Keyboard
with Numeric Keypad
(Wireless, Rechargeable)
(US English) - Silver
 10,840
\$111.05
✓prime FREE Delivery
In stock soon.



Apple Magic Mouse 2
(Wireless, Rechargeable) -
Space Gray
 25,072
\$89.55
Prime FREE Delivery
In stock soon.



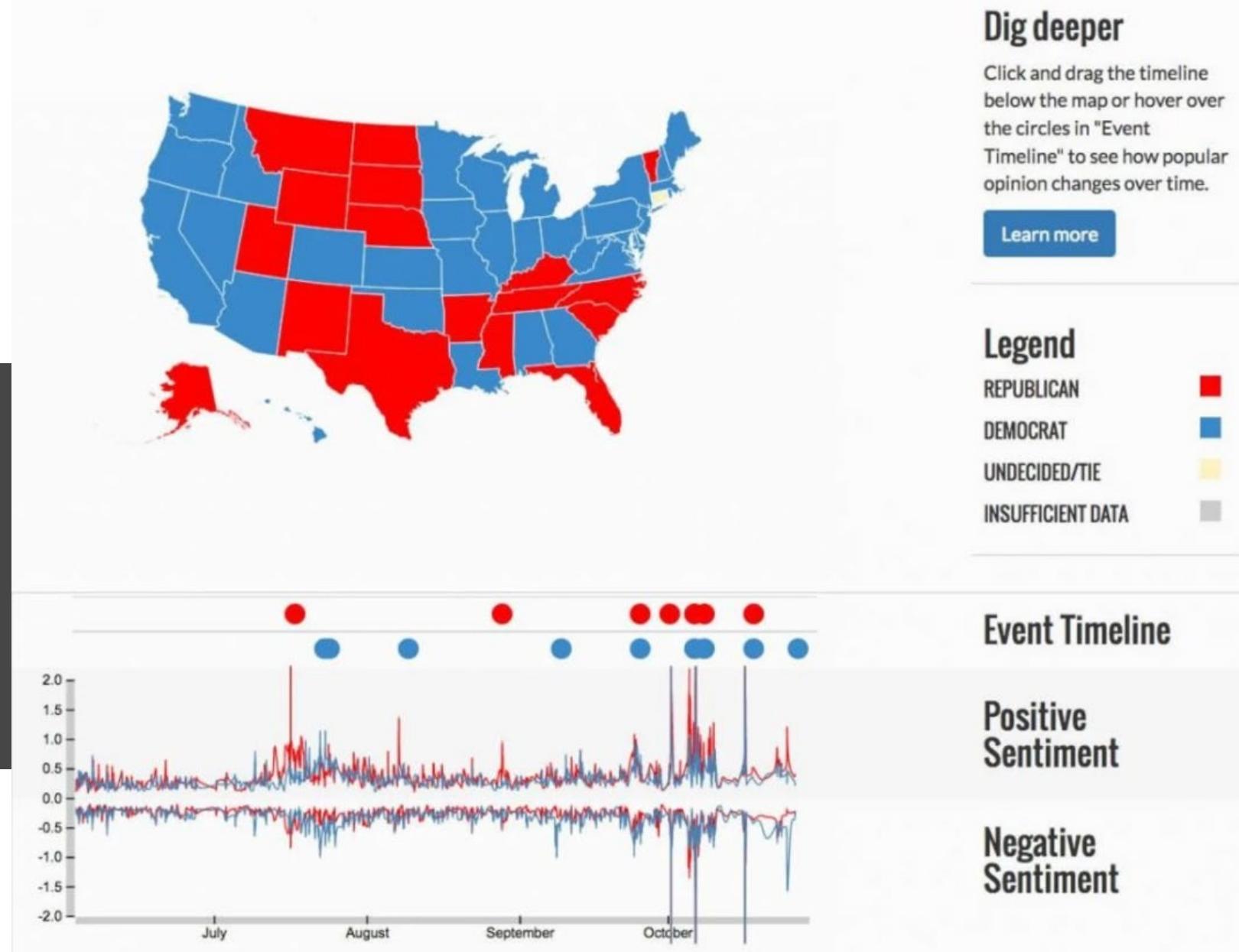
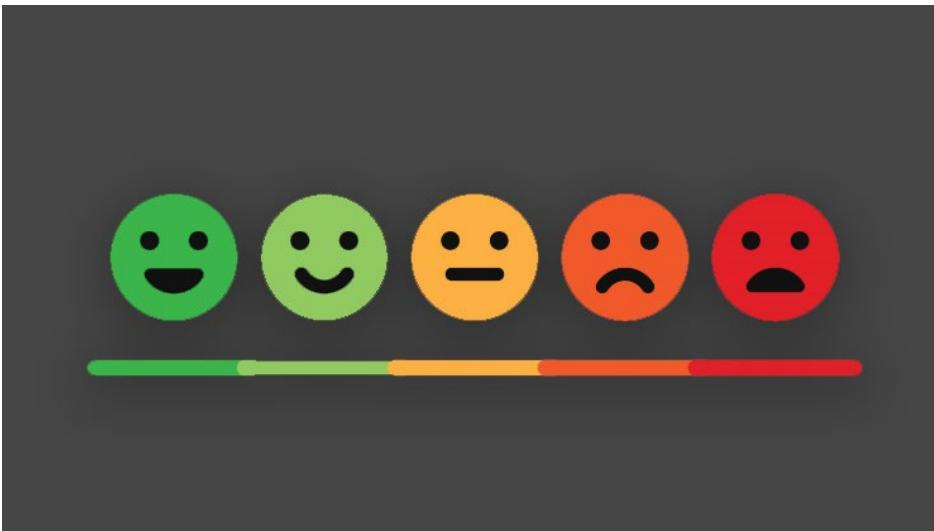
USB C to DisplayPort
Cable for Home Office
(4K@60Hz, 2K@165Hz),
uni Sturdy Aluminum
DisplayPort to USB C...
 3,578
\$15.99
✓prime FREE Delivery
In stock soon.



Logitech C920x HD Pro
Webcam, Full HD
1080p/30fps Video
Calling, Clear Stereo...
 5,621
#1 Best Seller in
Webcams
\$59.99
Prime FREE Delivery
In stock soon.



Sentiment Analysis

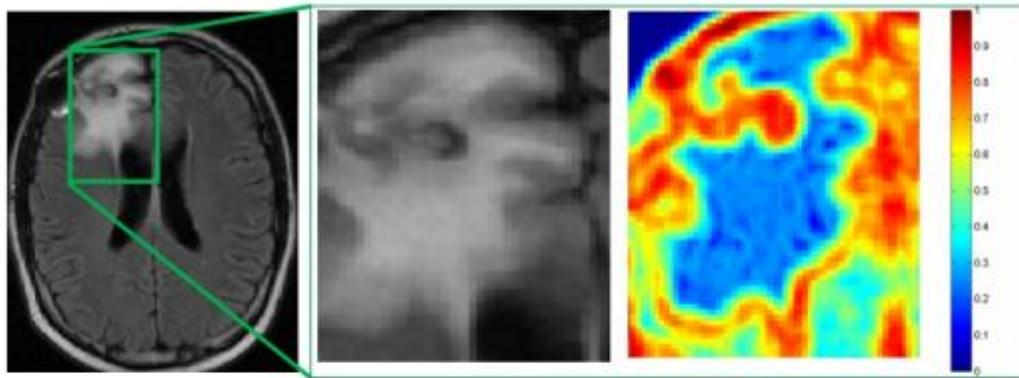


Price Forecasting

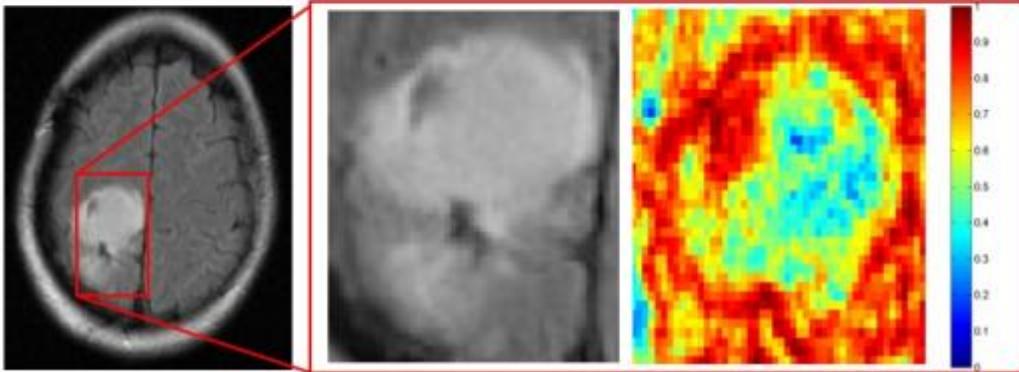


Medical Diagnosis

Radiation necrosis

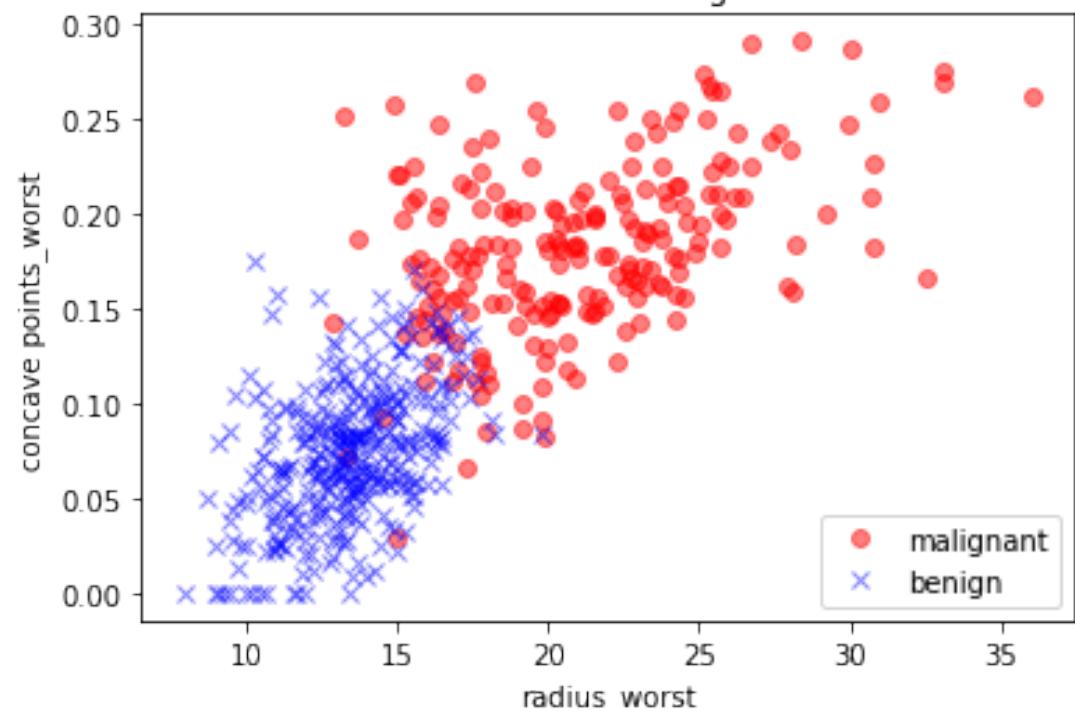


Tumor recurrence

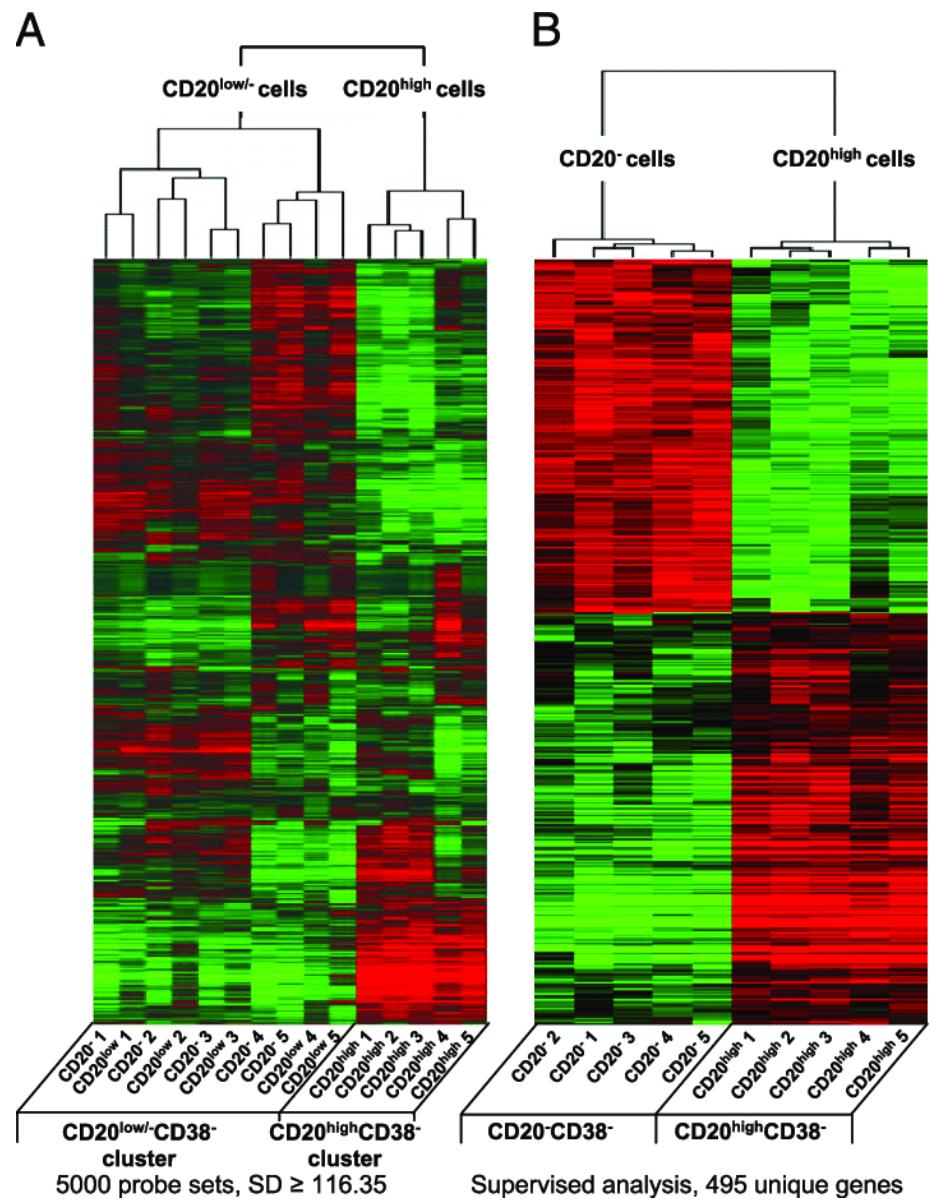
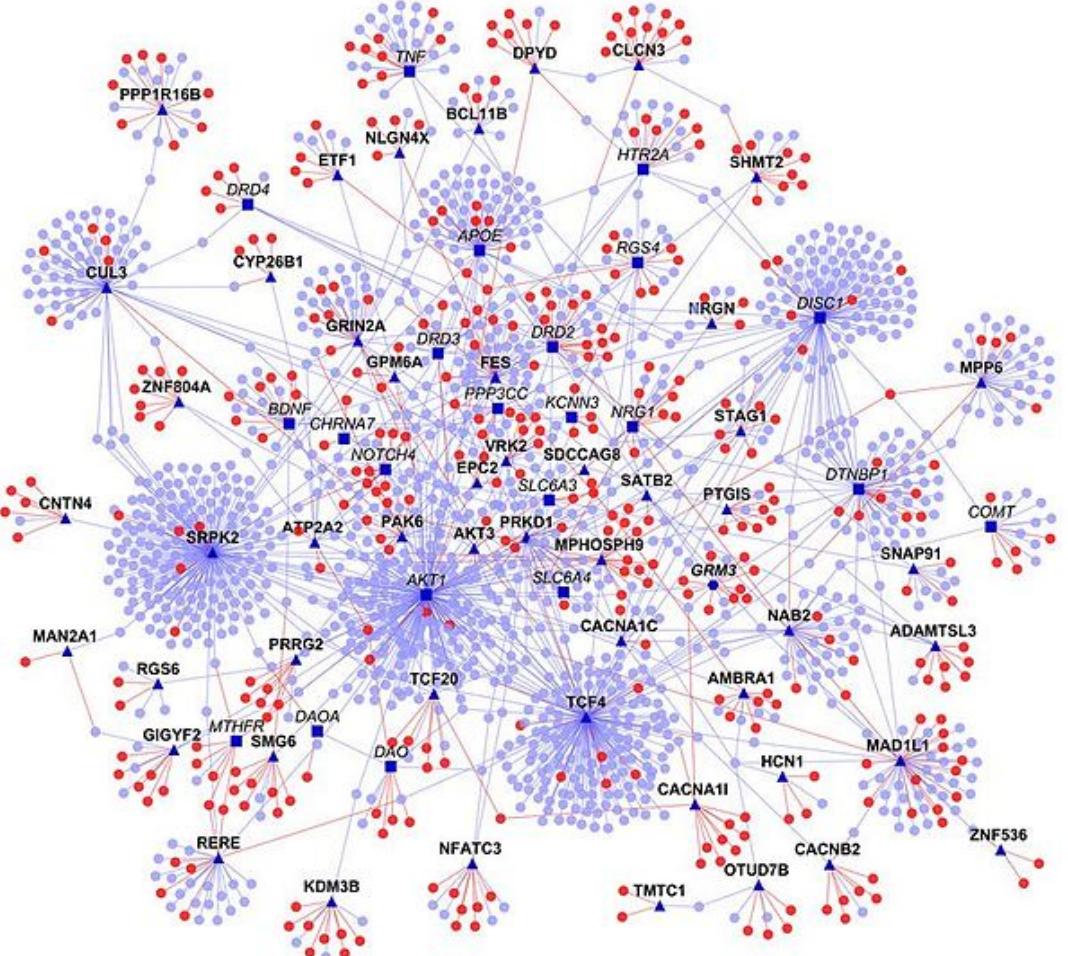


id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean
842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760
842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864
84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990
84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390
84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280

Breast Cancer Diagnosis



Bioscience Research



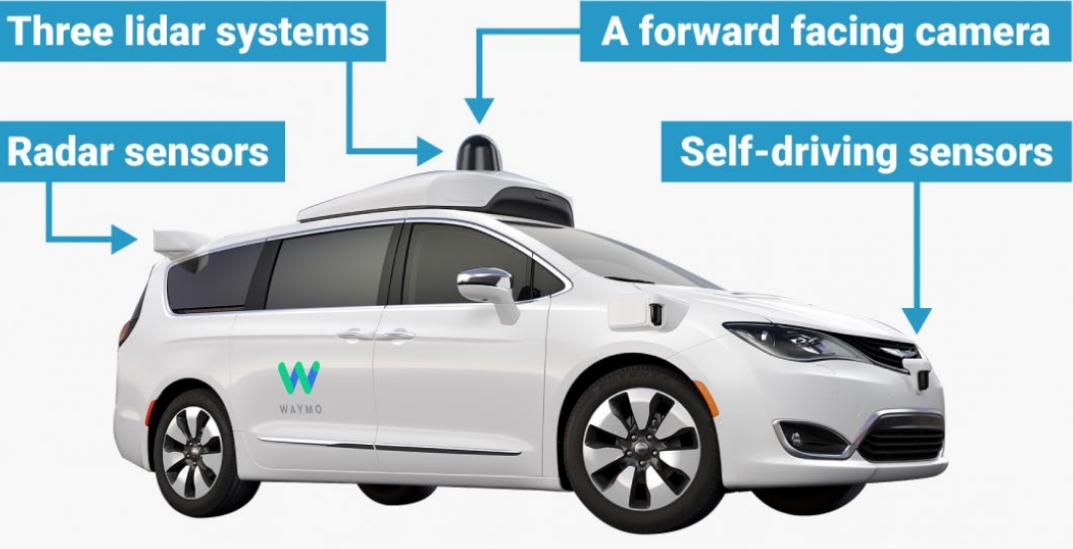
Internet of Things



IoT & Smart City



Self-driving Car



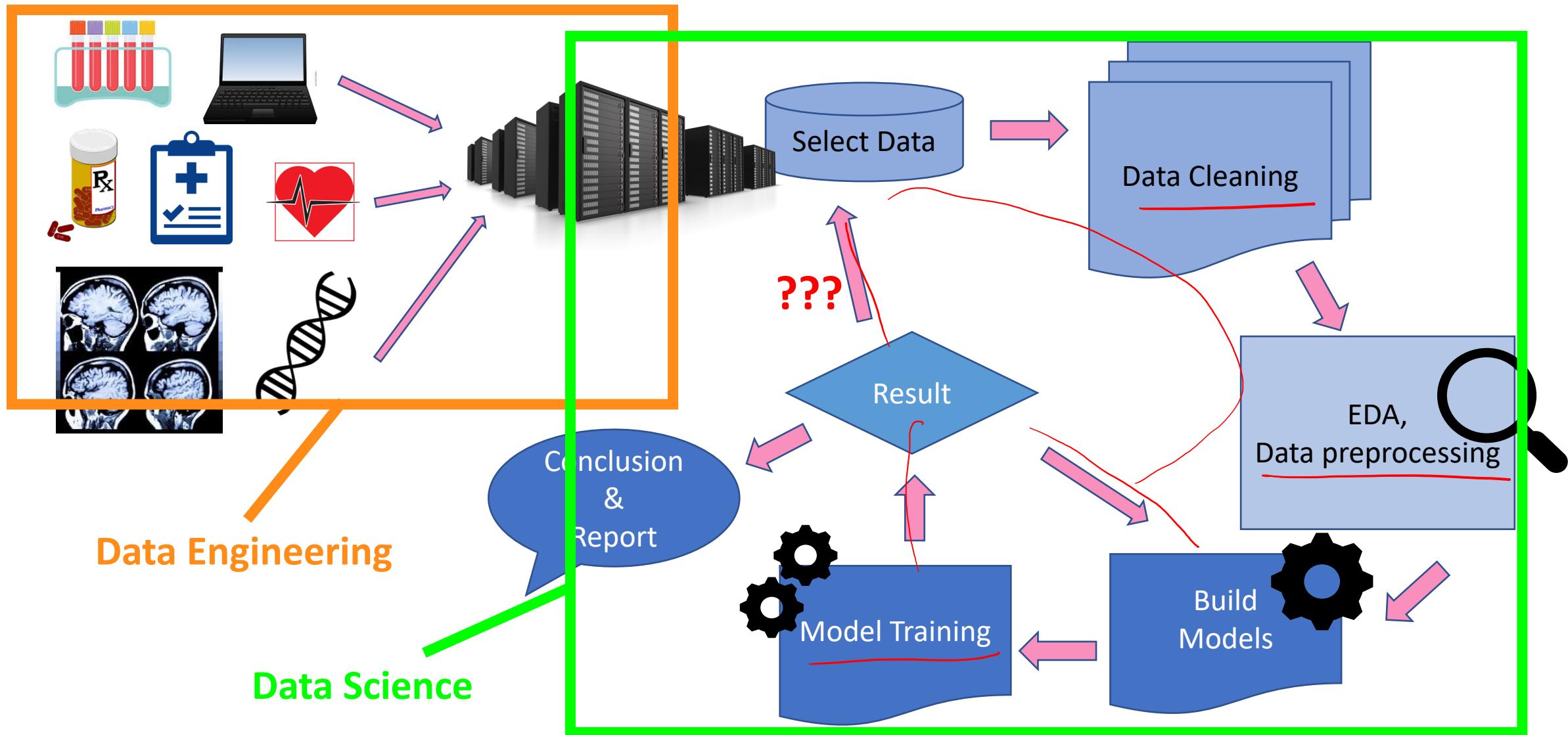
The diagram illustrates a white self-driving car from a three-quarter rear perspective. Several blue callout boxes point to specific sensors: "Three lidar systems" points to the top of the roof-mounted sensor array; "A forward facing camera" points to the front of the vehicle; "Radar sensors" points to the side of the roof; and "Self-driving sensors" points to the front grille area.



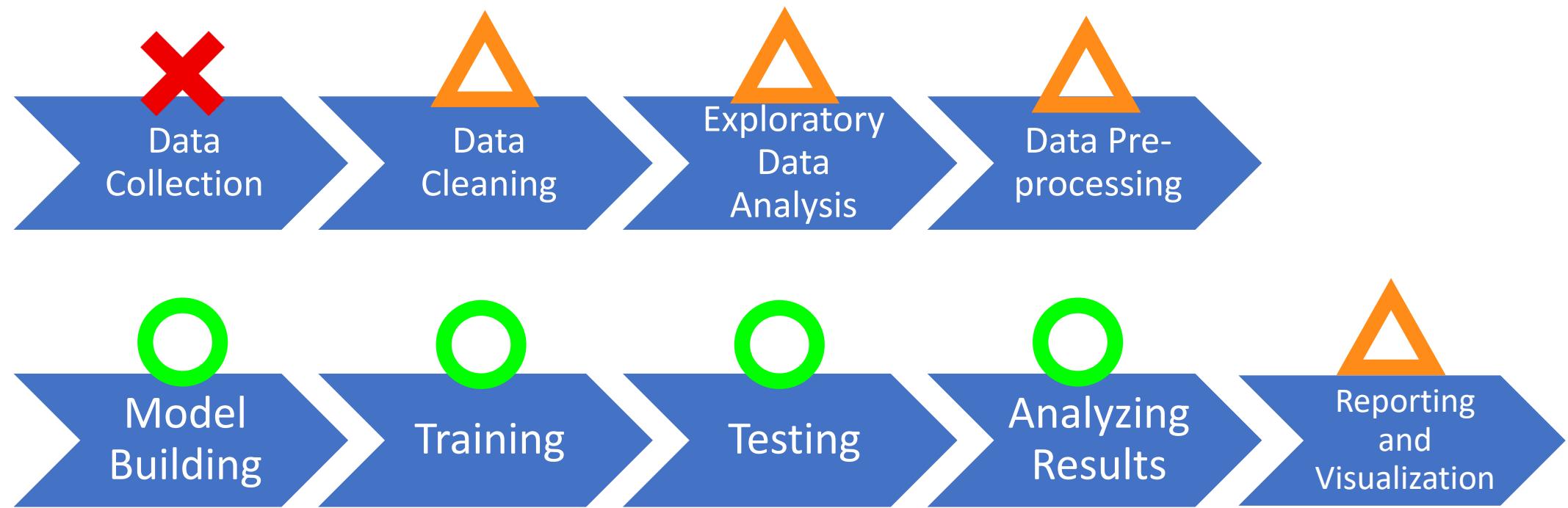
What will we
Learn in this
Course?



Data Science Project Life Cycle



What is this course about?



What is Learning?



Learn to generalize

M M \mathcal{M}

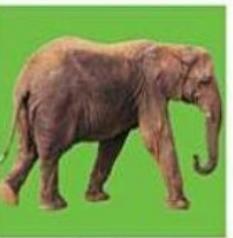
 $\textcolor{red}{M}$ \sim
 \mathcal{M}
Mother

Supervised Learning

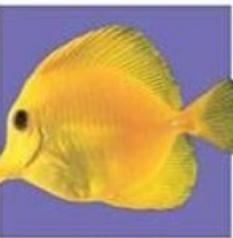
My big animal book



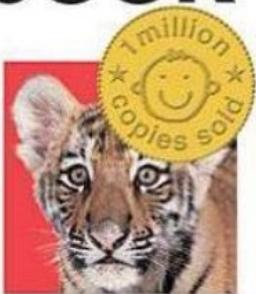
macaw



elephant



fish



tiger cub



zebra



kitten



bear



sheep



duckling



dog



rabbit



cockerel

- Learn the data with labels
- Learn to predict



Unsupervised Learning



- Learn the data without labels
- Learn the underlying features/information

Reinforcement Learning

- Learn how to act from experience
- Experience = Reward/Punishment



Definitions- Data

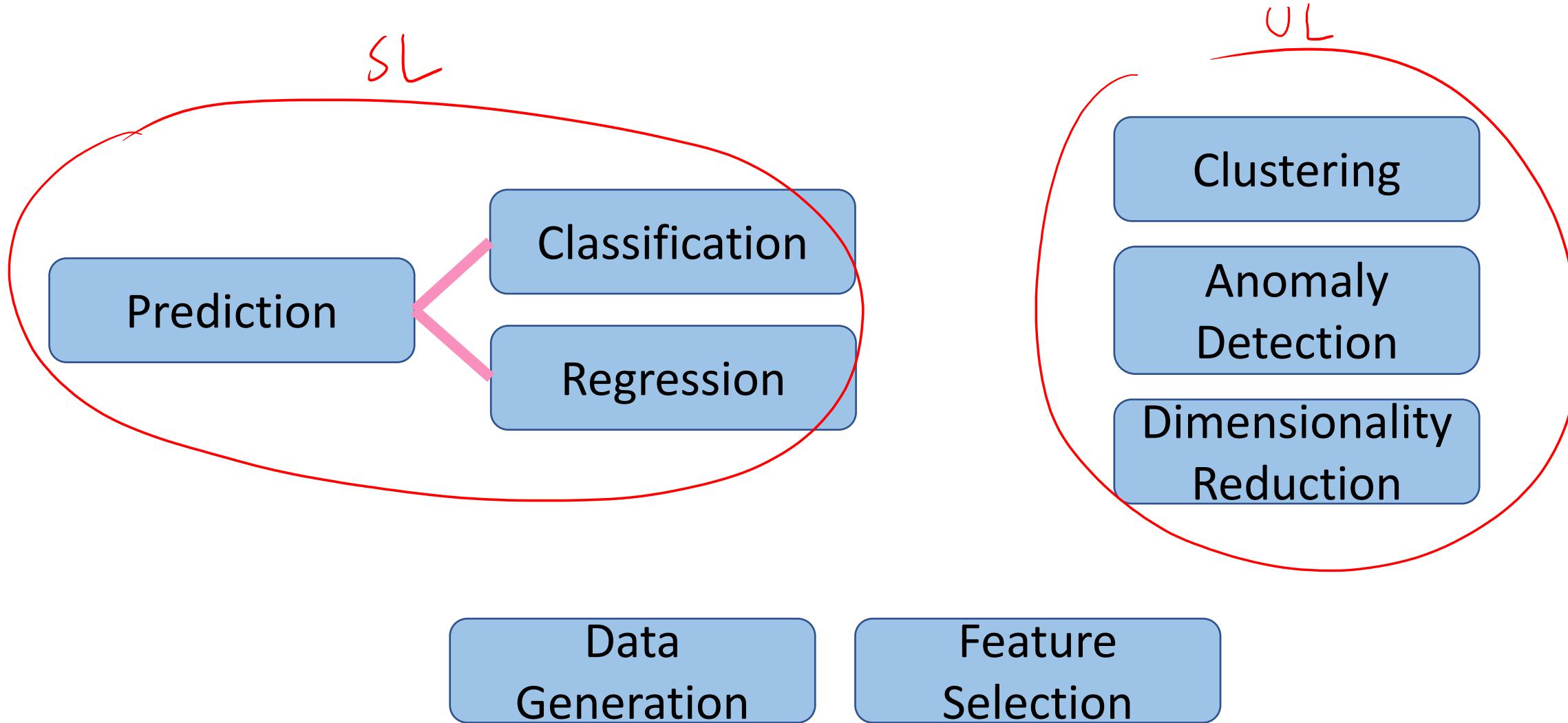
Labels = targets

features: predictors

price	bedrooms	bathrooms	sqft_living	sqft_lot	floors
221900	3	1.00	1180	5650	1.0
538000	3	2.25	2570	7242	2.0
180000	2	1.00	770	10000	1.0
604000	4	3.00	1960	5000	1.0
510000	3	2.00	1680	8080	1.0

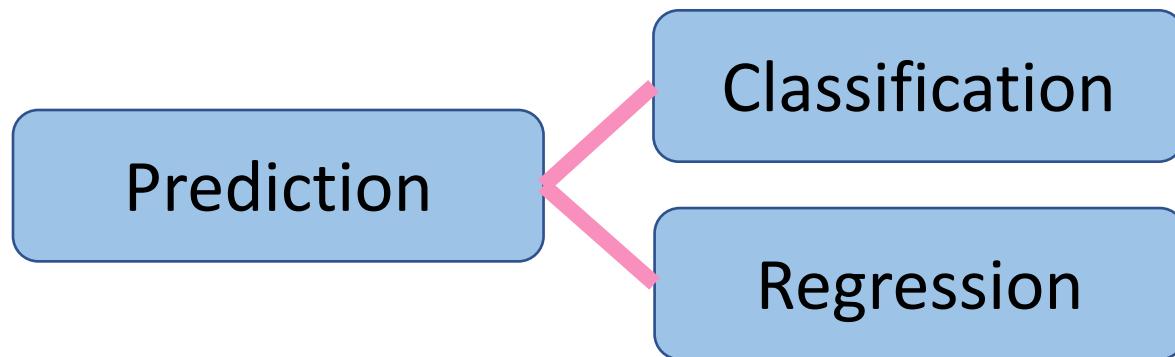
Observations
samples

Definitions- Machine Learning Tasks

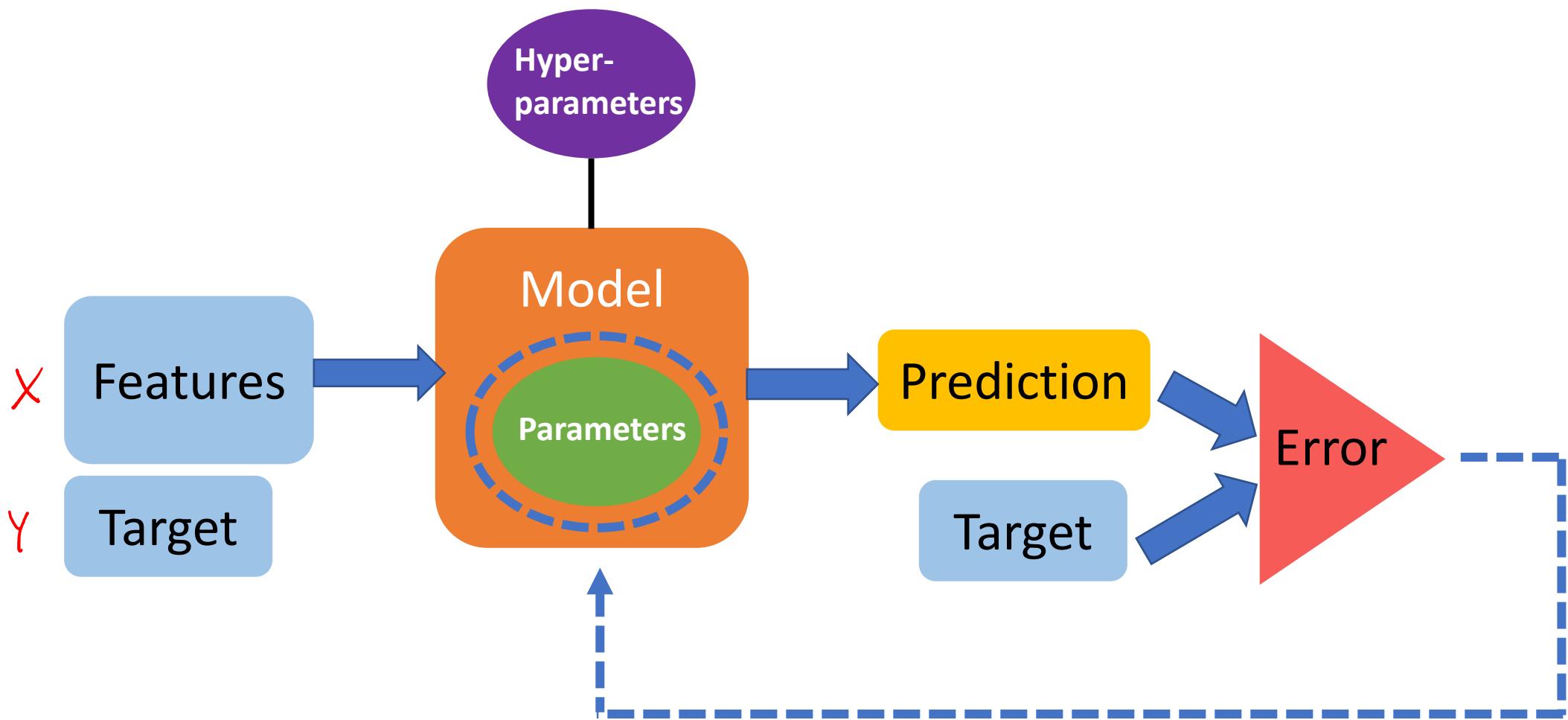


Definitions- Supervised Machine Learning Tasks

y 0/1 1, 2, 3, A, B, C
binary multi-class
0.1, 0.999, 3.4



Supervised Learning



Machine Learning Models

Supervised

Parameteric

- Linear Regression (R)
- Logistic Regression (C)
- Neural Networks (R, C)
- Naïve Bayes Classifier (C)

Non-parameteric

- k-Nearest Neighbors (R, C)
- Decision Trees (R,C)
- Random Forest (R, C)
- Gradient Boosting (R, C)
- Support Vector Machine (R,C)