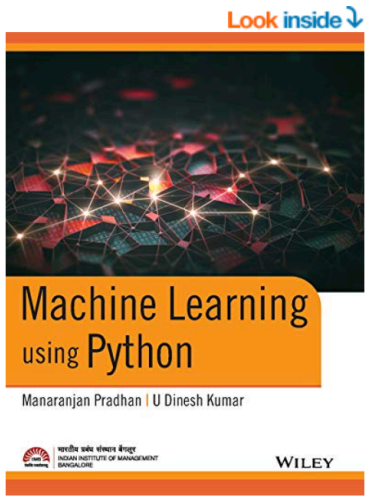
ML Systems Design

Manaranjan Pradhan

About Me

- Consulting and teaching on Big Data, Machine Learning, Deep Learning and MLOps
- Adjunct Faculty for IIM Bangalore, ISB Hyderabad
- https://www.linkedin.com/in/manaranjanpradhan

Manaranjan Pradhan (Manu)



Machine Learning using Python [Print Replica] Kindle Edition

by Manaranjan Pradhan (Author), U Dinesh Kumar (Author) | Format: Kindle Edition

363 ratings

See all formats and editions

Kindle Edition	Paperback
₹386.40	₹494.00 prime
Read with Our Free App	13 New from ₹483.00

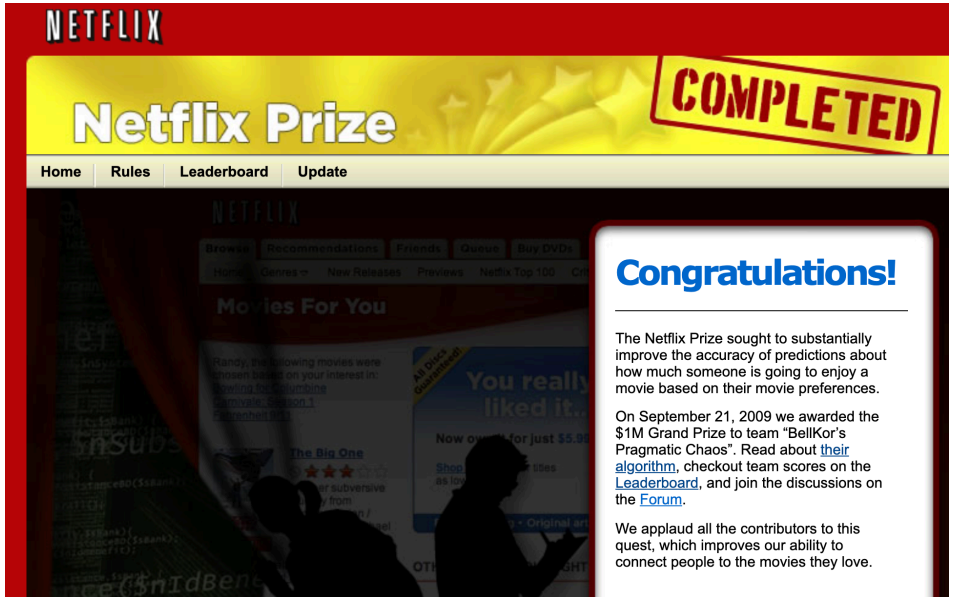
This book is written to provide a strong foundation in Machine Learning using Python libraries by providing real-life case studies and examples. It covers topics such as Foundations of Machine Learning, Introduction to Python, Descriptive Analytics and Predictive Analytics. Advanced Machine Learning concepts such as decision tree learning, random forest, boosting, recommender systems, and text analytics are covered. The book takes a balanced approach between theoretical understanding and

‹ Read more

https://www.amazon.in/Machine-Learning-Python-Manaranjan-Pradhan-ebook/dp/B07RLQPNRX/

Why ML Systems Design?

Netflix 1 Million Dollar Challenge!



Leaderboard

Showing Test Score. Click here to show quiz score

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace_	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11

https://www.netflixprize.com/

Netflix 1M USD Prize

CASEY JOHNSTON, ARS TECHNICA BUSINESS 04.16.2012 08:20 AM

Netflix Never Used Its $1 Million Algorithm Due To Engineering Costs

Netflix awarded a $1 million prize to a developer team in 2009 for an algorithm that increased the accuracy of the company's recommendation engine by 10 percent. But it doesn't use the million-dollar code, and has no plans to implement it in the future, Netflix announced on its blog Friday. The post goes on to explain why: [...]

- https://www.wired.com/2012/04/netflix-prize-costs/
- https://netflixtechblog.com/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429

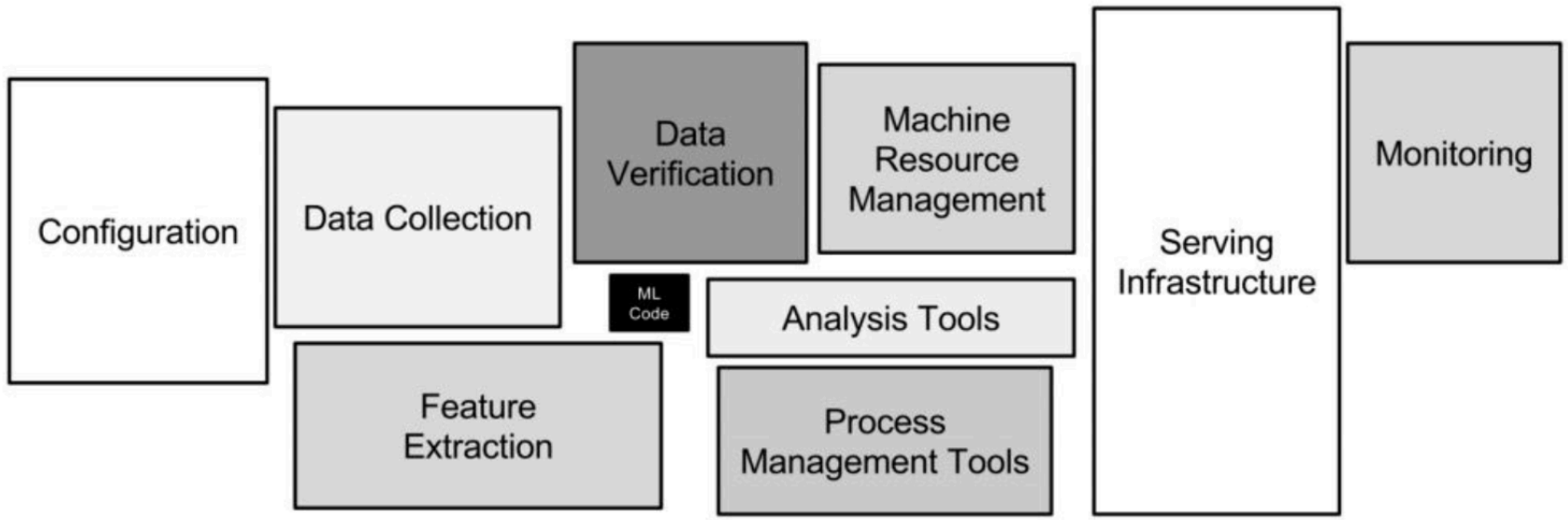
Netflix Blog

A year into the competition, the Korbell team won the first Progress Prize with an 8.43% improvement. They reported more than 2000 hours of work in order to come up with the final combination of 107 algorithms that gave them this prize. And, they gave us the source code. We looked at the two underlying algorithms with the best performance in the ensemble: *Matrix Factorization* (which the community generally called SVD, *Singular Value Decomposition*) and *Restricted Boltzmann Machines* (RBM). SVD by itself provided a 0.8914 RMSE, while RBM alone provided a competitive but slightly worse 0.8990 RMSE. A linear blend of these two reduced the error to 0.88. To put these algorithms to use, we had to

with the final Grand Prize ensemble that won the $1M two years later. This is a truly impressive compilation and culmination of years of work, blending hundreds of predictive models to finally cross the finish line. We evaluated some of the new methods offline but the additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment. Also, our focus on improving Netflix personalization

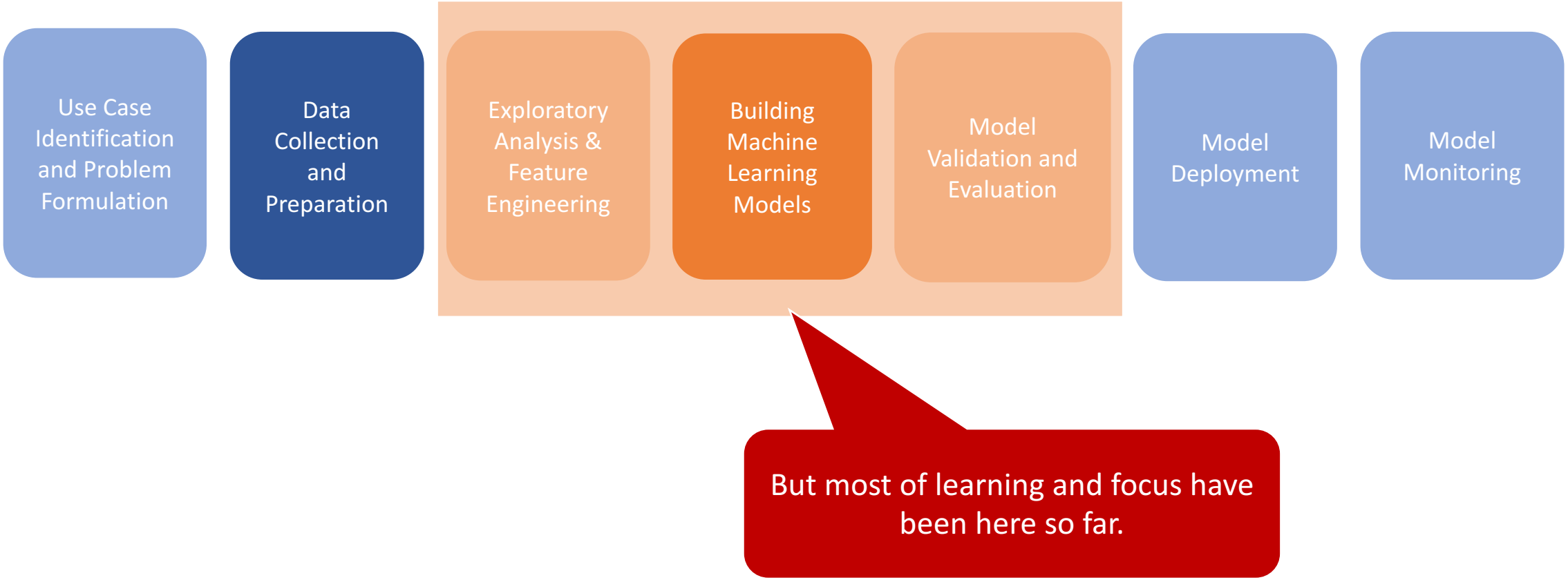
Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.

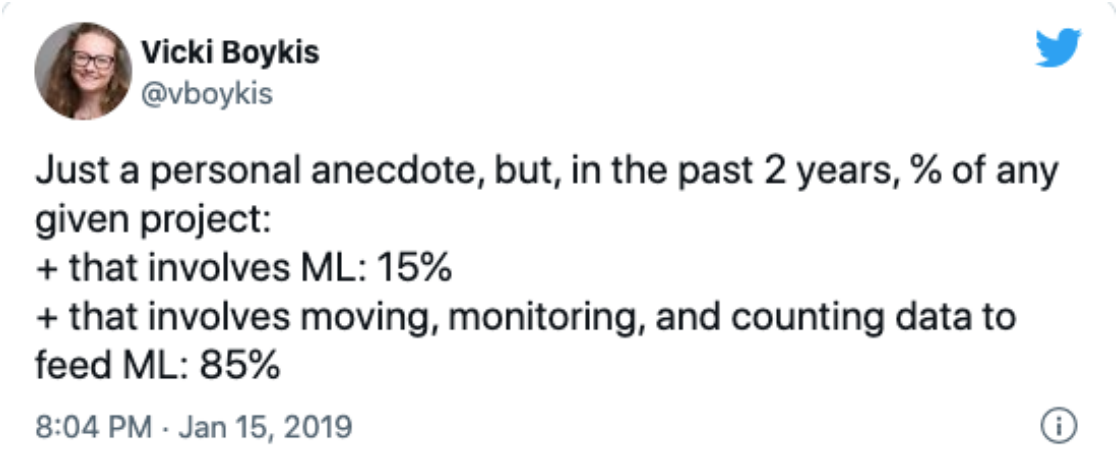


https://papers.nips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf

ML Lifecycle

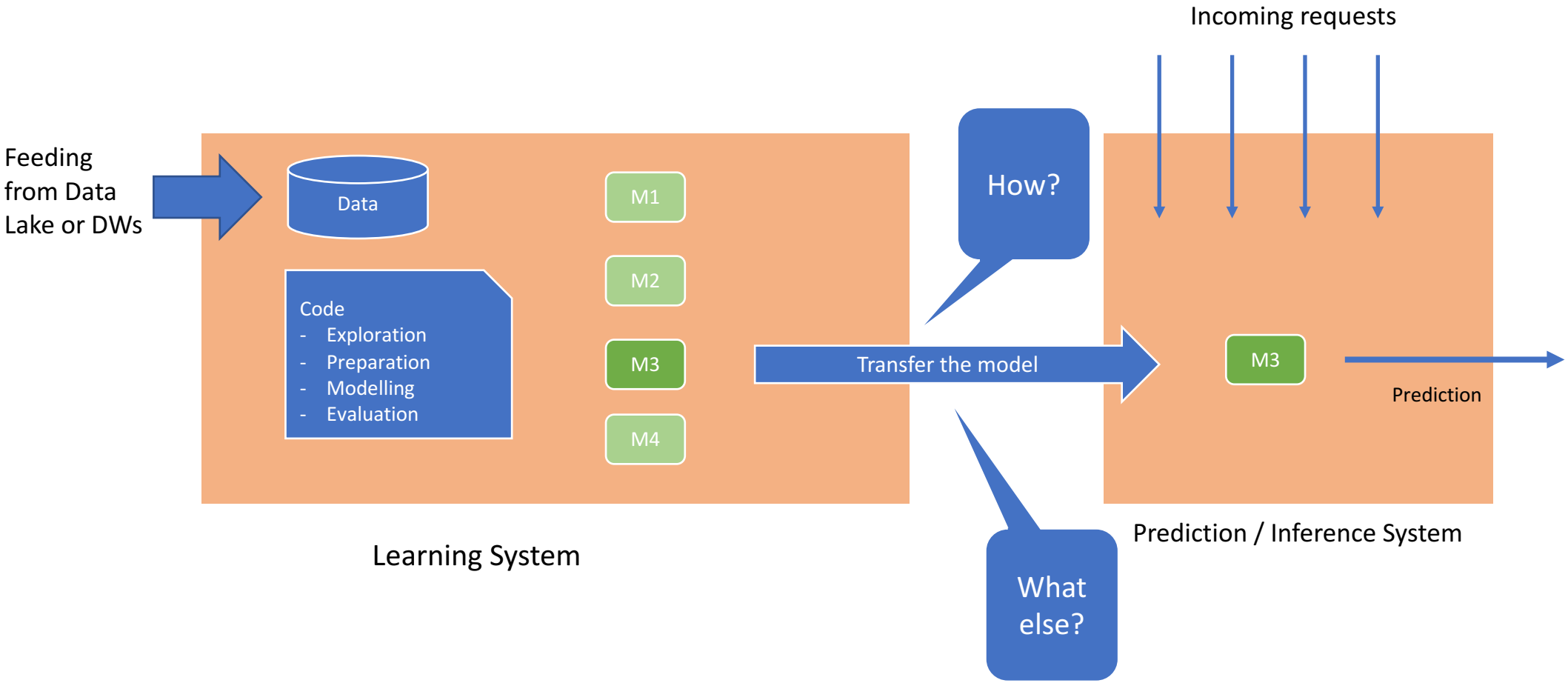


What Practitioners Say?



What exactly is a ML System?

When you say ML systems

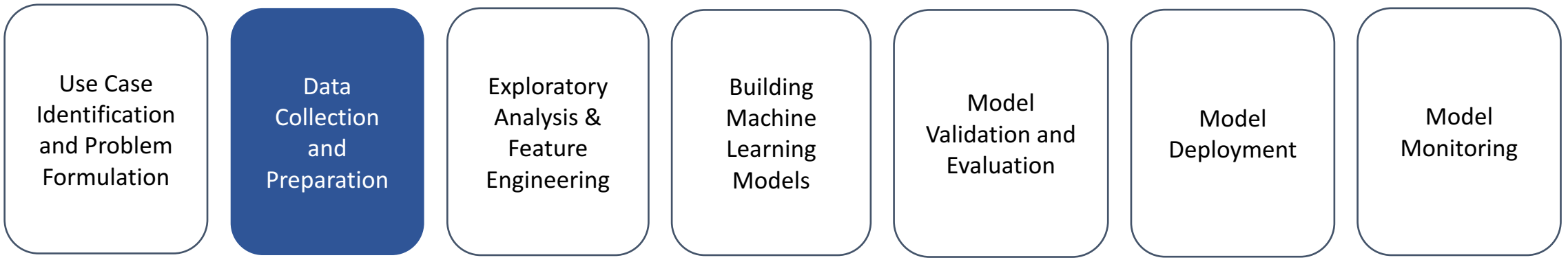


Different Types of ML systems

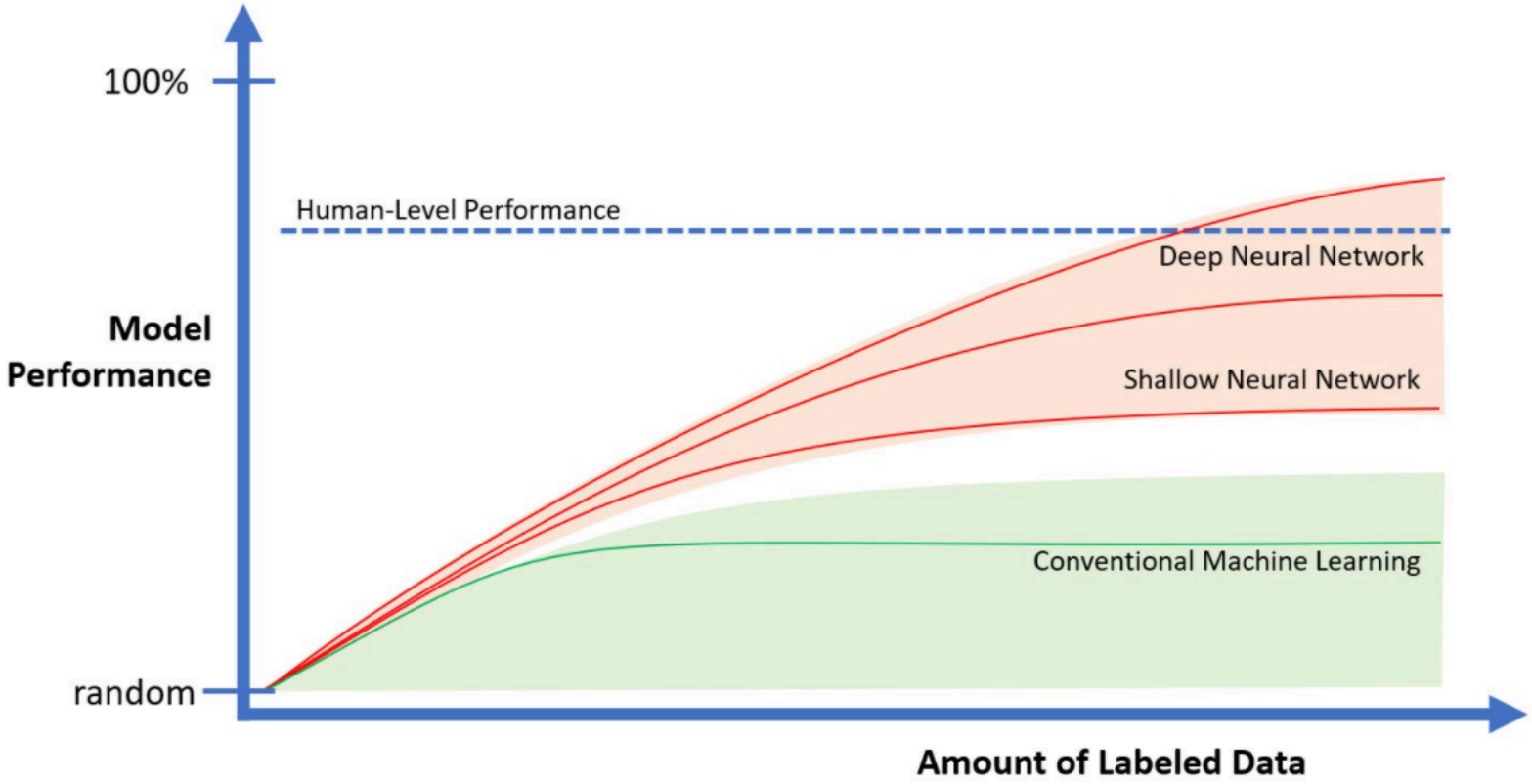
	Batch prediction	Online prediction
Frequency	Periodical (e.g. every 4 hours)	As soon as requests come
Useful for	Processing accumulated data when you don't need immediate results (e.g. recommendation systems)	Need prediction result immediately
Optimized	High throughput	Low latency
Examples	• TripAdvisor hotel ranking • Netflix recommendations • Customer Churn Analysis	• Google Assistant speech recognition • Fraud Detection

https://stanford-cs329s.github.io/

ML Lifecycle



Need for more Labelled Data



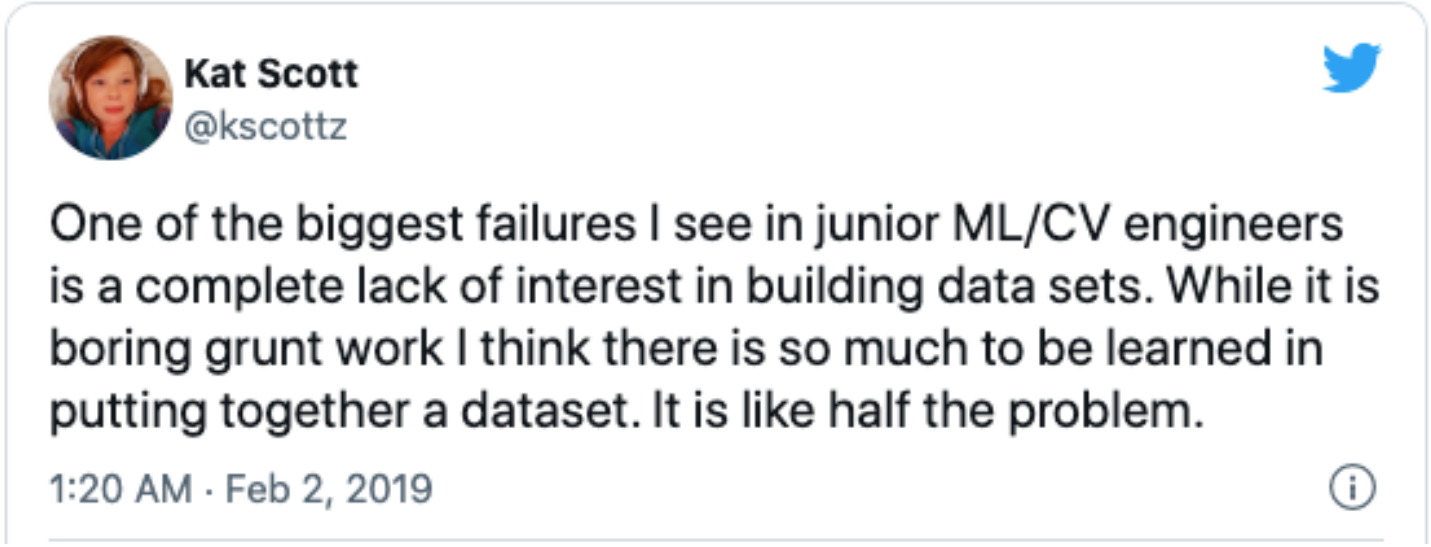
- Labelled data may not available or scarce.
- It is time consuming and expensive to create labelled data.

- Weak Supervision
- Data Augmentation

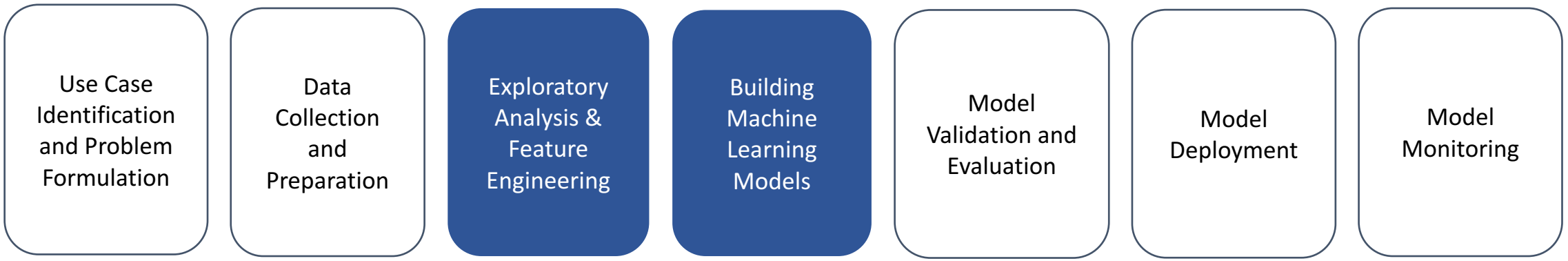
https://towardsdatascience.com/how-managers-should-prepare-for-deep-learning-new-paradigms-28de63054ea6

How to create so much labelled data?

What Practitioners Say?

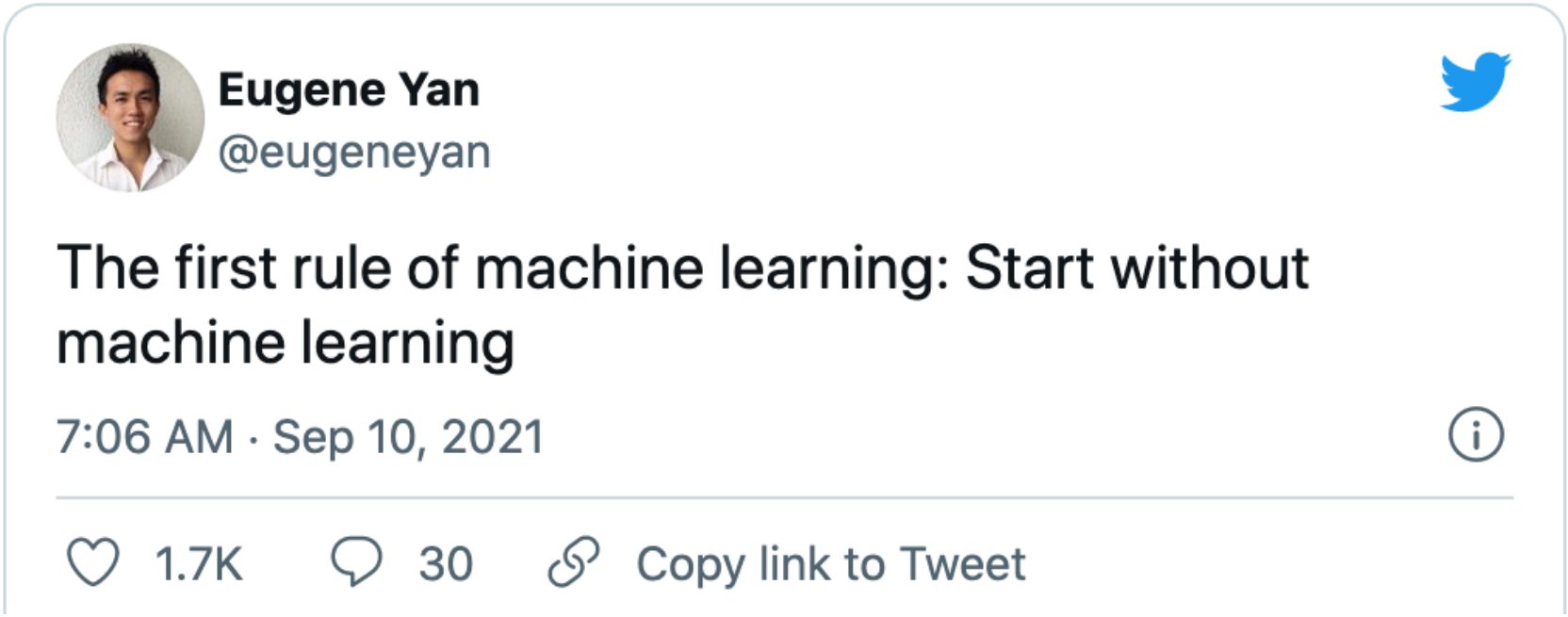


ML Lifecycle



Do we always need to build a ML Model?

Model Baselining



https://eugeneyan.com/writing/first-rule-of-ml/

Google's Rule for Machine Learning

Rule #1: Don't be afraid to launch a product without machine learning.

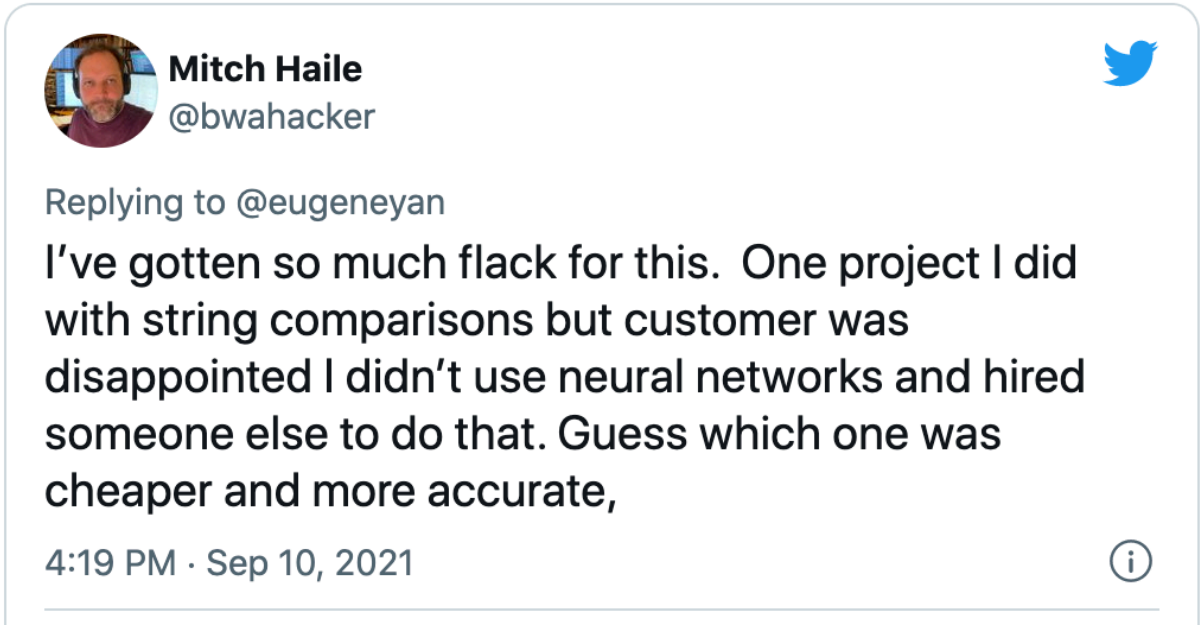
Machine learning is cool, but it requires data. Theoretically, you can take data from a different problem and then tweak the model for a new product, but this will likely underperform basic heuristics. If you think that machine learning will give you a 100% boost, then a heuristic will get you 50% of the way there.

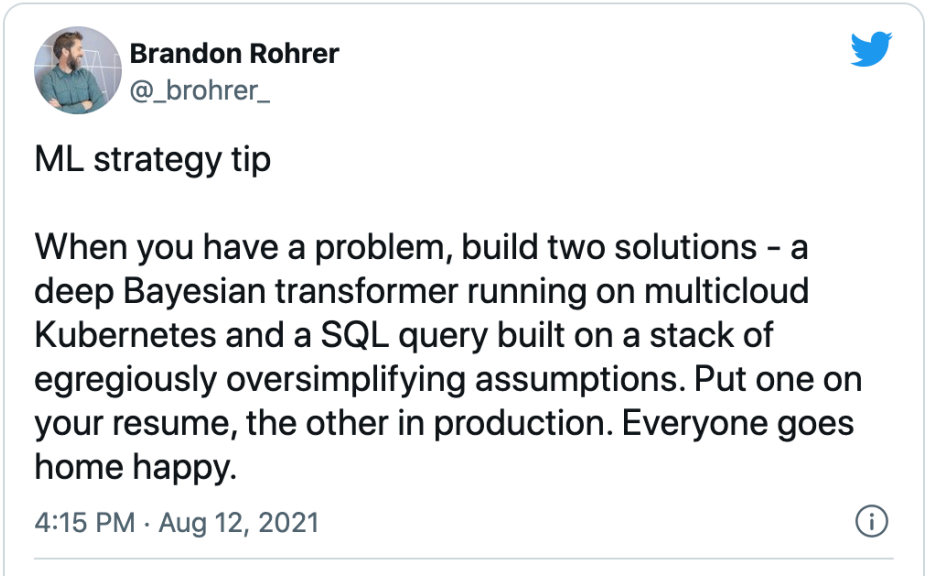
https://developers.google.com/machine-learning/guides/rules-of-ml

Baselining

- Create a system with if/else rules from heuristics
- Build a simple ML (linear regression) first
- Build a system with regex (hand crafted regular expressions) for classifying text data
- Benefits of creating a heuristics system

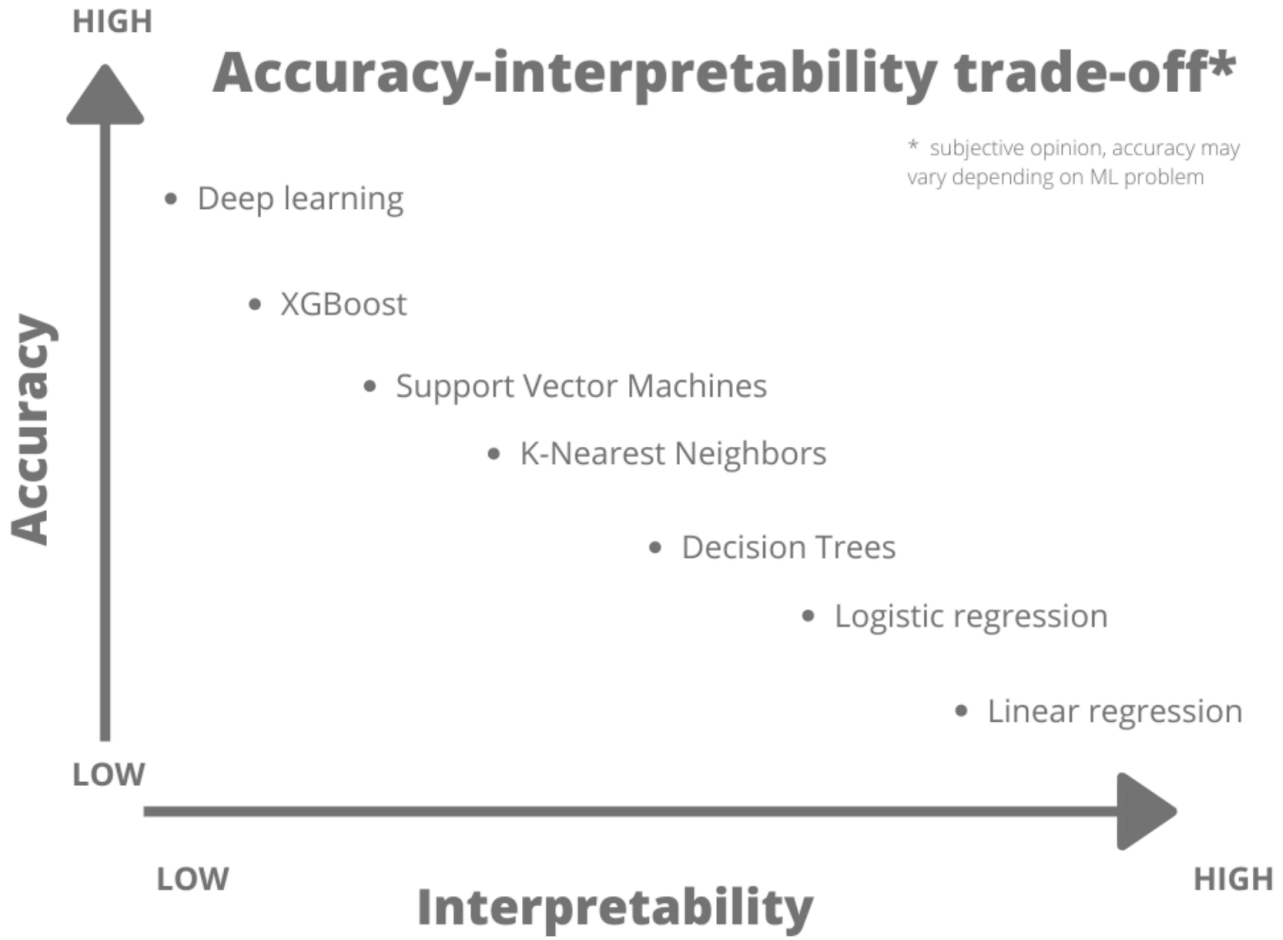
When you are forced to build ML systems?





Which model need to be built?

Accuracy – Interpretability Tradeoff



Accuracy Vs. Explainability

Amazon scraps secret AI recruiting tool that showed bias against women

https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

FORTUNE

Amazon Reportedly Killed an AI Recruitment System Because It Couldn't Stop the Tool from Discriminating Against Women

BY DAVID MEYER
October 10, 2018 3:30 PM GMT+5:30

https://fortune.com/2018/10/10/amazon-ai-recruitment-bias-women-sexist/

Amazon scrapped 'sexist AI' tool

https://www.bbc.com/news/technology-45809919

- White box Vs. Black box models
- Explainable AI (XAI)
- Bias and Fairness

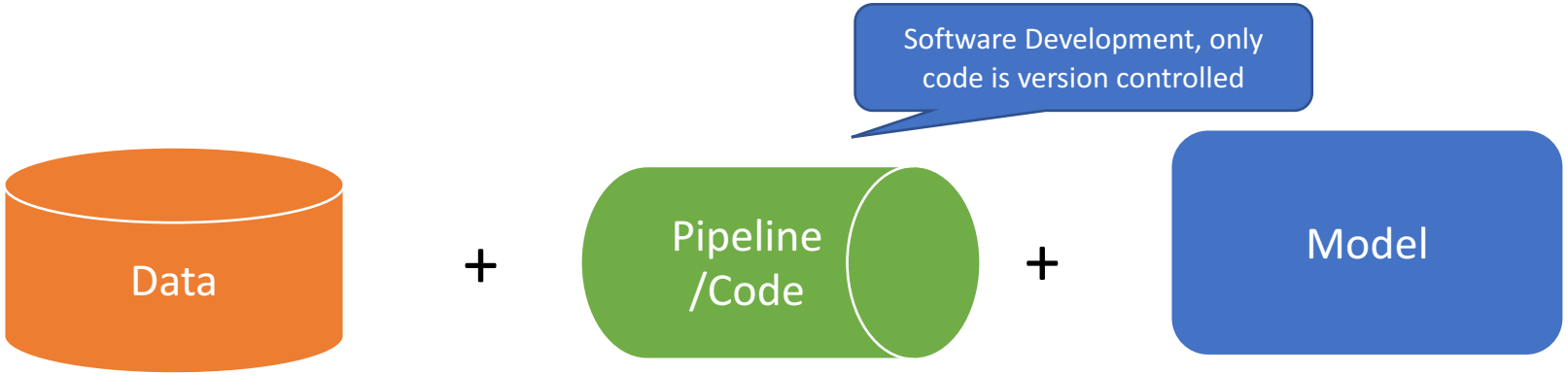
Model Development is messy!

- Run number of experiments to refine your model
- Experiments may involve
  - Different Transformations
  - Different Models
  - Different Hyperparameters
- Easy to lose track of code, hyperparameters, and artifacts
- Fail to reproduce experiments (reproducibility)

Experiment Tracking

Name (84 visualized)	acc	human_iou	iou	State	Notes	User	Tags	Created	Runtime	Sweep	batch_size.(	batch_size.1	batch_size.1	batch_size	bn_weight_	bn_weight_	encoder	epochs	framework	img_size
good-cosmos-425	0.4031	8.015e-8	0.3154	failed	Add notes	stacey		1y ago	43s	-	-	-	-	8	-	true	resnet34	10	fast.ai	[360,640]
logical-energy-420	0.626	9.030e-8	0.4297	finished	Add notes	stacey	test_or	1y ago	2m 14s	-	-	-	-	8	-	true	resnet34	10	fast.ai	[360,640]
laced-dust-419	0.5968	8.677e-8	0.4701	finished	Add notes	stacey	test_or	1y ago	2m 4s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
whole-music-418	0.6139	7.775e-8	0.4728	finished	Add notes	stacey	test_or	1y ago	1m 40s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
grateful-glitter-417	0.2367	9.091e-8	0.1209	failed	Add notes	stacey	test_or	1y ago	21s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
clear-night-415	0.5403	8.878e-8	0.3764	killed	Add notes	stacey	test_or	1y ago	1m 17s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
glorious-night-414	0.7627	7.359e-8	0.5818	killed	Add notes	stacey	test_or	1y ago	1m 33s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
smart-sponge-413	0.6517	7.501e-8	0.4796	finished	Add notes	stacey	test_or	1y ago	1m 46s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
atomic-feather-412	0.6913	8.319e-8	0.4658	finished	Add notes	stacey	test_or	1y ago	1m 46s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
sunny-cloud-411	0.6291	8.615e-8	0.485	finished	Add notes	stacey	test_or	1y ago	1m 50s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
fragrant-bee-410	0.346	8.774e-8	0.2778	failed	Add notes	stacey	test_or	1y ago	22s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
soft-eon-408	0.3354	5.887e-8	0.3211	failed	Add notes	stacey	test_or	1y ago	21s	-	-	-	-	8	-	true	resnet18	10	fast.ai	[360,640]
glad-sweep-197 (resr	0.8519	0.0000182	0.7681	finished	9	stacey		1y ago	10m 26s	lnrd5iw8	-	-	-	7	-	true	resnet34	10	fast.ai	[360,640]
major-sweep-196	0.8698	0.0000198	0.7909	finished	best m...	stacey		1y ago	10m 16s	lnrd5iw8	-	-	-	7	-	true	resnet34	10	fast.ai	[360,640]
grateful-sweep-193	0.8562	2.811e-7	0.7739	finished	Add notes	stacey		1y ago	10m 24s	lnrd5iw8	-	-	-	8	-	true	resnet34	10	fast.ai	[360,640]
restful-sweep-190	0.8049	7.120e-8	0.6796	failed	stoppe...	stacey		1y ago	5m 15s	lnrd5iw8	-	-	-	8	-	true	resnet34	10	fast.ai	[360,640]
dark-sweep-189 (resr	0.8457	0.002082	0.7547	finished	5	stacey		1y ago	10m 16s	lnrd5iw8	-	-	-	8	-	true	resnet34	10	fast.ai	[360,640]
snowy-sweep-185	0.8631	0.0000150	0.7667	finished	Add notes	stacey		1y ago	10m 37s	lnrd5iw8	-	-	-	7	-	true	resnet34	10	fast.ai	[360,640]
expert-sweep-184	0.8365	7.606e-8	0.7594	finished	Add notes	stacey		1y ago	10m 37s	lnrd5iw8	-	-	-	7	-	true	resnet34	10	fast.ai	[360,640]
sleek-sweep-181	0.8609	7.710e-8	0.7728	finished	Add notes	stacey		1y ago	10m 18s	lnrd5iw8	-	-	-	8	-	true	resnet34	10	fast.ai	[360,640]

Keeping track of all things!

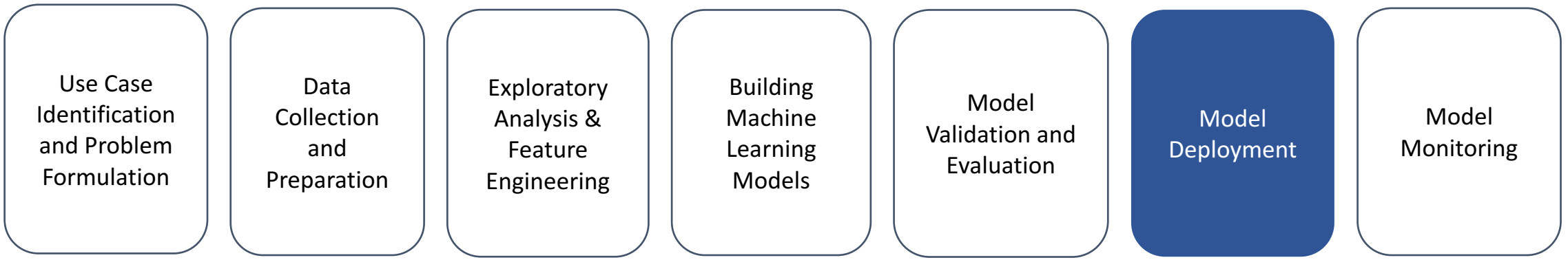


Any change in samples will effect the performance of the model.

Any change in imputation, scaling and encoding techniques will change the performance of the model.

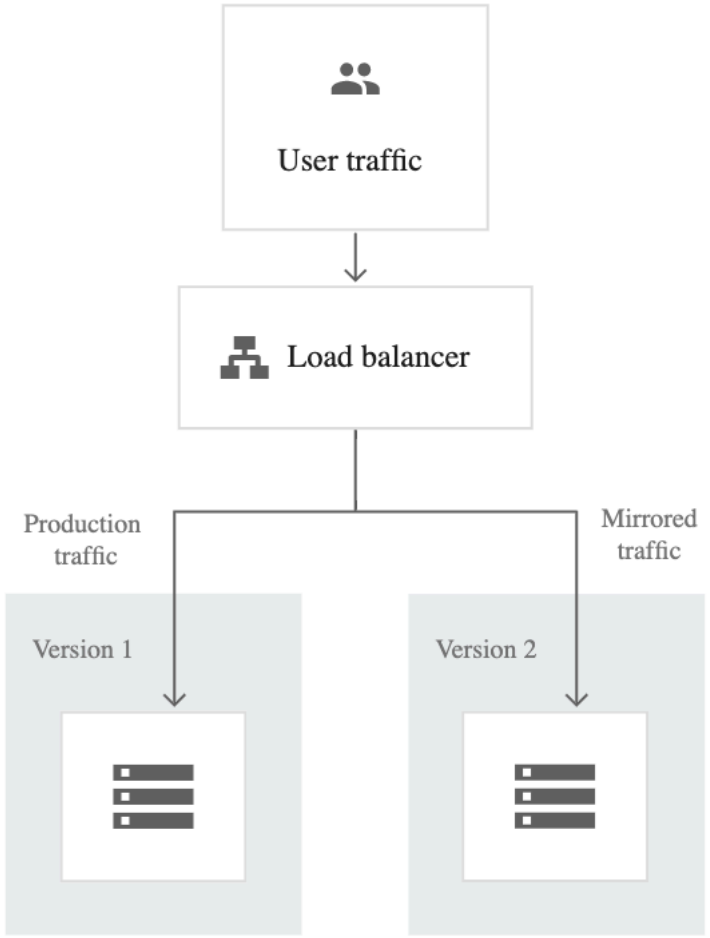
Any change in hyperparameters will effect the performance of the mode.

ML Lifecycle



What if there is no trust in the models?

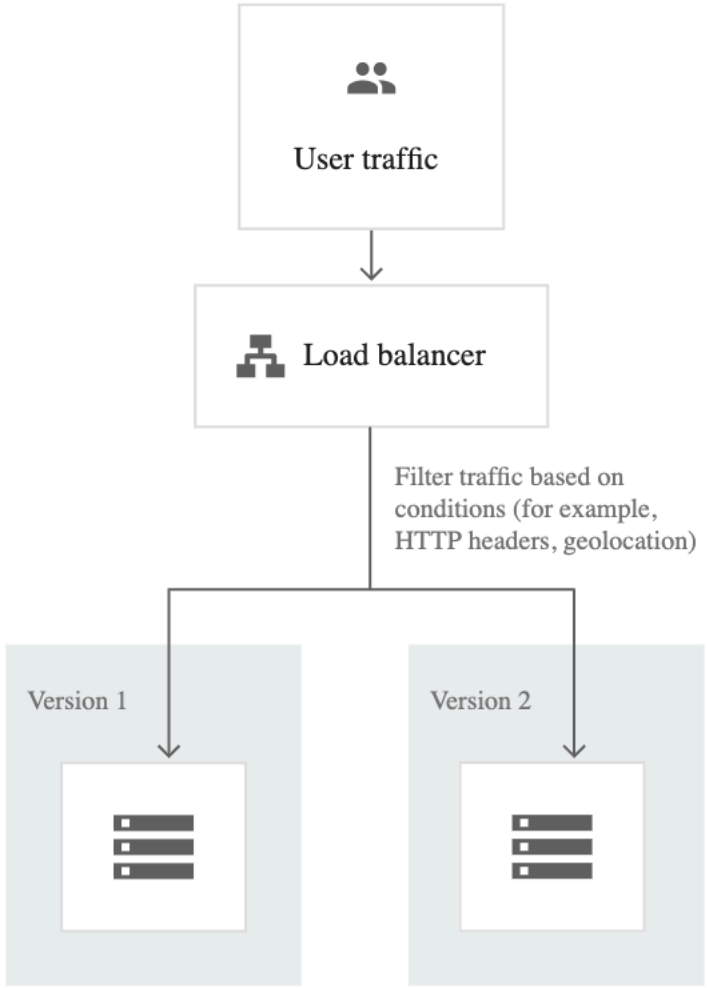
Shadow Deployment



- **Shadow deployment**:
  - Route all inferences through both old and new system.
  - Use old system's prediction for decision making, but new system's prediction for monitoring it's performance against old system

https://cloud.google.com/architecture/application-deployment-and-testing-strategies

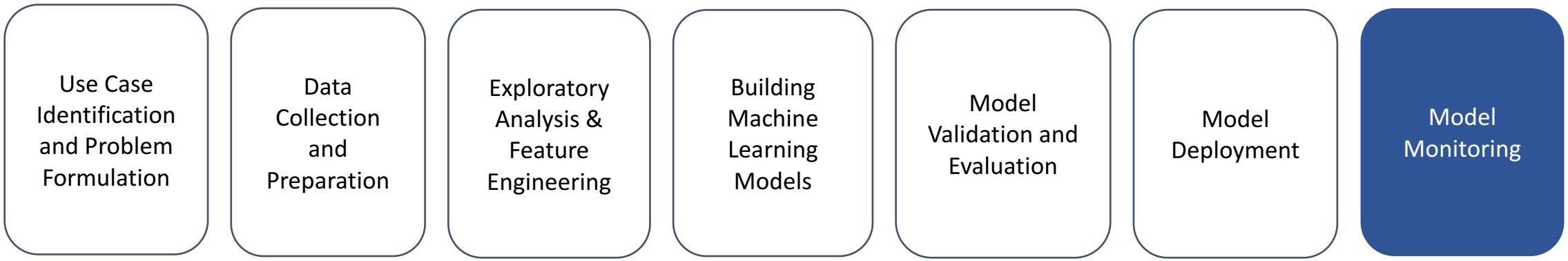
A/B Deployment



- **A/B Deployment:**
  - Run a more principled statistical test like t-test or chi-square test
  - Random route old and new systems users and monitor the performance

https://cloud.google.com/architecture/application-deployment-and-testing-strategies

ML Lifecycle



What is the life expectancy of the models?

Model Monitoring

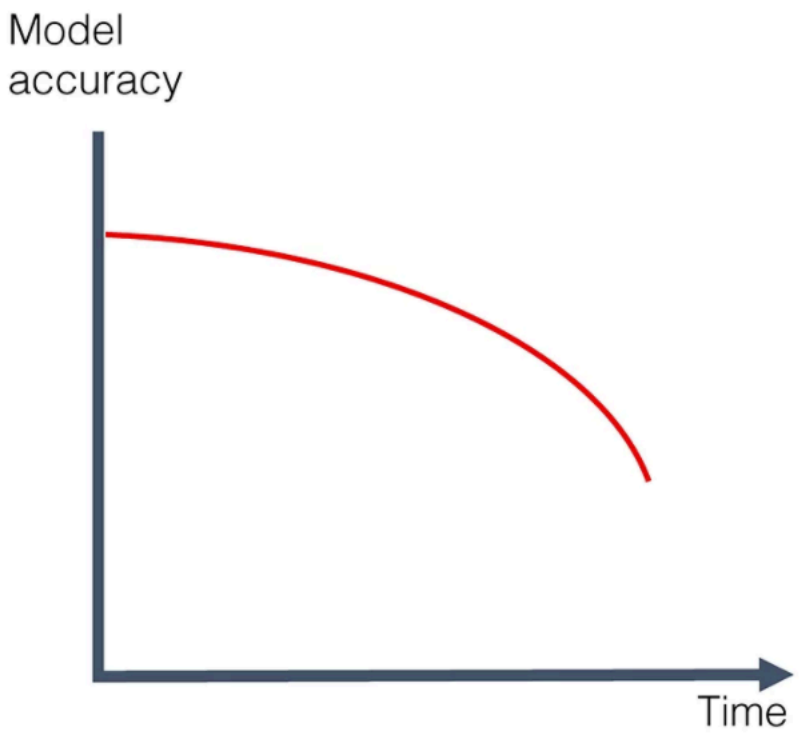
NEWSLETTERS • EYE ON A.I.

How Instacart fixed its A.I. and keeps up with the coronavirus pandemic

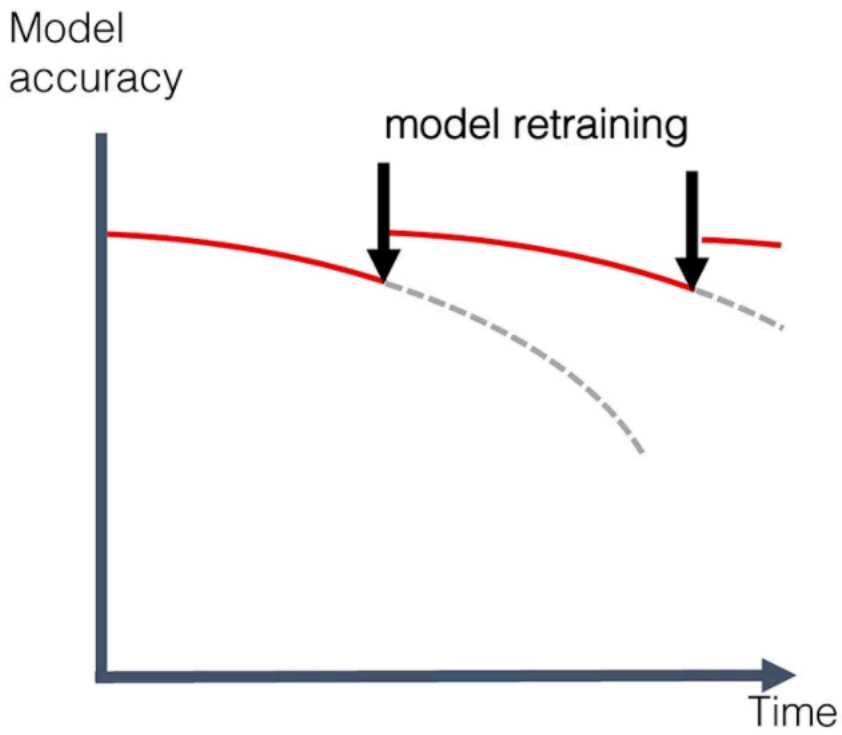
Starting in mid-March, Instacart's all-important technology for predicting whether certain products would be available at specific stores became increasingly inaccurate. The accuracy of a metric used to evaluate how many items are found at a store dropped to 61% from 93%, tipping off the Instacart engineers that they needed to re-train their machine learning model that predicts an item's availability at a store. After all, customers could get annoyed being told one thing—the item that they wanted was available—when in fact it wasn't, resulting in products never being delivered. "A shock to the system" is how Instacart's machine learning director Sharath Rao described the problem to *Fortune*.

https://fortune.com/2020/06/09/instacart-coronavirus-artificial-intelligence/

Model Decays over time



Model decay over time

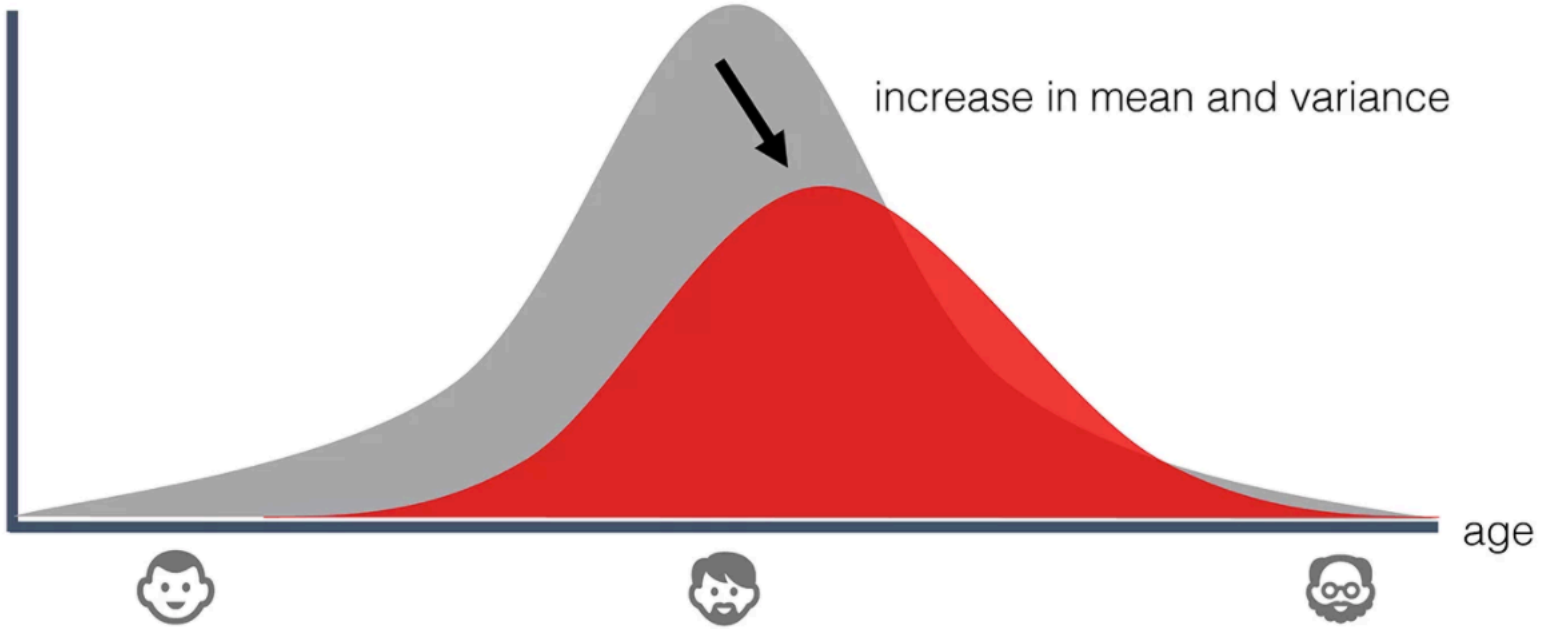


Regularly updated model

https://evidentlyai.com/blog/machine-learning-monitoring-data-and-concept-drift

Data Drift

- Distribution of features change in production compared to training data



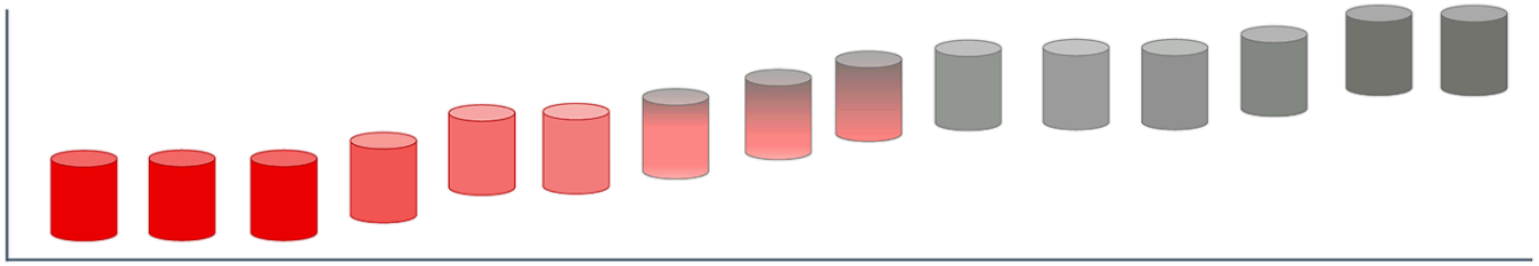
https://evidentlyai.com/blog/machine-learning-monitoring-data-and-concept-drift

Why the model decay need to be monitored

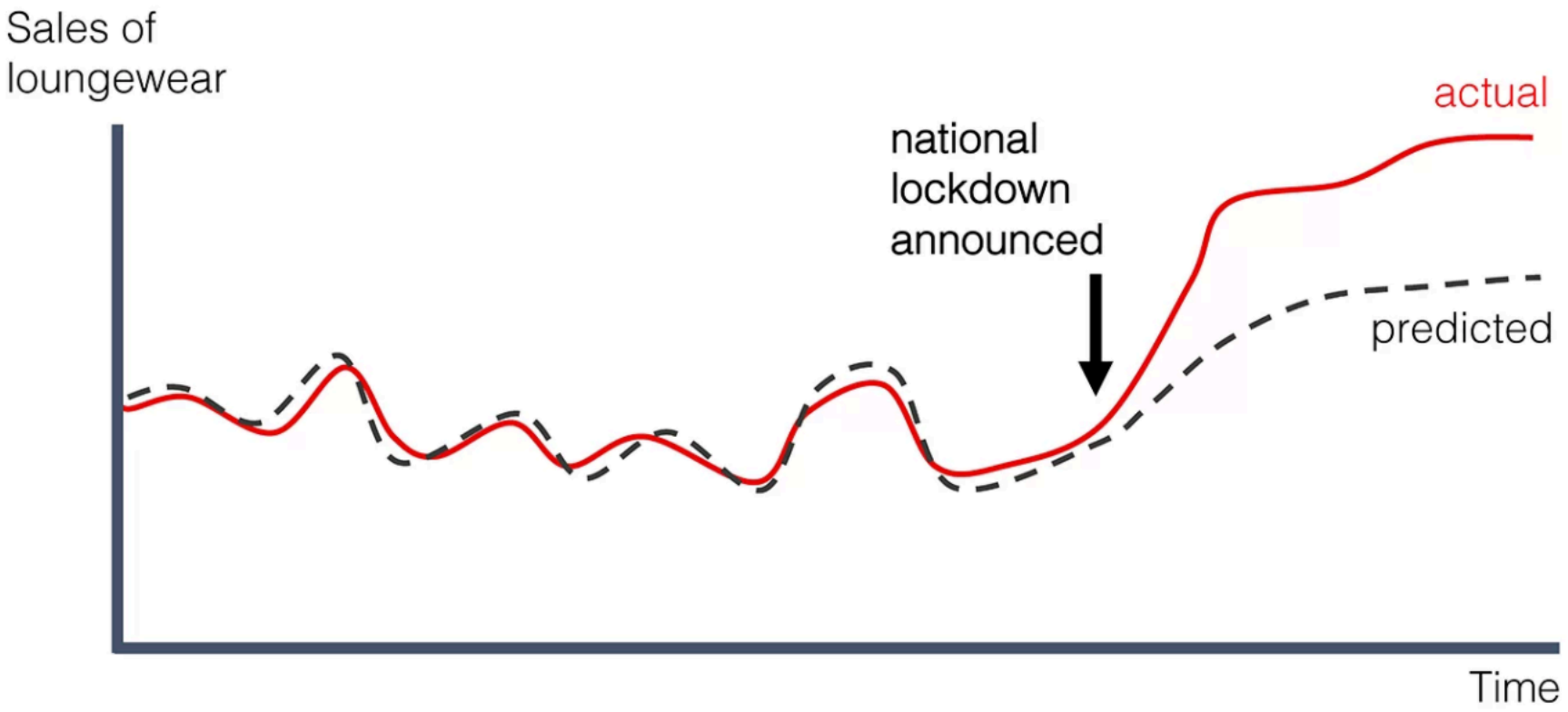
- Risk arising out of model predictions
  - Fraud detection
- May lead to wrong decision making
  - Demand Prediction
- Loss of revenue
  - Customer churn
- The system may loose trust among it’s users
  - Health care diagnosis or predictions

Concept Drift: Gradual

- **Change in Macroeconomic conditions**: As some borrowers default on their loans, the credit risk is redefined. Scoring models need to learn it.
- **Mechanical wear of equipment**: Under the same process parameters, the patterns are now slightly different. It affects quality prediction models in manufacturing.



Sudden Concept Drift



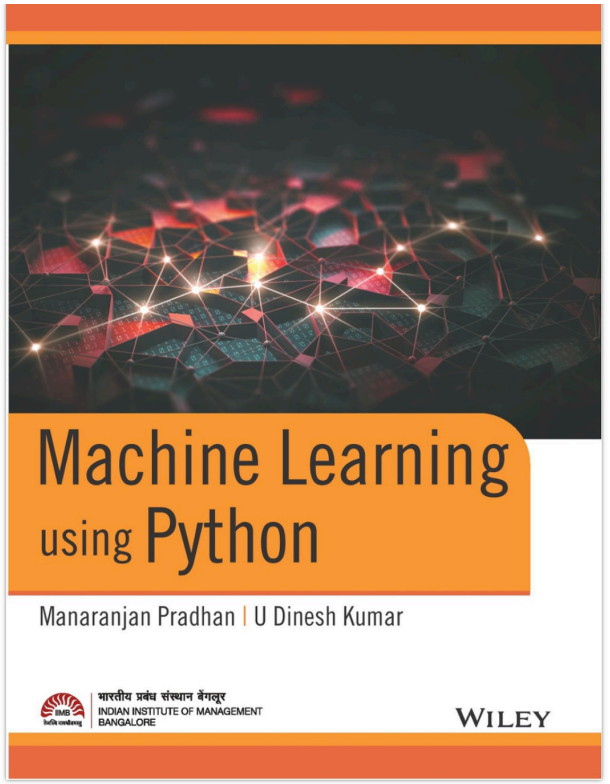
Dealing with Drift

- Retrain the model
  - using all available data, both before and after the change.
  - Use everything, but assign higher weights to the new data so that model gives priority to the recent patterns.
  - If enough new data is collected, we can simply drop the past.
- Naive retraining is not always enough
- Sometimes it's best to modify the model scope or the business process

Corporate Training Offerings

- Machine Learning Foundation
- Advanced Machine Learning
- Deep Learning Foundation
- Deep Learning and NLP (Natural Language Processing)
- MLOps and ML Systems Design

Q&A



Available on Amazon and Flipkart