CS 329P : Practical Machine Learning (2021 Fall)

1.1 Course Introduction

Qingqing Huang, Mu Li, Alex Smola
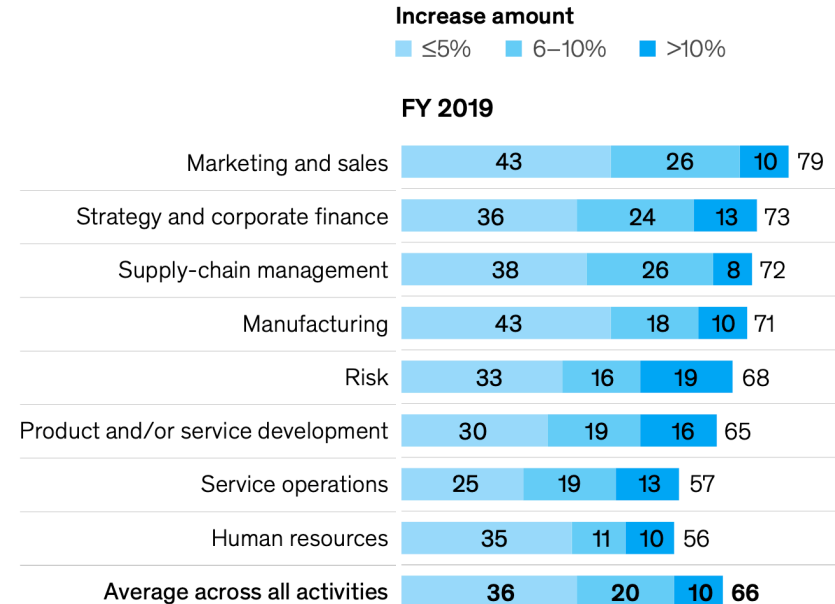
https://c.d2l.ai/stanford-cs329p

Machine Learning (ML) in Industry

- A decade ago ML was mainly used by "Big Tech"

- It's common for companies using ML to drive revenues

  - Top segments are: high-tech, automotive, manufacturing, retail, finance, healthcare

  - Covid-19 accelerated this process

Revenue increase from AI adoption in the previous year (source McKinsey)

Increase amount
■ ≤5%　■ 6–10%　■ >10%

FY 2019

|	≤5%	6–10%	>10%	Total
Marketing and sales	43	26	10	79
Strategy and corporate finance	36	24	13	73
Supply-chain management	38	26	8	72
Manufacturing	43	18	10	71
Risk	33	16	19	68
Product and/or service development	30	19	16	65
Service operations	25	19	13	57
Human resources	35	11	10	56
Average across all activities	**36**	**20**	**10**	**66**

Industrial ML Applications

Manufacturing | Predictive maintenance, quality control

Retail | Recommendation, chatbot, demand forecasting

Healthcare | Alerts from real-time patient data, disease identification

Finance | Fraud detection, application processing

Automobile | Breakdown prediction, self-driving

House Sales Prediction

- The goal is to predict the bid price for the winning buyer

Estimated price
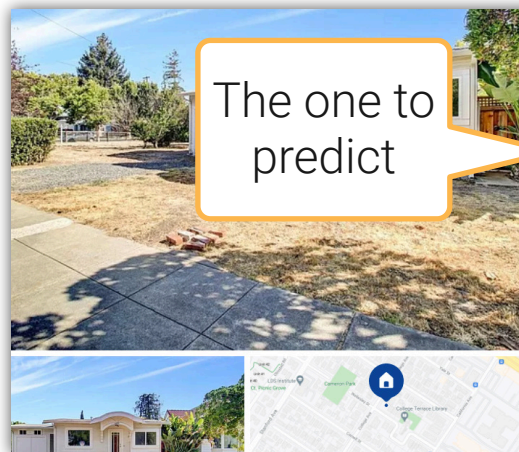
The one to predict

8,873 Square Feet
2239 Wellesley St, Palo Alto, CA 94306

Sold: $3,395,000 | Sold on 08/30/21 | Zestimate®: $3,410,600

Est. refi payment: $14,558/mo | $ Refinance your loan

Home value   Owner tools   Home details   Neighborhood details

7/12/2021   Listed for sale $3,295,000 (+11.3%) $3,168/sqft
Source: MLSListings Inc Report

Ask price

ML Workflow
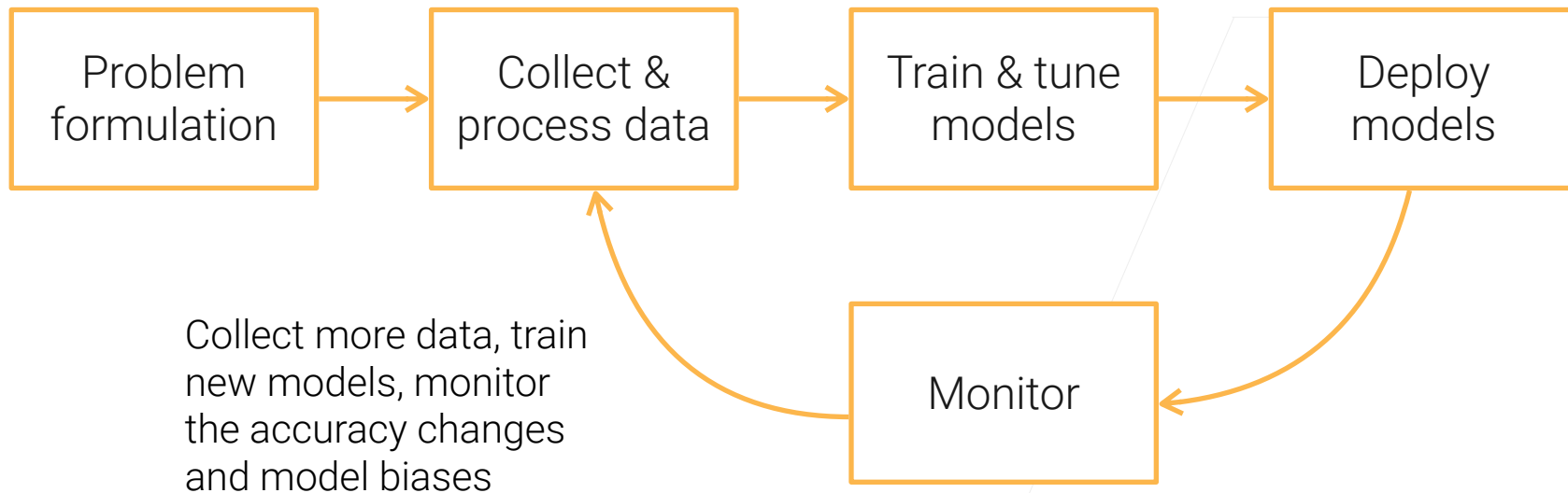
8,873 Square Feet
2239 Wellesley St, Palo Alto, CA 94306

● **Sold: $3,395,000** | Sold on 08/30/21 | Zestimate®: **$3,410,600**

Example:
Predict house sale
prices

Collect past house
sale records,
extract features

Train a linear
regression model

Deploy model
online to predict a
future sale

```
Problem
formulation  →  Collect &
                process data  →  Train & tune
                                 models  →  Deploy
                                            models
                                               ↓
                              Monitor
```

Collect more data, train
new models, monitor
the accuracy changes
and model biases

Industrial ML Applications
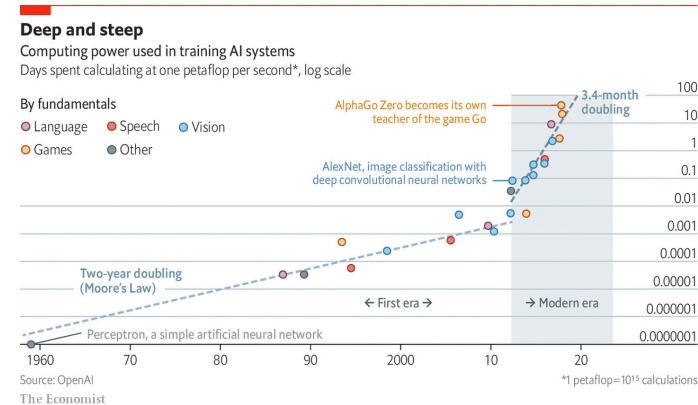
Manufacturing		Predictive maintenance, quality control
Retail		Recommendation, chatbot, demand forecasting
Healthcare		Alerts from real-time patient data, disease identification
Finance		Fraud detection, application processing
Automobile		Breakdown prediction, self-driving

Challenges at ML stages

- <u>Formulate problem:</u> focus on the most impactful industrial problems

- <u>Data:</u> high-quality data is scarce, privacy issues

- <u>Train models:</u> ML models are more and more complex, data-hungry, expensive

- <u>Deploy models</u>: heavy computation is not suitable for real-time inference

- <u>Monitor</u>: data distributions shifts, fairness issues

Deep and steep
Computing power used in training AI systems
Days spent calculating at one petaflop per second*, log scale

By fundamentals
○ Language ● Speech ○ Vision
○ Games ● Other

AlphaGo Zero becomes its own
teacher of the game Go

3.4-month doubling

AlexNet, image classification with
deep convolutional neural networks

Two-year doubling
(Moore's Law)

← First era →   → Modern era

Perceptron, a simple artificial neural network

1960   70   80   90   2000   10   20

Source: OpenAI   *1 petaflop=10¹⁵ calculations
The Economist

Course Topics

- Techniques a data scientist needs but often not taught in university ML/stats/programming courses

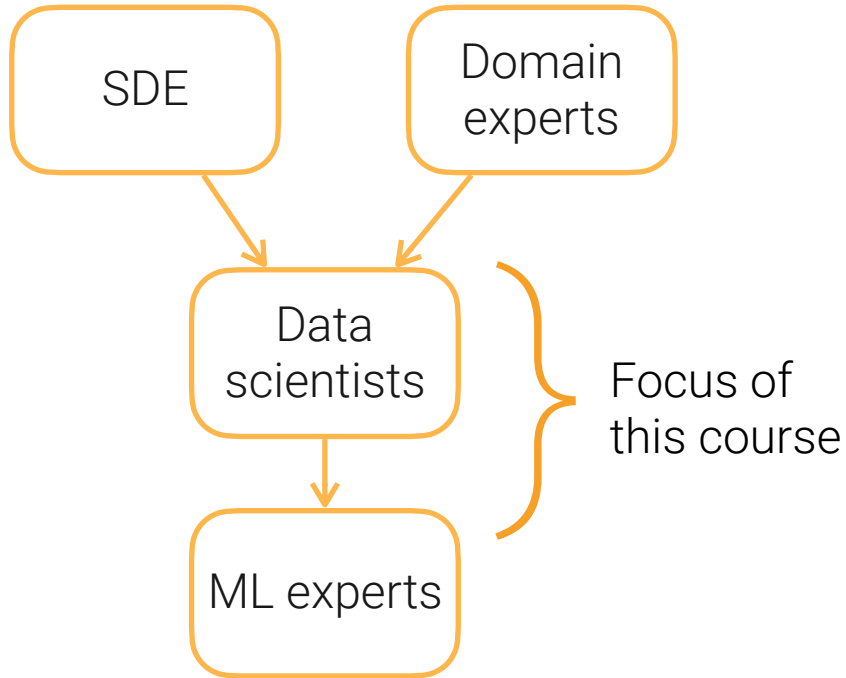| Data	Train	Deploy	Monitor
• Collect/ preprocess data • Covariate/ concepts/label shifts • Data beyond IID	• Model validation/ combinations/ tuning • Transfer learning • Multi-modality	• Model deployment • Distillation	• Fairness • Explainability

Roles

- **Domain experts:** have business insights, know what data is important and where to find it, identify the real impact of a ML model

- **Data scientists:** full stack on data mining, model training and deployment

- **ML experts:** customize SOTA ML models

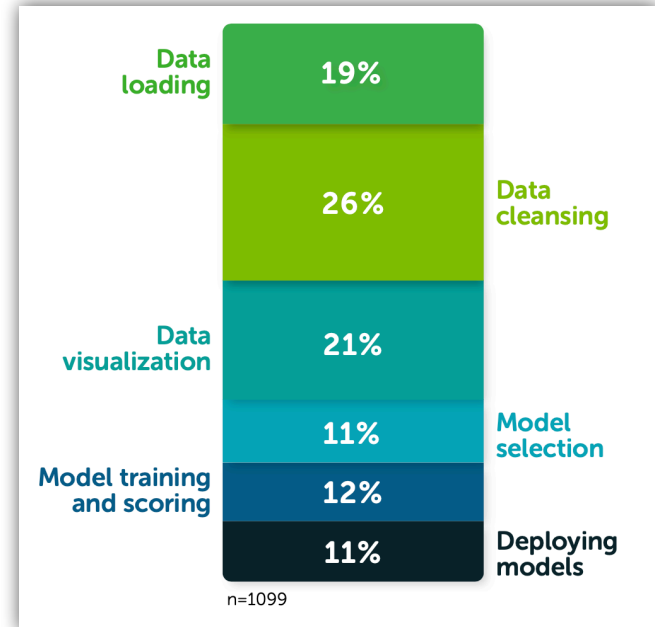- **SDE:** develop/maintain data pipelines, model training and serving pipelines

Roles

Skill improvement

Focus of this course

How data scientists spent their time (source: Anaconda survey 2020)

Data loading 19%

Data cleansing 26%

Data visualization 21%

Model selection 11%

Model training and scoring 12%

Deploying models 11%

n=1099

Summary

- ML has become a staple of modern business.

- A ML workflow includes: formulating the problem, preparing data, training and deploying ML models, monitoring

- This course will teach technologies a data scientist needs in each ML workflow stage