Applied Machine Learning for Business Analytics

Lecture 12: Why do ML Projects Fail in Business

Lecturer: Zhao Rui

Logistics

- Your Voice Matters
  - BT5153 requires training data for better performance in the future
  - https://es.nus.edu.sg/blue/
  - Thanks ahead for your time!
- Group projects submission is due @ 11:59 pm, 24 Apr
- Appreciate if you keeps video on!

Course Projects can also be cool

Neural networks for fashion image classification and visual search

F Li, S Kant, S Araki, S Bangera, SS Shukla - arXiv preprint arXiv …, 2020 - arxiv.org

We discuss two potentially challenging problems faced by the ecommerce industry. On relates to the problem faced by sellers while uploading pictures of products on the platf for sale and the consequent manual tagging involved. It gives rise to misclassifications leading to its absence from search results. The other problem concerns with the potent bottleneck in placing orders when a customer may not know the right keywords but has visual impression of an image. An image based search algorithm can unleash the true

☆ Save 99 Cite Cited by 3 Related articles All 2 versions »

https://arxiv.org/pdf/2005.08170.pdf

Neural Networks for Fashion Image Classification and Visual Search

Fengzi Li | Shashi Kant | Shunichi Araki | Sumer Bangera | Swapna Samir Shukla *

Abstract

We discuss two potentially challenging problems faced by the ecommerce industry. One relates to the problem faced by sellers while uploading pictures of products on the platform for sale and the consequent manual tagging involved. It gives rise to misclassifications leading to its absence from search results. The other problem concerns with the potential bottleneck in placing orders when a customer may not know the right keywords but has a visual impression of an image. An image based search algorithm can unleash the true potential of ecommerce by enabling customers to click a picture of an object and search for related products without the need for typing. In this paper, we explore machine learning algorithms which can help us solve both these problems.

1. Introduction

E-commerce has revolutionized the world of consumerism and unleashed a greater demand of products by providing a trouble-free buying experience and delivery to the user. We present two challenges confronting the industry - one from seller's perspective and the other from buyer's perspective.

A seller who wishes to sell his product on an ecommerce platform has to upload pictures of his product and tag appropriate labels which render the product in search results. Human intervention makes this process prone to errors. Any misclassification thus caused may prevent the product from appearing in the search results thereby being responsible for less sales or no sale at all. Machine learning models can classify the images with high accuracy and prompt the sellers to do appropriate tagging.

There is another challenge that throttles the demand of a customer due to his lack of knowledge of the right keywords. In a typical e-commerce website, a user enters the keywords for the product he wants to purchase. Based on the keywords, the search algorithm matches it with product labels in its database and renders relevant products to the user. A user then explores among the search results to find the product which she wants to buy and places an order. A text-based search relies on the pre-requisite that the customer knows the product very well and knows the right keywords to punch into the search toolbar. However, this is not always the case. We come across many different things in our day-to-day life about which we are not aware. Such cases restrict our ability to search for the product on the e-commerce website. A visual search is an answer to this problem.

When customers undertake a visual search, they look for a product with a photo or other image instead of the keywords normally used in search engines. Shoppers can take a picture of something they want to buy, upload it to the visual search engine and immediately see visually similar items available to purchase. This idea is already being implemented by a lot of AI solution providers such as Visenze, Google Lens etc.

Any visual search algorithm is essentially a unsupervised problem wherein machine learning models can be deployed to learn features about the new images and search for similar products. Once the target image is uploaded to the visual search engine, more products of the similar features can be rendered as search results on the website. Typical visual search algorithms such as autoencoders can be used to generate the latent features of images. Besides, transfer learning using pre-trained deep neural network models are also employed to extract the embedding features of images.

To summarise, this paper will pursue two broad objectives:

- **Image Classification**: To train different neural network models to learn from large sets of images of products from an e-commerce website. Use transfer learning with pre-trained models such as VGG19 to do image classification.
- **Image Search**: Use autoencoders and cosine similarity to identify similar images.

*All authors are graduate students of MSBA program Class of 2020 at NUS Business School. Source code can be found at https://github.com/swapnasamirshukla/BT5153-Term-Project .

BT5153 Applied Machine Learning Term Paper, MSBA @ NUS, 2020.

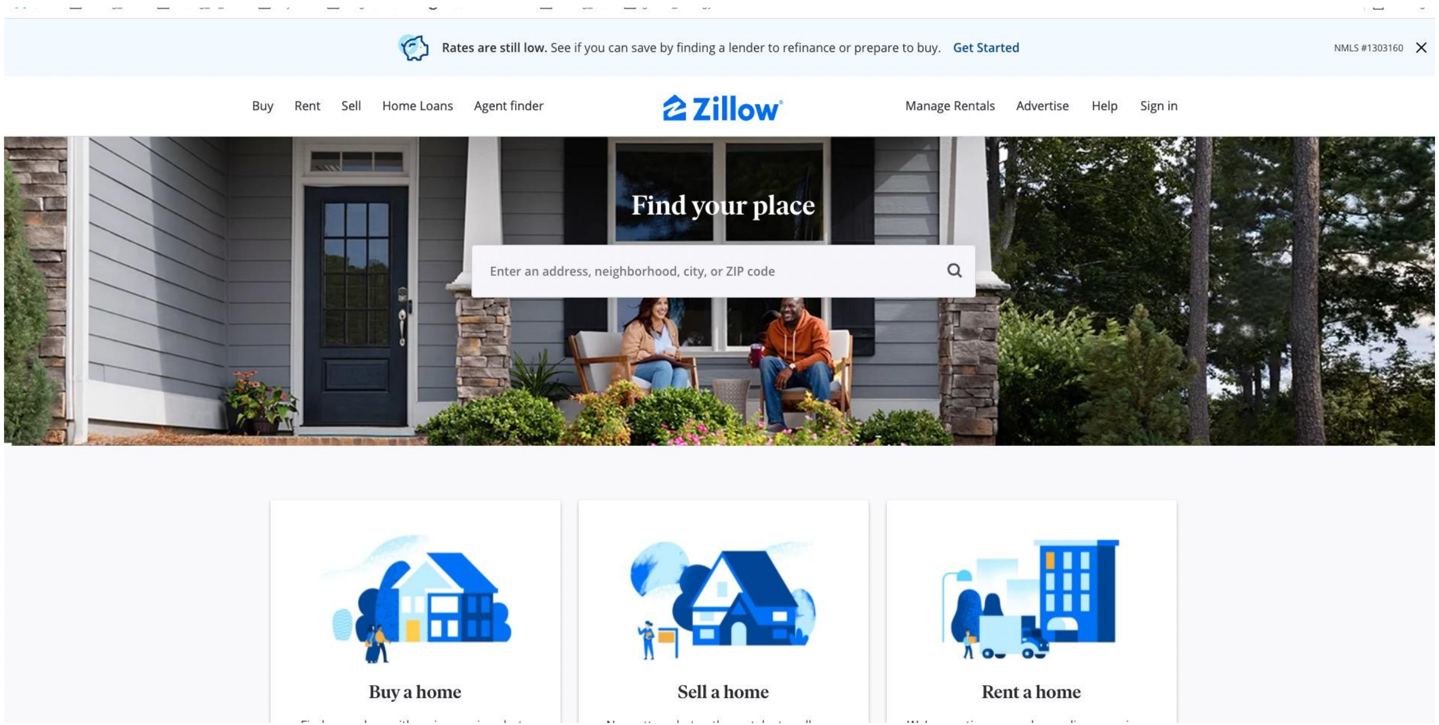
Agenda

1. Zillow Offers
2. Causes of ML Failures
3. Understanding Data Science and Analytics Roles
4. Courser Summary

1. Zillow Offers

Zillow

- an America online real-estate marketplace company



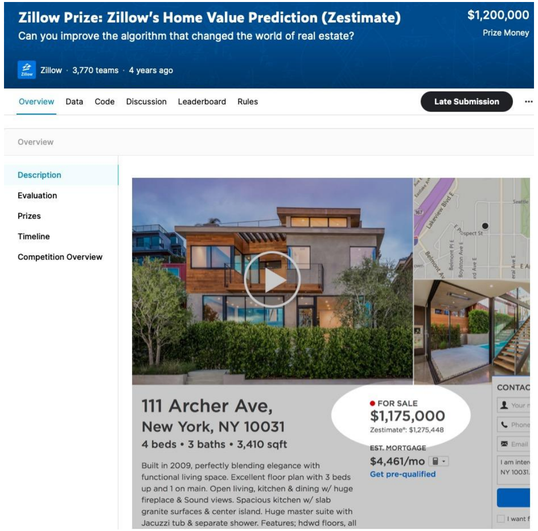
Zillow’s Business Model

- How does Zillow make money?
  - https://seoaves.com/zillow-business-model-how-does-zillow-make-money
- Zillow Offers:
  - The system called “Zestimate” will analyze multiple data and predict the housing price as the bids for the seller
  - A home sale can be completed in a matter of hours
  - Due to the speed and convenience of the sale process, zillow can purchase houses below market value. After repairs and simple renovation, the houses could be sought to a new buyer at a higher price.
  - The delta in the price + commission fees charged from buyer and seller are the gross profits

1M Kaggle Competition

- At 2018, Zestimate launched a one-million kaggle competition
- In the competition, you will build machine learning models to predict the log error between the actual sale price and the Zestimate based on all the features of a home

$$logerror = log(Zestimate) - log(SalePrice)$$



https://www.kaggle.com/c/zillow-prize-1

1M Kaggle Competition

- The winning solution pushed the Zestimate’s current nationwide error rate of 4.5% to below 4%

Public | Private

The private leaderboard is calculated with approximately 49% of the test data.
This competition has completed. This leaderboard reflects the final standings.

Prize Winners

#	△	Team	Members	Score	Entries	Last	Code
1	▲ 9	Zensemble		0.07408	253	4Y	
2	▲ 1	Juan Zhai 卷宅		0.07421	53	4Y	

Meet the ‘Zillow Prize’ winners who get $1M and bragging rights for beating the Zestimate

BY KURT SCHLOSSER on January 30, 2019 at 5:00 am

Share 102 | Tweet | Share | Reddit | Email

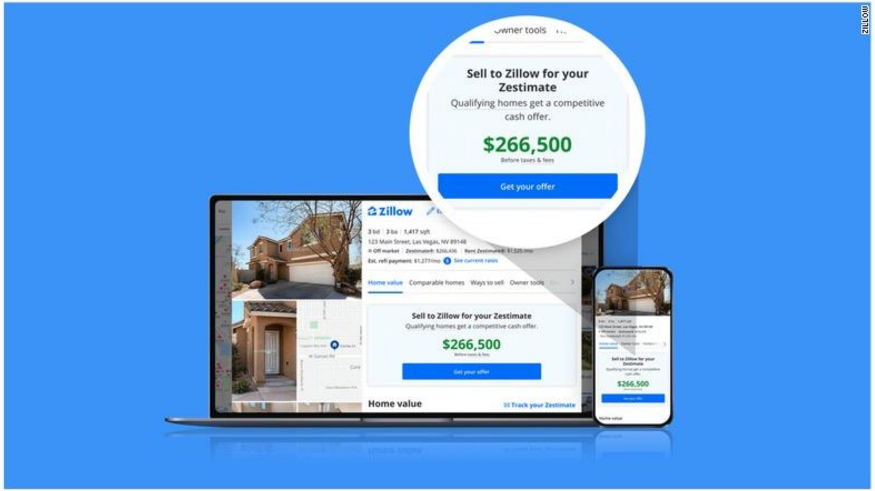
Zillow Chief Analytics Officer Stan Humphries, left, presents a check to Nima Shahbazi, a member of the team that won the Zillow Prize. (Zillow Photo)

Zestimates was deployed to make offers

Zillow will now make cash offers for homes based on its 'Zestimates'

By Clare Duffy, CNN Business

Updated 1550 GMT (2350 HKT) February 25, 2021



Zillow's "Zestimates" will now represent initial cash offers to homeowners in some markets.

https://edition.cnn.com/2021/02/25/tech/zillow-zestimate-cash-offer/index.html

Stock Price in the past 6 months

What went wrong with Zillow Offers?

Zillow, facing big losses, quits flipping houses and will lay off a quarter of its staff.

The real estate website had been relying on its algorithm that estimates home values to buy and resell homes. That part of its business lost about $420 million in three months.

Zillow is sitting on thousands of houses worth less than what the company paid for them. Caitlin O'Hara for The New York Times

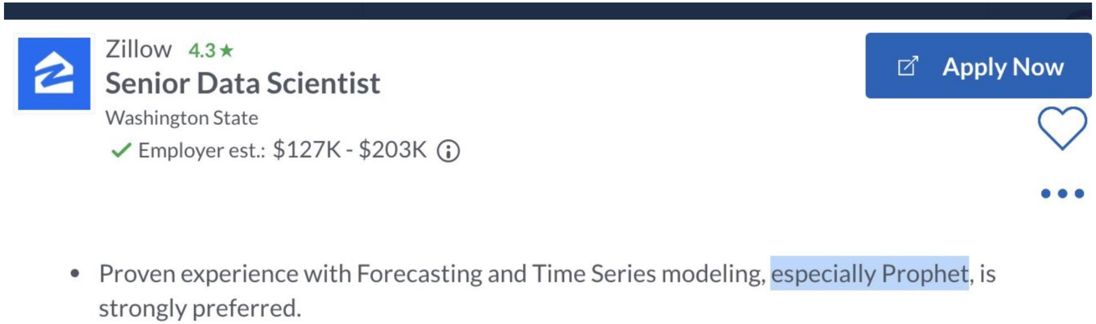
https://www.nytimes.com/2021/11/02/business/zillow-q3-earnings-home-flipping-ibuying.html

What went wrong with Zillow Offers?

1. Use ML to predict home prices
2. Use predicted prices to flip houses
3. ML models over-predict house prices
4. Buy houses at higher prices

Blaming game

1. Prophet: A python library for forecasting
2. Kaggle-style data science
3. Leadership
4. ML/DS team



More insightful discussion:
https://ryxcommar.com/2021/11/06/zillow-prophet-time-series-and-prices/

Prophet

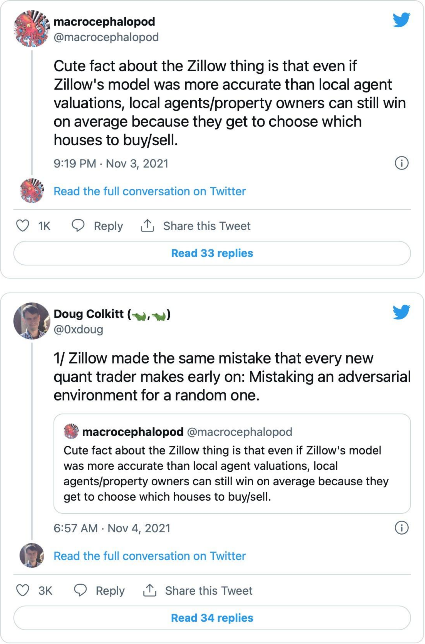
- The model is developed by Facebook to predict the web traffic
  - Housing price do not have strong seasonality pattern which is different from web traffic

This is an accurate description of what Prophet's model is. To get a little more in the weeds, Prophet does the following linear decomposition:

- `g(t)`: Logistic or linear growth trend with optional linear splines (linear in the exponent for the logistic growth). The library calls the knots "change points."
- `s(t)`: Sine and cosine (i.e. Fourier series) for seasonal terms.
- `h(t)`: Gaussian functions (bell curves) for holiday effects (instead of dummies, to make the effect smoother).

Adverse Selection

- Even the model accuracy is high, property owners will only sell when the predicted price is higher than their expected price



macrocephalopod
@macrocephalopod

Cute fact about the Zillow thing is that even if Zillow's model was more accurate than local agent valuations, local agents/property owners can still win on average because they get to choose which houses to buy/sell.

9:19 PM · Nov 3, 2021

Read the full conversation on Twitter

1K Reply Share this Tweet

Read 33 replies

Doug Colkitt (🐢,🐢)
@0xdoug

1/ Zillow made the same mistake that every new quant trader makes early on: Mistaking an adversarial environment for a random one.

macrocephalopod @macrocephalopod
Cute fact about the Zillow thing is that even if Zillow's model was more accurate than local agent valuations, local agents/property owners can still win on average because they get to choose which houses to buy/sell.

6:57 AM · Nov 4, 2021

Read the full conversation on Twitter

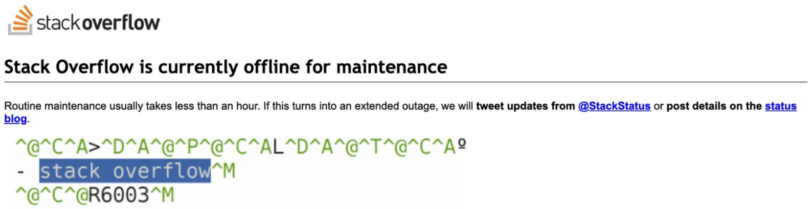
3K Reply Share this Tweet

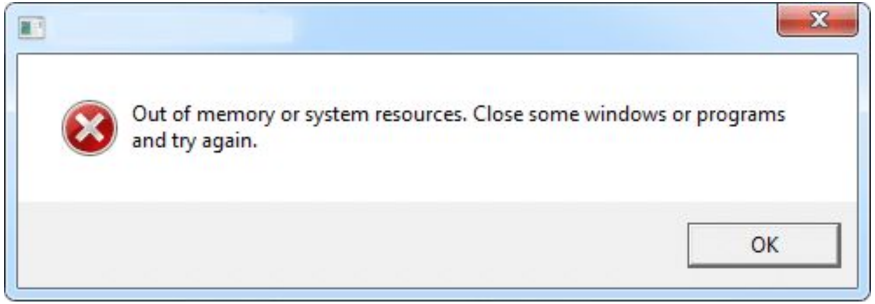
Read 34 replies

2. Causes of ML Failures

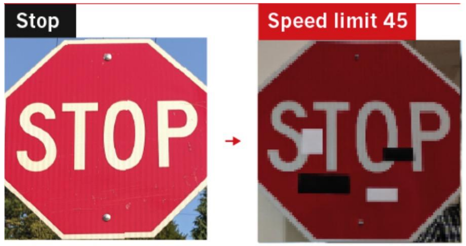
ML systems fail silently

Normal softwares fail





ML systems fail



Amazon scraps secret AI recruiting tool that showed bias against women

That is because Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry.

In effect, Amazon's system taught itself that male candidates were preferable. It penalized resumes that included the word "women's," as in "women's chess club captain." And it downgraded graduates of two all-women's colleges, according to people familiar with the matter. They did not specify the names of the schools.

[Amazon scraps secret AI recruiting tool that showed bias against women](#) (Reuters, 2018)

Japan’s Henn na Hotel fires half its robot workforce

“Guests complained their robot room assistants thought snoring sounds were commands and would wake them up repeatedly during the night.”

https://www.hotelmanagement.net/tech/japan-s-henn-na-hotel-fires-half-its-robot-workforce

What is an ML failure?

A failure happens when one or more expectations of the system is violated.

Two types of expectations:

- Operational metrics: e.g. response time, downtime
- ML metrics: e.g. accuracy, MSE, BLUE score (machine translation)

What is an ML failure?

A failure happens when one or more expectations of the system is violated

- Traditional software: mostly operational metrics
- ML systems: operational + ML metrics
  - Ops: returns the risk scores of users within 800ms latency on average
  - ML: Accuracy as 80%

ML system failures

- If you call API to infer the user’s risk score and get no response-> ops failure
- If the prediction is incorrect -> ML failure?

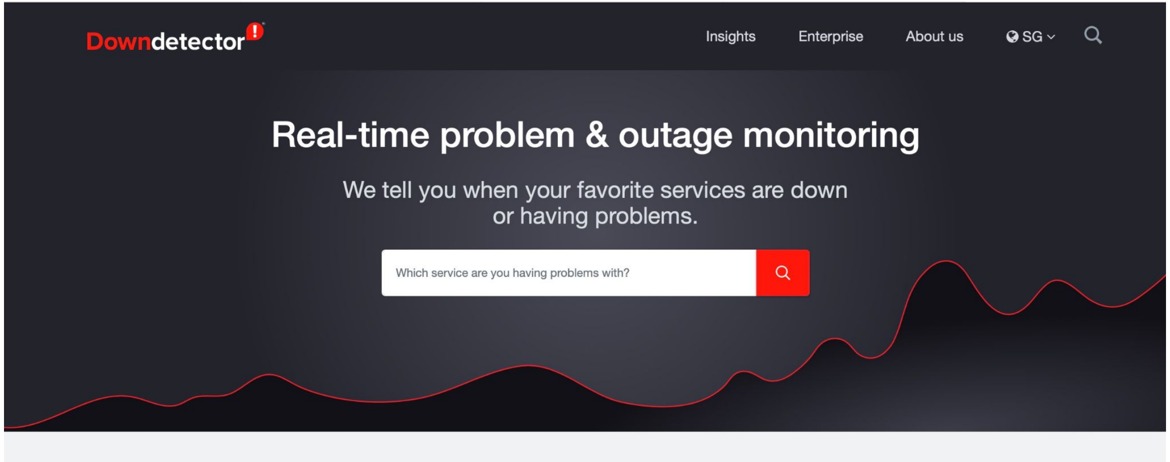
ML system failures

- If you call API to infer the user's risk score and get no response-> ops failure
- If the prediction is incorrect -> Might not be

ML failure when the predictions are consistently wrong

What are Ops Failures

- They are normal software systems' failures:
  - Network issues: downtime / crash
  - Deployment issues
  - Hardware issues
  - Dependencies issues



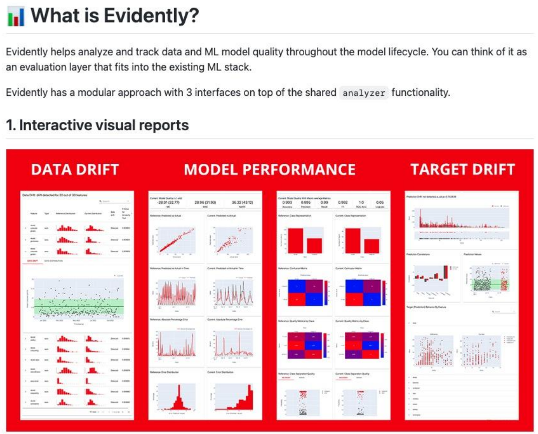
ML-specific failures (during/post deployment)

1. Production data differing from training data
2. Degenerate feedback loops

What are the potential issues for the pre-deployment stage?

Production data differing from training data

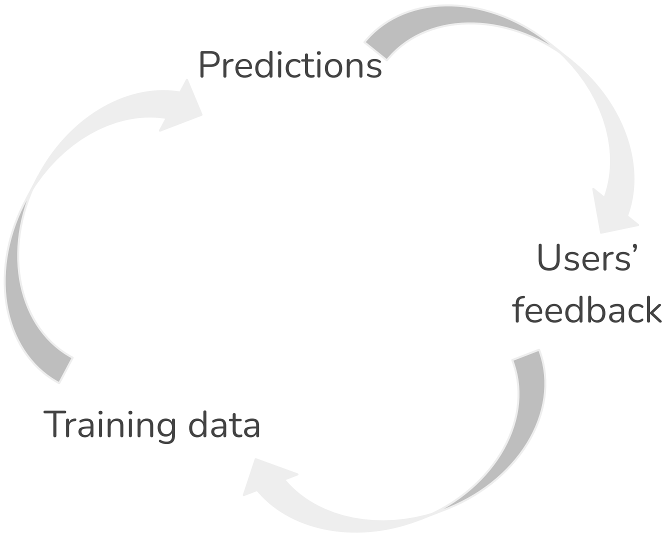
- Train-serving skew:
  - Model performing well during development but poorly after production
- Data distribution shifts
  - Model performing well when first deployed, but poorly over time



https://github.com/evidentlyai/evidently

Degenerate feedback loops

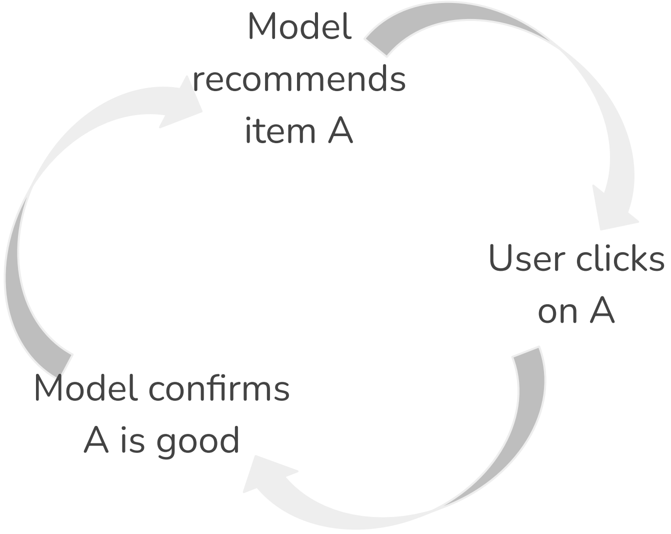
- When predictions influence the feedback, which is then used to extract labels to train the next iteration of the model
- Common in tasks with natural labels



It is not only dangerous for the business, but also for society

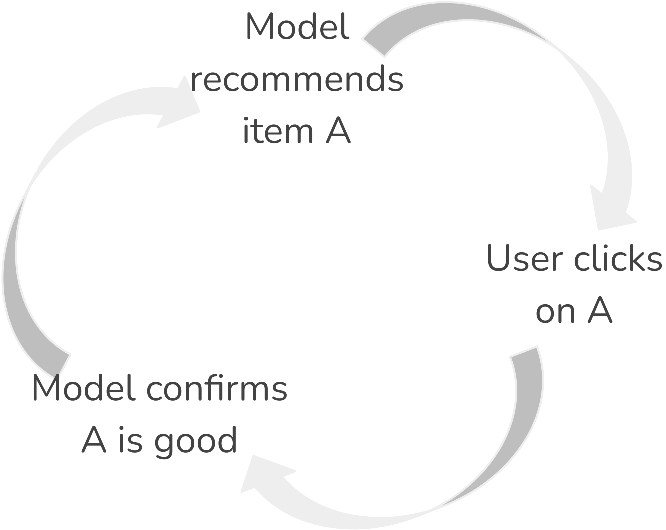
Degenerate feedback loops: recsys

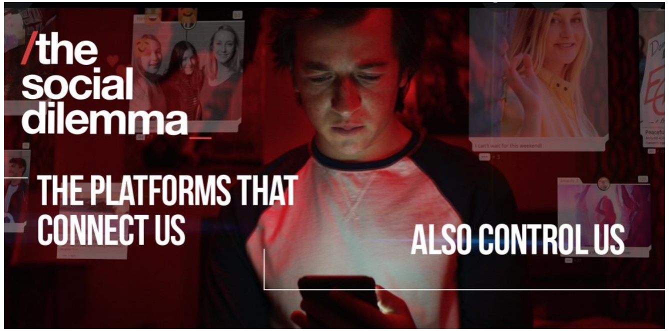
- Originally, A is ranked marginally higher than B -> model recommends A
- After a while, A is ranked much higher than B



Degenerate feedback loops: recsys

- Originally, A is ranked marginally higher than B -> model recommends A
- After a while, A is ranked much higher than B





Degenerate feedback loops: resume screening

- Originally, model thinks X is a good feature
- Model only picks resumes with X
- Hiring managers only see resumes with X, so only people with X are hired
- Model confirms that X is good

Replace X with:

- Has a name that is typically used for gender A
- Went to NUS, MSBA

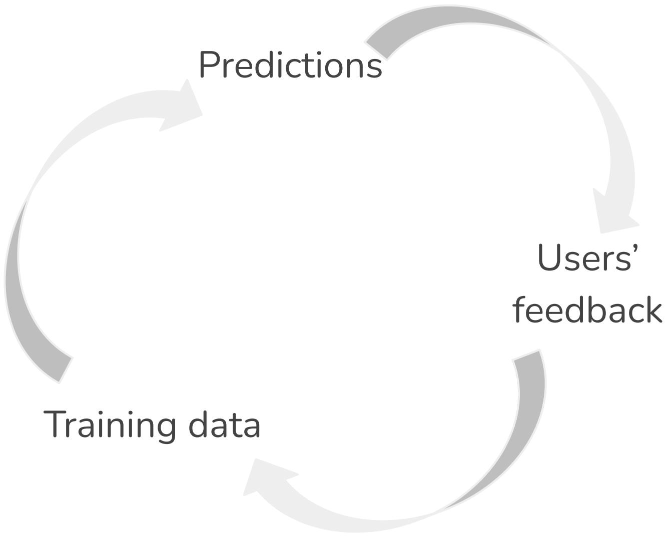
Degenerate feedback loops: resume screening

- Originally, model thinks X is a good feature
- Model only picks resumes with X
- Hiring managers only see resumes with X, so only people with X are hired
- Model confirms that X is good

Tracking feature importance might help!

Detecting degenerate feedback loops

Only arise once models are in production -> hard to detect during training



Well studied in Recommendation System

Degenerate feedback loops: mitigate

1. Randomization
2. Positional features

Degenerate Feedback Loops in Recommender Systems

Ray Jiang, Silvia Chiappa, Tor Lattimore, András György, Pushmeet Kohli
{rayjiang,csilvia,lattimore,agyorgy,pushmeet}@google.com
DeepMind London, UK

https://arxiv.org/abs/1902.10730

3. Understanding Data Science and Analytics Role

Data-Driven Decision

- Understand your business problem
- Identify key challenges and hypothesis in the business problem
- Use data analytics and science methodologies to **test** the hypothesis and **solve** the challenges

How can we make impacts?

- Four stages:
  - How to track/log data?
    - To solve the biz problem, which kind of informations/data are required? -> design trackers
    - It all comes from your business understanding
    - We might need to talk with tech team to design the robust mechanism to make sure the data collection is correct

How can we make impacts?

- Four stages:
  - How to track/log data?
  - How to process data?
    - data cleaning
    - data quality check
    - schema design
    - DE might help to design the raw tables from multiple sources of logs while DSA team needs to design intermedia tables or data mart.

How can we make impacts?

- Four stages:
  - How to track/log data?
  - How to process data?
  - How to analyze data?
    - Dashboard
    - Attribution
    - Metrics design
    - Experimentation
    - Modeling (Data Scientist)

Myth: Modeling is more advanced or DS is on top of DA.

How can we make impacts?

- Four stages:
    - How to track/log data?
    - How to process data?
    - How to analyze data?
        - Dashboard
        - Attribution
        - Metrics design
        - Experimentation
        - Modeling (Data Scientist)

Problem Formulation First, Less Methodology-focused

Simple Tool to solve important problem > Complex tool to solve important problem >> Solve trivial problem

How can we make impacts?

- Four stages:
  - How to track/log data?
  - How to process data?
  - How to analyze data?
  - How to automate the decision-making process?
    - Machine learning is the answer (rule-based system, supervised models and unsupervised models): data -> pattern -> decision

Toy Example

- E-commerce A: CAC in Google Ads per surface enter is 0.01 sgd and the trx conversion rate is 0.1%

Toy Example

- E-commerce A: CAC in Google Ads per surface enter is 0.01 sgd and the trx conversion rate is 0.1%
- Basic Analysis:
  - 10 sgd per transaction, average profit is 20 sgd
  - Double down Google Ads

Toy Example

- E-commerce A: CAC in Google Ads per surface enter is 0.01 sgd and the trx conversion rate is 0.1%
- Advanced Analysis:
  - Product Sense:
    - Another competitor: CAC is 0.1 sgd while the conversion rate is 5%
  - Funnel Analysis
  - A/B Testing
  - User Survey
  - Better Data
    - Users Income
    - Users Demographics Data
    - Other alternative data to support your hypothesis

Advanced analysis is targeted at improving the conversion rate

Open Discussion

- Steve Jobs introduced iphone in 2007
  - It is an art.
  - No data support
- Classify dog vs Cat
  - It is a science or engineering problem
  - Backed by ImageNet data
- Data Science is both an art and a science
  - ML can only solve the engineering problem
  - No need to be obsessed with machine learning

“Import Sklearn” and “Model Fit” are easy which can be replaced by AutoML

4. Course Summary

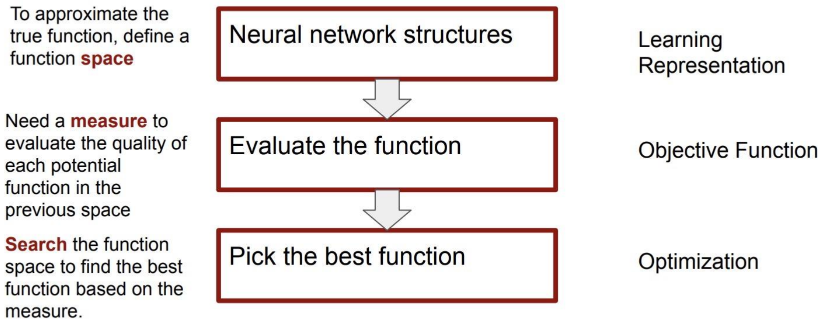
Recap: Bridge the Gap

- Introduction to Machine Learning and its **Application**
    - Gap between theory and practice
- Machine Learning **Practices**
    - Prevent potential issues before model deployment
- **Explainable** Machine Learning
    - Interpretability (remember the case: pokemon vs digimon or jpeg vs png)
- Model **Evaluation** in Machine Learning
    - Modeling
- Model **Deployment** in Machine Learning
    - Production
- Why do ML Projects **Fails** in Business
    - Business Understanding/Product Sense First

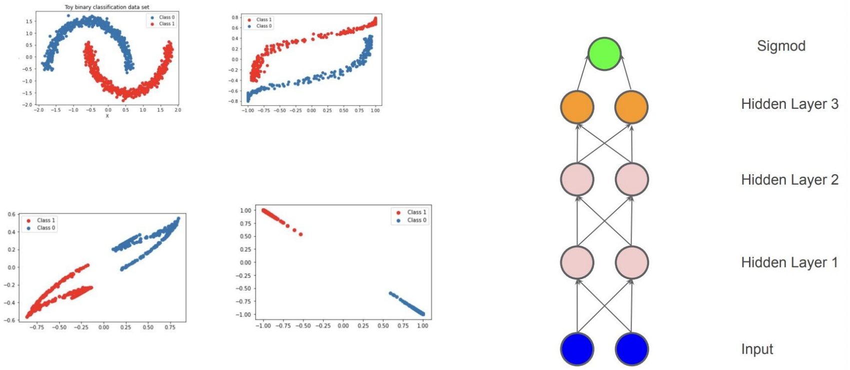
Recap: DL

- Neural Networks and Deep Learning
- Deep Learning Practices
- Auto-encoders
- Convolutional Neural Networks
- Frontiers in NLP
    - **Self-supervised** Training

Three Steps in Deep Learning



Fully-Connected Neural Network



Recap: Data-driven Decision

- Causal Inference for **Decision** Making
    - Data -> Insights -> Decisions
    - Without decisions, data and Insights are both cheap
    - As a data analyst or data scientist, we need to make Impact, Impact, Impact
        - Identify product-market fit
        - Improve strategies
        - Find directions
        - Fix issues
        - Quantify targets
        - Prioritization

Thank You

- Immense thanks to Sanjay and Xiaohui!!!
- Enjoy having all of **you** in BT5153 this year. Appreciate your hard work!