

Data Science + Content

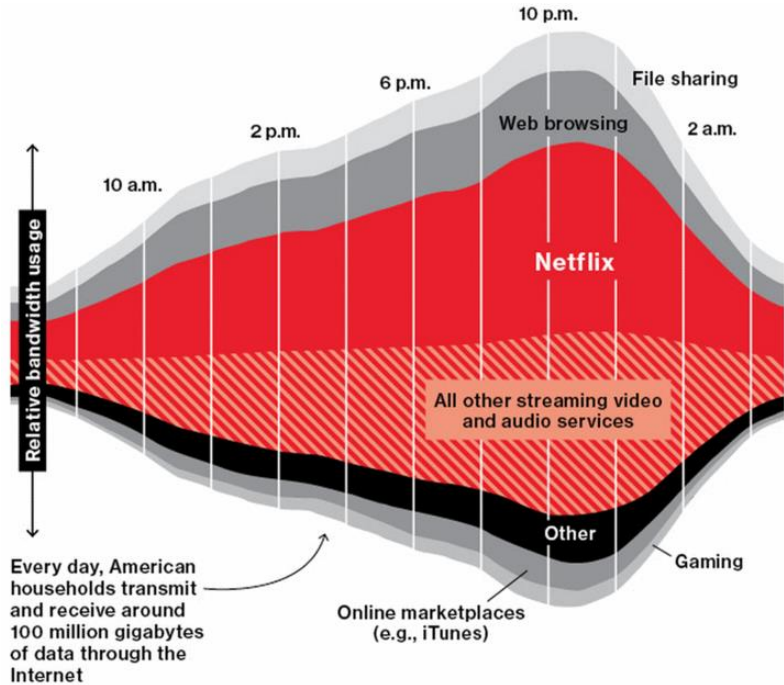
Todd Holloway, Director of Content Science & Algorithms
for Smart Content Summit, 3/9/2017

NETFLIX



NETFLIX

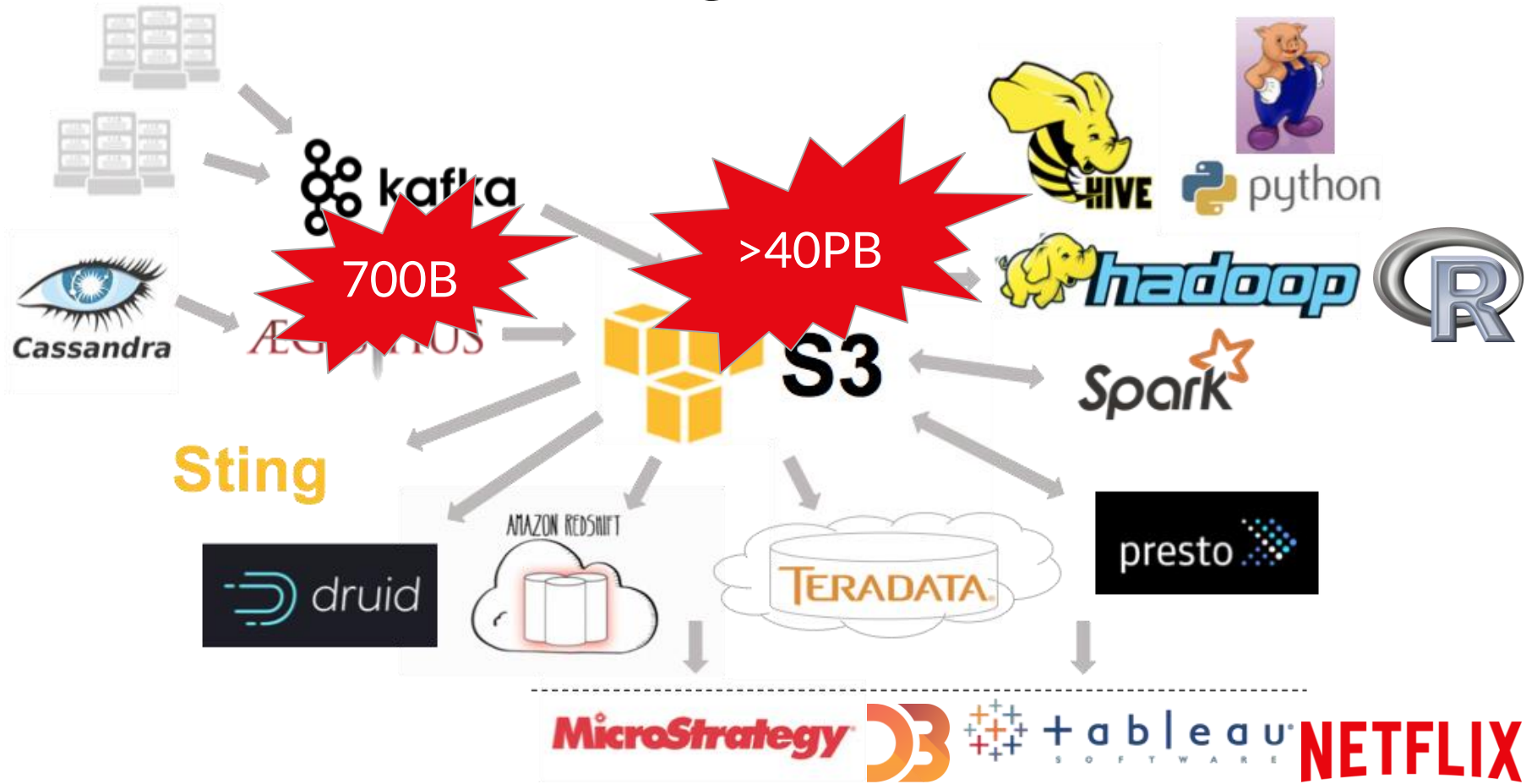
Netflix *by the Numbers...*



- > 90M members
- Available worldwide (except China)
- > 1000 device types
- Hours: > 3B per month
- Log 100B events/day
- 36.5% of peak US downstream traffic

NETFLIX

Data @ Netflix



Evolution of Netflix

DVD service



1999

Streaming service



2007

Originals



2012

Going Global



2016

NETFLIX

Evolution of Machine Learning @ Netflix (2000)

The collage consists of five screenshots from the Netflix website in 2000:

- Top Left:** A window titled "Other Movies You Might Enjoy" with a "Close" button. It displays movie cards for "Amelie", "Y Tu Mama Tambien", "Guys and Balls", "Mostly Martha", "Only Human", and "Russian Dolls". Each card has an "Add" button and a star rating. A message states: "Eiken has been added to your Queue at position 2. This movie is available now. Move To Top Of My Queue".
- Top Right:** A vertical stack of five rating prompts. Each prompt shows a star rating (e.g., 1 star, 2 stars, 3 stars, 4 stars, 5 stars) and a "Not Interested" button. A yellow callout box says "Click to rate the movie 'Hated It'", "Click to rate the movie 'Didn't Like It'", "Click to rate the movie 'Liked It'", "Click to rate the movie 'Really Liked It'", and "Click to rate the movie 'Loved It'".
- Bottom Left:** A "Your DVD Queue" section. It has a table with columns: Title, Star Rating, and Shipped. It lists "Doctor Who: End of Time: Parts 1 & 2: Disc 1" and "The Big Bang Theory: Season 1: Disc 3". Below is a "DVD Queue (6)" table with columns: List Order, Title, Star Rating, and Genre. It lists items like "Doctor Who: End of Time: Parts 1 & 2: Disc 2", "Best of The Muppet Show: Steve Martin", "Narnia: The Voyage of the Dawn Treader", "The Chronicles of Narnia: Prince Caspian", and "The Lion, the Witch & the Wardrobe".
- Bottom Center:** A "movielens" logo with the tagline "helping you find the right movies". It includes a "Shortcuts" section and a "Basic Search" form with fields for Title, Genre, and Domain.
- Bottom Right:** A search results page for "Dark Knight, The (2008)". It shows a "Prediction or Rating" column with 5 stars, a "Your Rating" column with "Not seen", and a "Movie Information" column. It lists popular tags like "superheroes", "Batman", "serial killer", "DVD", "info", "IMDb", "flag", "Crime, Drama, Thriller", "organized crime", "Martin Scorsese", "heist", "ending", "action", "best ending", "New York City", "heist", "Batman Begins (2005)", "DVD", "VHS", "info", "IMDb", "flag", "Action, Crime", "superheroes", "comic book", "Batman", "Gene Baby Gene (2007)", "DVD", "info", "IMDb", "flag".

<http://classic.movielens.org/>

NETFLIX

Evolution of Machine Learning @ Netflix (2007-09)

https://en.wikipedia.org/wiki/Netflix_Prize

Netflix Prize

Home Rules Leaderboard Register Update Submit Download

Leaderboard 10.05% Display top 20 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8558	10.05	2009-06-26 18:42:37
Grand Prize - RMSE <= 0.8563				
2	PragmaticTheory	0.8582	9.80	2009-06-25 22:15:51
3	BellKor in BigChaos			
4	Grand Prize Team			
5	Dace			
6	BigChaos			

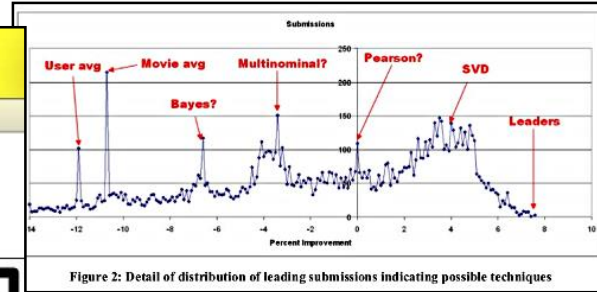
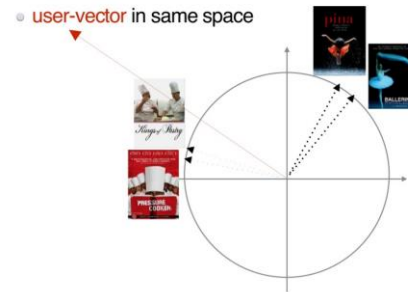


Figure 2: Detail of distribution of leading submissions indicating possible techniques

Videos in a Latent Space Representation



NETFLIX

Machine Learning & Data Science Tools (2017)



NETFLIX

Machine Learning Algorithms (2017)

- Regression models (Logistic, Linear, Elastic nets)
- GBDT/RF
- SVD & other MF models
- Factorization Machines
- Clustering (from k-means to HDP)
- Deep Learning
- LDA
- Association Rules



ML plays a role in every phase of the content lifecycle.

1. Acquisition
2. Quality control
3. Localization
4. Marketing
5. Streaming
6. Presentation to users

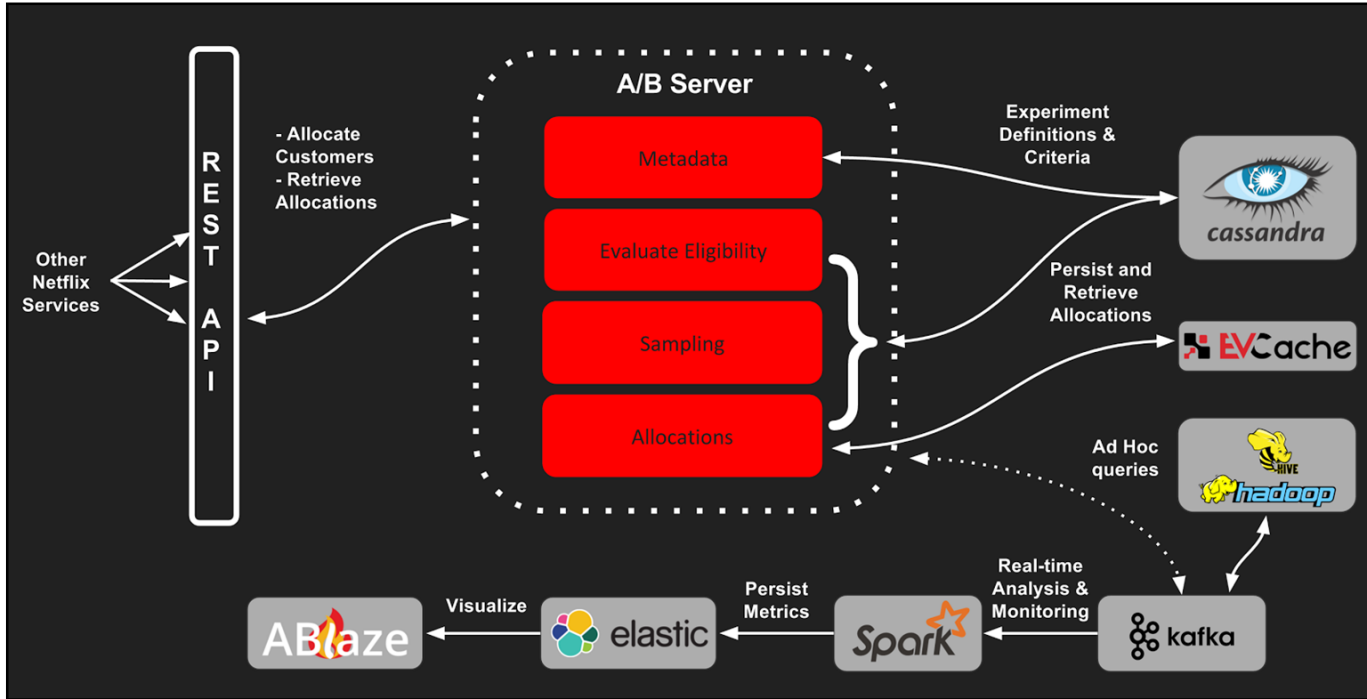
Let's work backward in the lifecycle...

User Interface



Promotion
Layout
Imagery
Metadata
Search Functionality
Row / Content Selection
... and more

User Interface Experimentation Platform



<http://techblog.netflix.com/2016/04/its-all-about-testing-netflix.html>

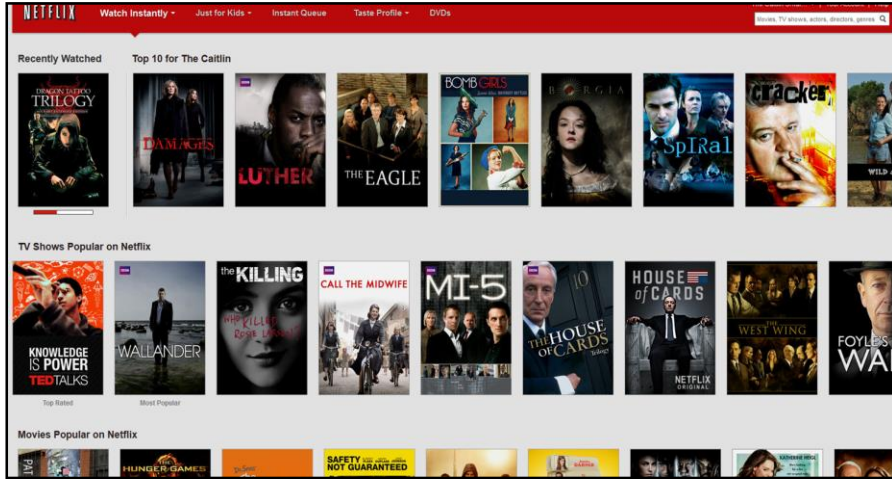
NETFLIX

User Interface Experimentation Platform

Cover Art Optimization



User Interface Experimentation Platform



2013



2015

NETFLIX

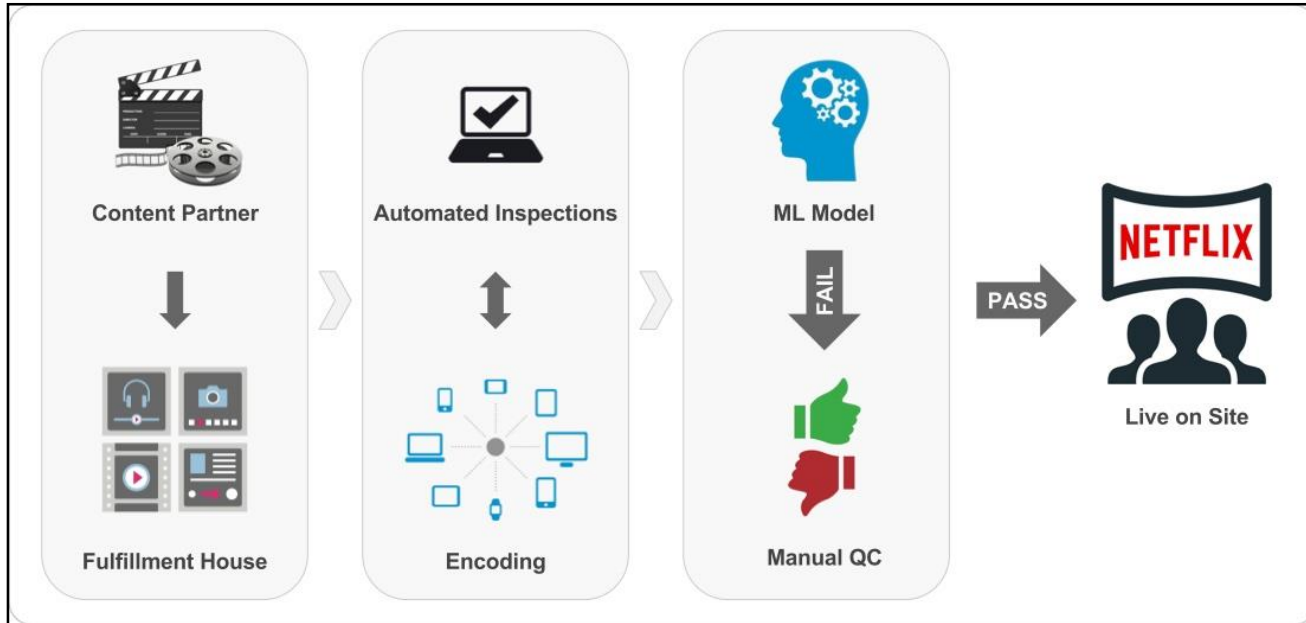
Streaming Optimization



<https://media.netflix.com/en/company-blog/how-netflix-works-with-isps-around-the-globe-to-deliver-a-great-viewing-experience>

NETFLIX

Digital Asset Quality Control



<http://techblog.netflix.com/2015/12/optimizing-content-quality-control-at-netflix-predictive-modeling.html>

NETFLIX

Data Science to Aid Selection of Content

Scripts and pitches



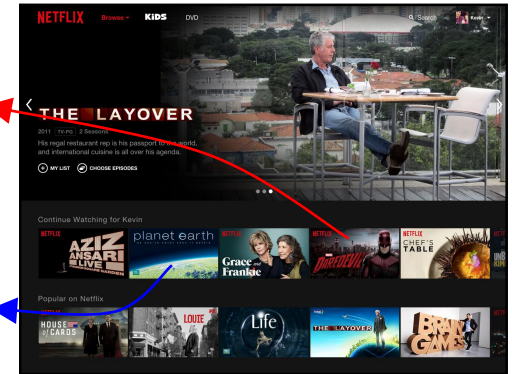
Studio productions

**Selection
process**

Original content



Licensed content



NETFLIX

Can data science create content?



Can a computer write a script that would win a competition? [benjamin.wtf]

NETFLIX

Can data science create content?



He is standing in the stars and sitting on the floor. He takes a seat on the counter and pulls the camera over to his back. He stares at it. He is on the phone. He cuts the shotgun from the edge of the room and puts it in his mouth. He sees a black hole in the floor leading to the man on the roof.

NETFLIX

Is Netflix doing this?

No.

It's the opposite.

We give creative freedom to the
creatives.



NETFLIX

But can data help in choosing content?

Yes.

All decisions are made by experienced creatives,
but analytic products can help.

NETFLIX

Netflix's Notion of Value

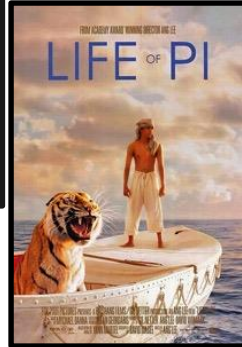
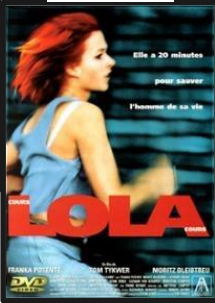
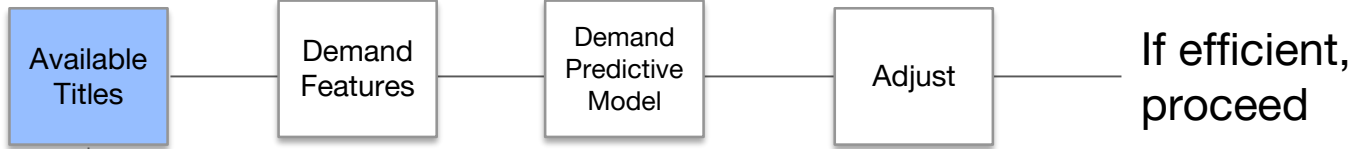
Content Efficiency = value / cost

Netflix's [quarterly shareholder letter](#), released overnight, contained lots of new information. But one section in particular stood out:

Looking at our original content performance over the last few years, there have been so many impressive aspects...

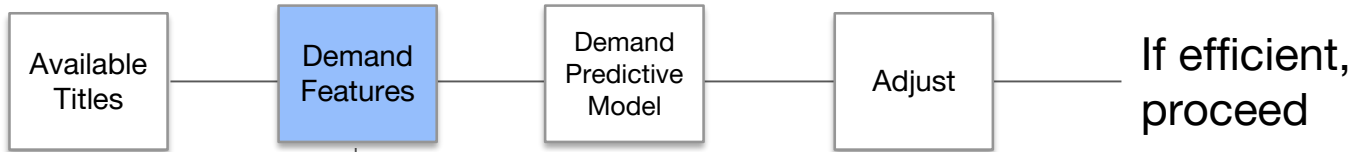
But there is one real shocker; **last year our original content overall was some of our most efficient content.** Our originals cost us less money, relative to our viewing metrics, than most of our licensed content, much of which is well known and created by the top studios.

Using Machine Learning to Predict Value



NETFLIX

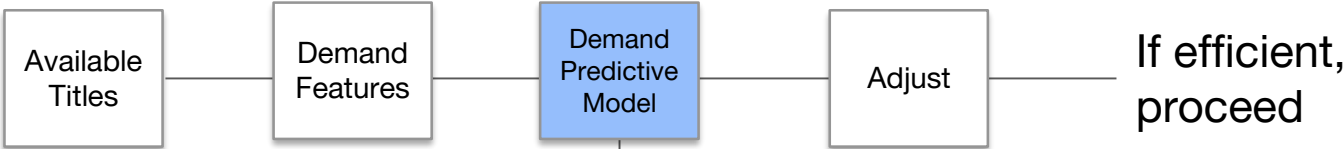
Using Machine Learning to Predict Value



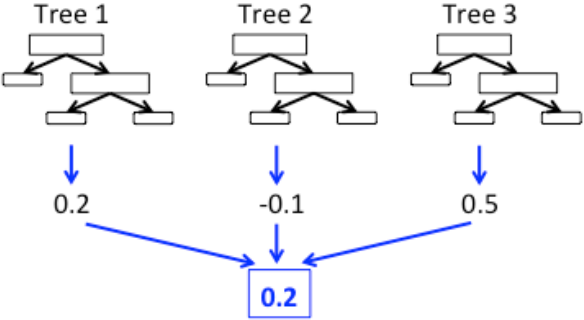
E.g.

- Past performance on Netflix (if previously licensed)
- Past performance of similar titles on Netflix
- Broadcast ratings
- Theatre ticket sales
- Talent involved
- Reviews
- Awards

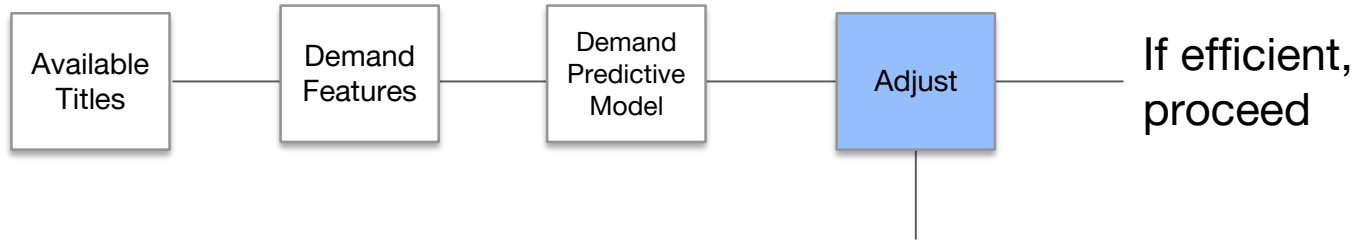
Using Machine Learning to Predict Value



Ensemble Model:
example for regression



Using Machine Learning to Predict Value

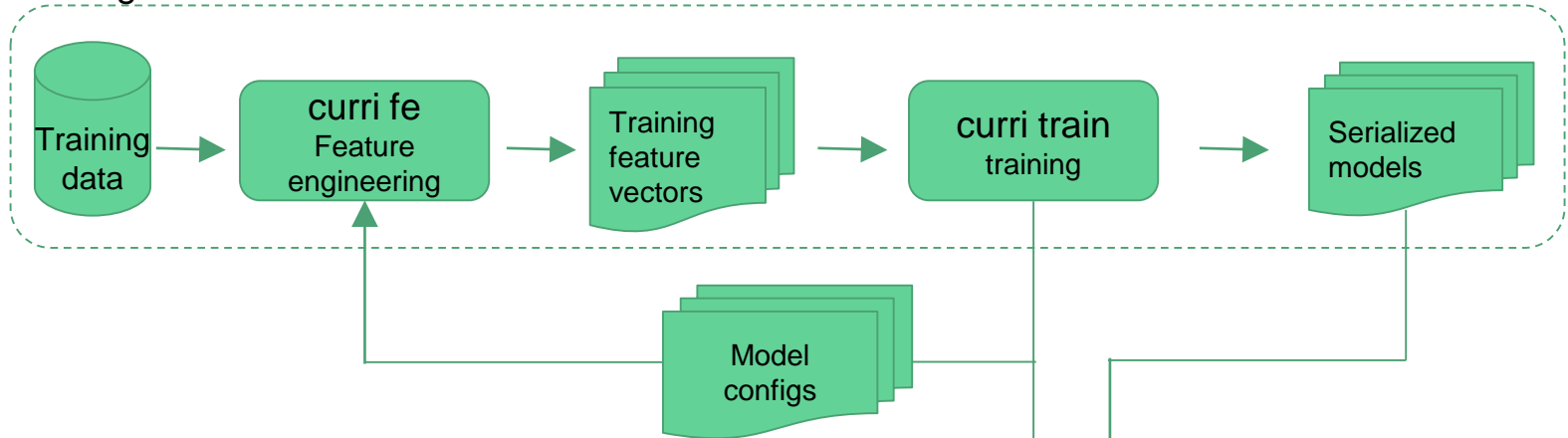


E.g.

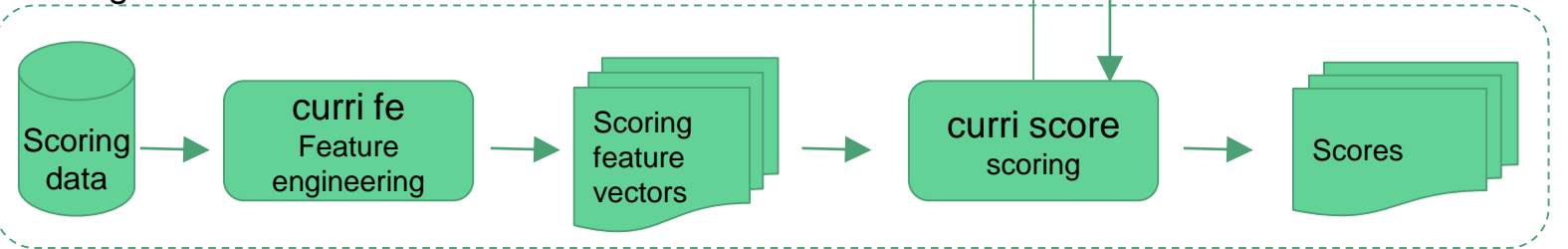
- Buyer judgements
- Deal term adjustments

Custom Machine Learning Framework

Training workflow



Scoring workflow



Using Machine Learning for Originals

Originals is a more difficult problem than licensing

- Less data - no box office or reviews
- Moving target - ideas and scripts can evolve
- Fungible - execution varies with talent and budget



NETFLIX

Finding Comparable Titles

'House of Cards' script =



X



Those members also watch:



NETFLIX

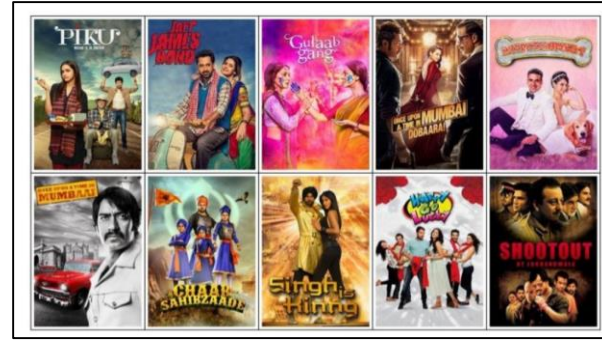
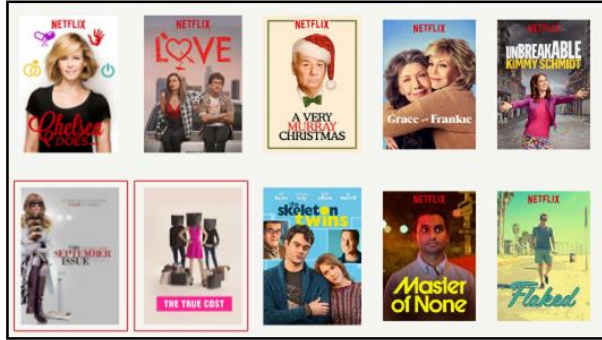
What's like **Twister** meets **Shark Week**?



NETFLIX

Programming to Tastes

(Something for Everyone)



NETFLIX

Programming to Tastes



NETFLIX

Data Science and Tech are in the DNA of Netflix
(and we'll keep looking for ways to leverage that DNA for content)



Thank You. Questions?

NETFLIX