Auto Content Moderation in C2C e-Commerce

Shunya Ueta, Suganprabu Nagarajan, Mizuki Sango
(Mercari, inc)

2020 USENIX Conference on Operational Machine Learning
JULY 27–AUGUST 7, 2020

Contents

mercari

Content Moderation

Identify potentially unsafe or inappropriate content in service

- App Discovery with Google Play, Part 3: Machine Learning to Fight Spam and Abuse at Scale
- YouTube Community Guidelines enforcement
- AI advances to better detect hate speech by Facebook
- Advances in content understanding, self-supervision to protect people by Facebook
- Facebook Transparency Report
- A Safe and Secure Marketplace by **Mercari**
- etc.

What is Mercari?

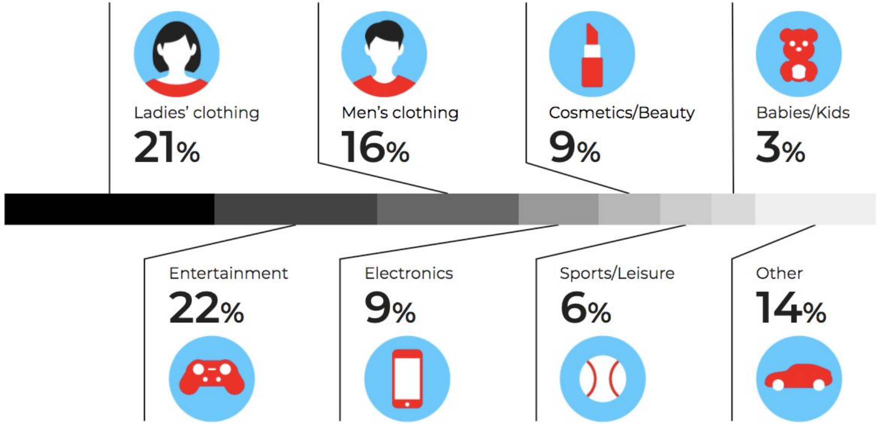
The Mercari app is a C2C marketplace where individuals can easily sell used items

Japan mercari

U.S. MERCARI

Monthly active users: **16+ Million**

Total number of items: **1.5+ Billion**



Source: Internal documents - percentages of FY2020,6 3Q GMV for Mercari's Japan business

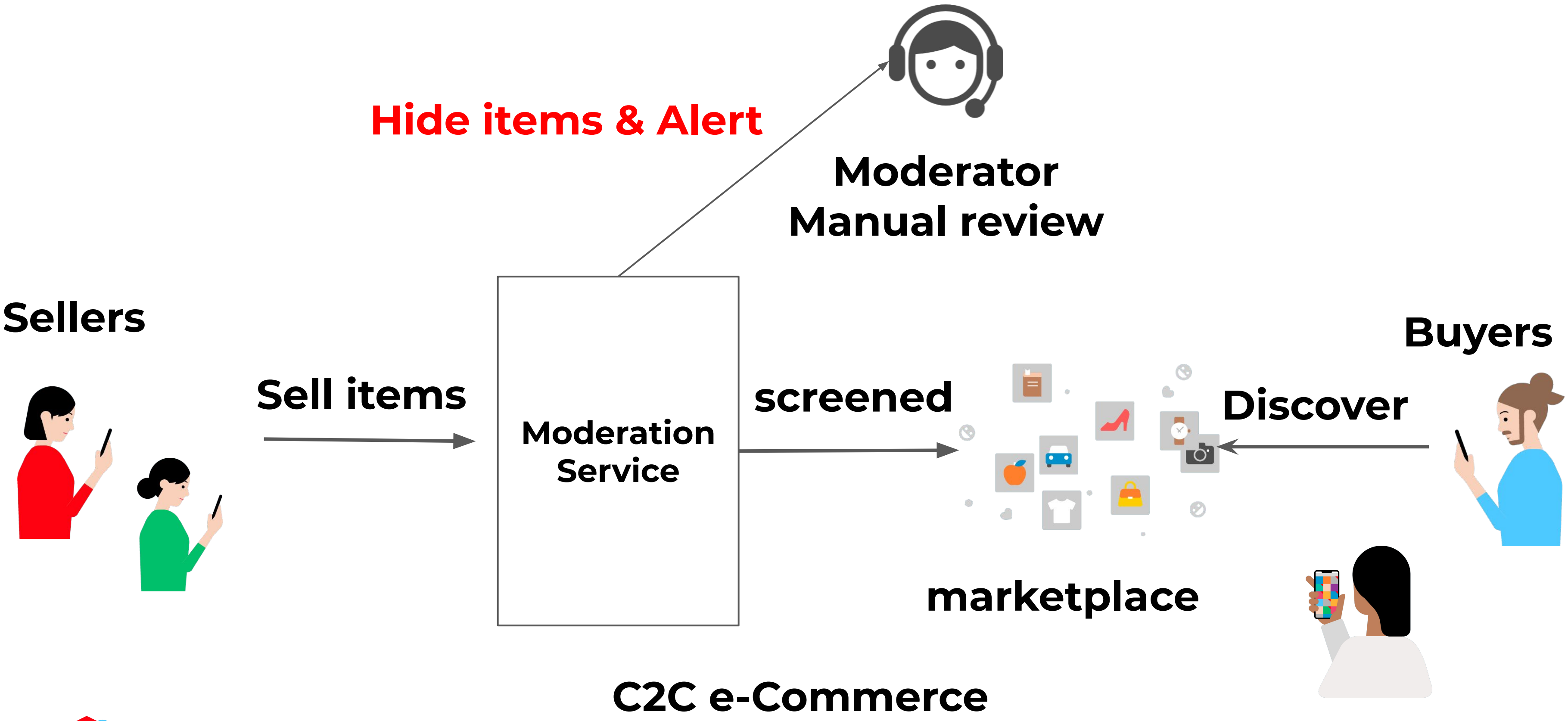
Why Content Moderation in C2C e-Commerce?

We want to decrease risk for customer and marketplace

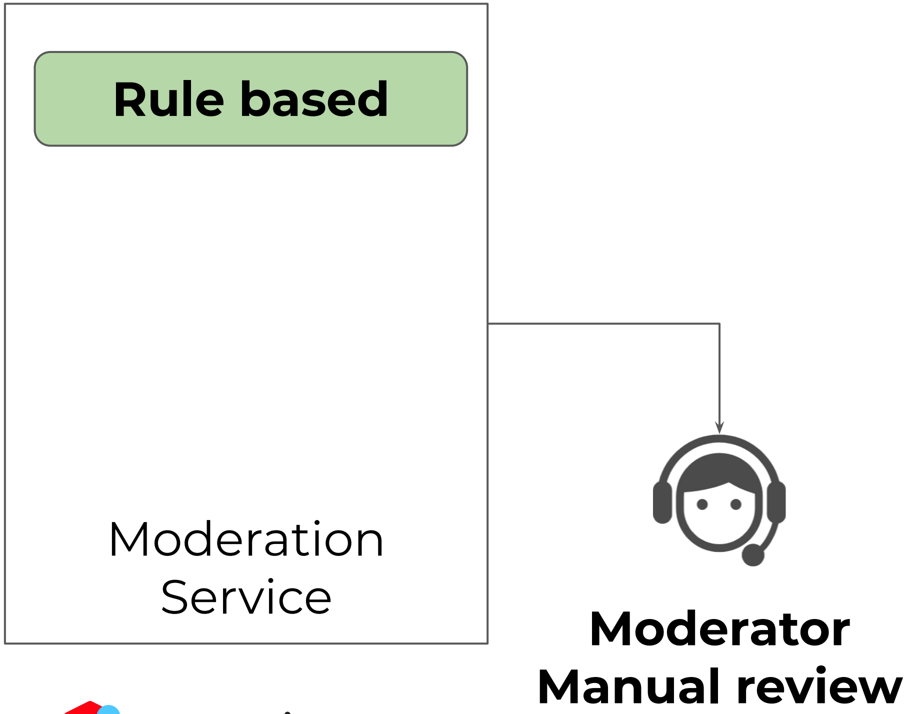
Sellers unintentionally violate policy. Buyers buy violated items without knowing

Policy case: counterfeits, weapons, etc.

Content Moderation system



Concept of Moderation Service: Rule based



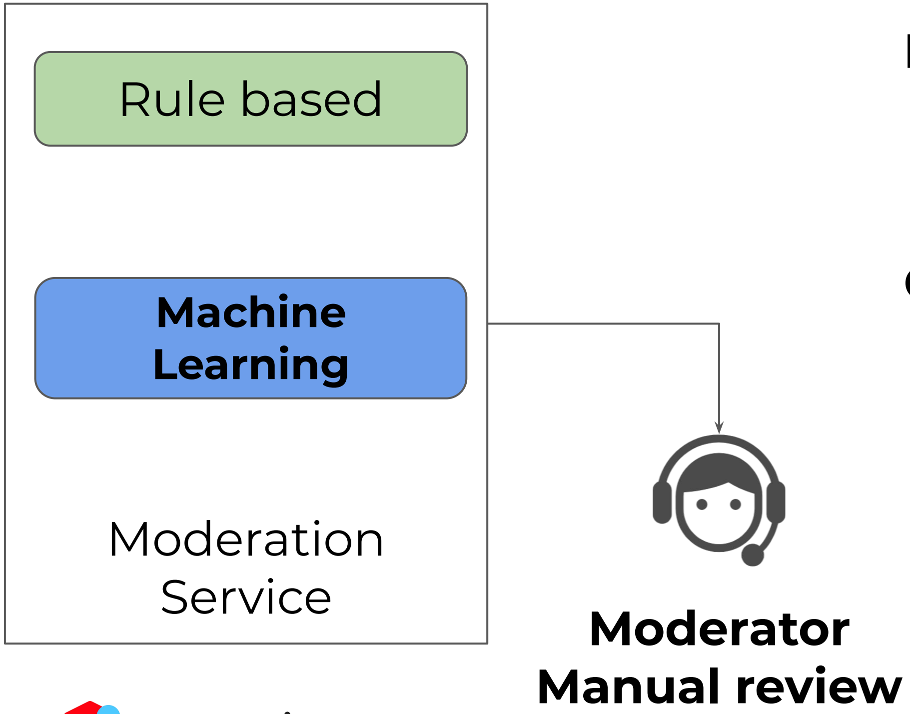
Pros

- Easy to develop and can be quickly released to production

Cons

- Hard to manage
- Difficult to cover the inconsistencies in spellings **e.g.** {NIKE, nike, ないき, ナイキ}

Concept of Moderation Service: ML



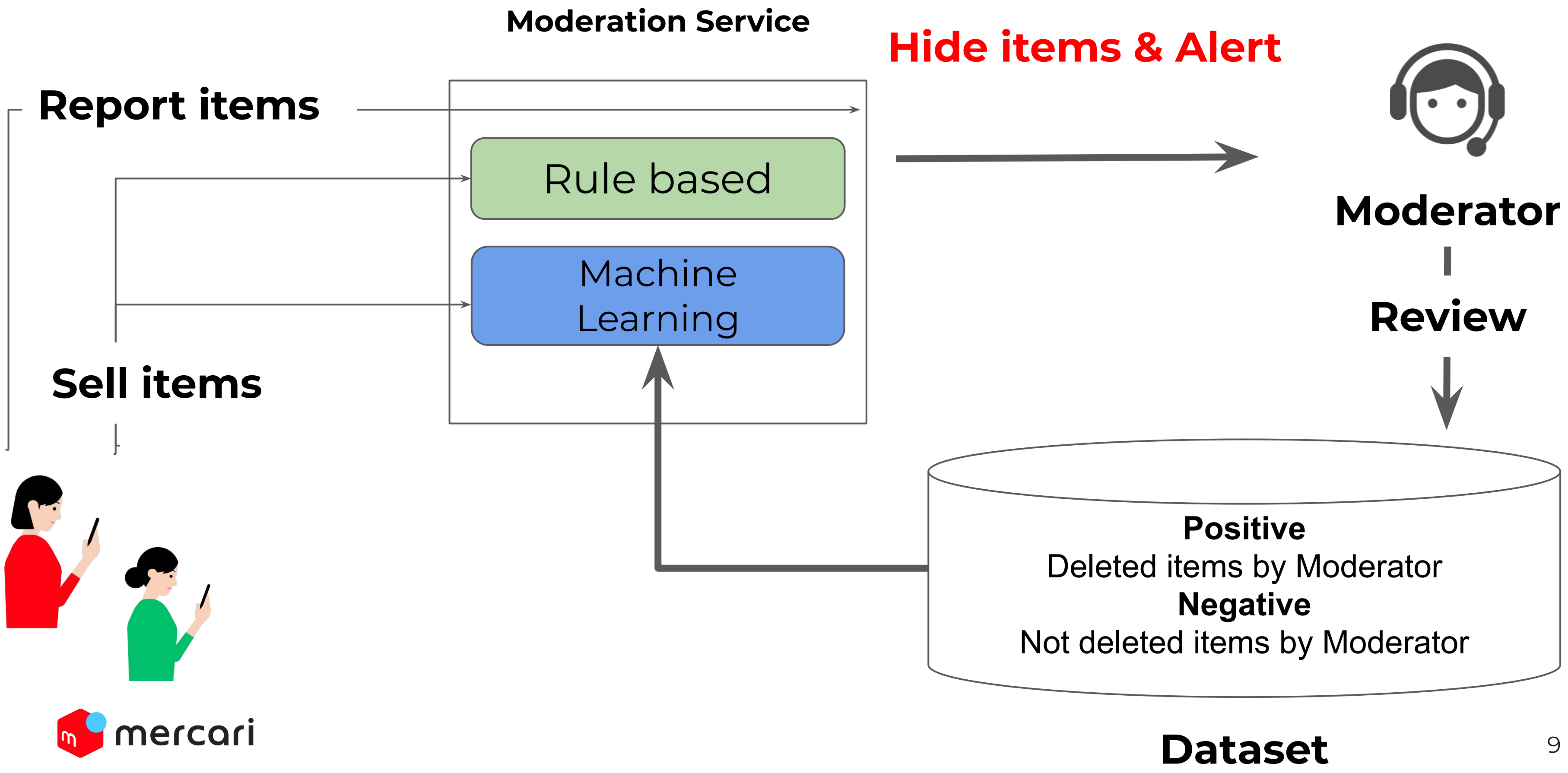
Pros

- Automatically learns the features of items deleted by moderators
- Adapts to spelling inconsistencies

Cons

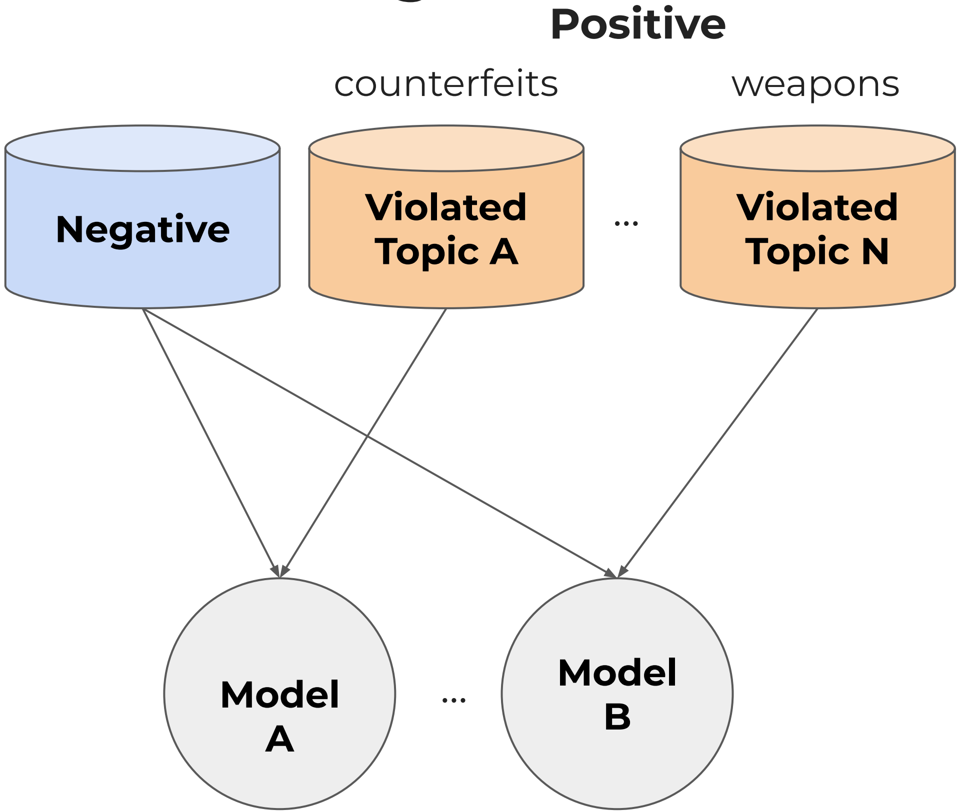
- Model update is hard
- Concept drift (a.k.a. training-serving skew)

How to create the data for ML



mercari

Task Design



- Data is highly imbalanced
- Each violated topic's total number of alerts is bounded by moderator team

All models trained as one-vs-all

- No side-effect when deploying a trained model to other class
- Hard to improve performance for each topic in a multi-class model

Multimodality of content

Items have multimodal data

- Image
- Text
- Category
- Brand
- Price, etc.

We use multimodal model to improve model performance.
See our article:
https://tech.mercari.com/entry/2019/09/12/130000

Case of items

Model selection based on dataset size

- Gradient Boosted Decision Trees (GBDT)

  → Efficient for training and inference when **training data size is not large**

  *****Image feature is not used in GBDT

- Gated Multimodal Unit (GMU)

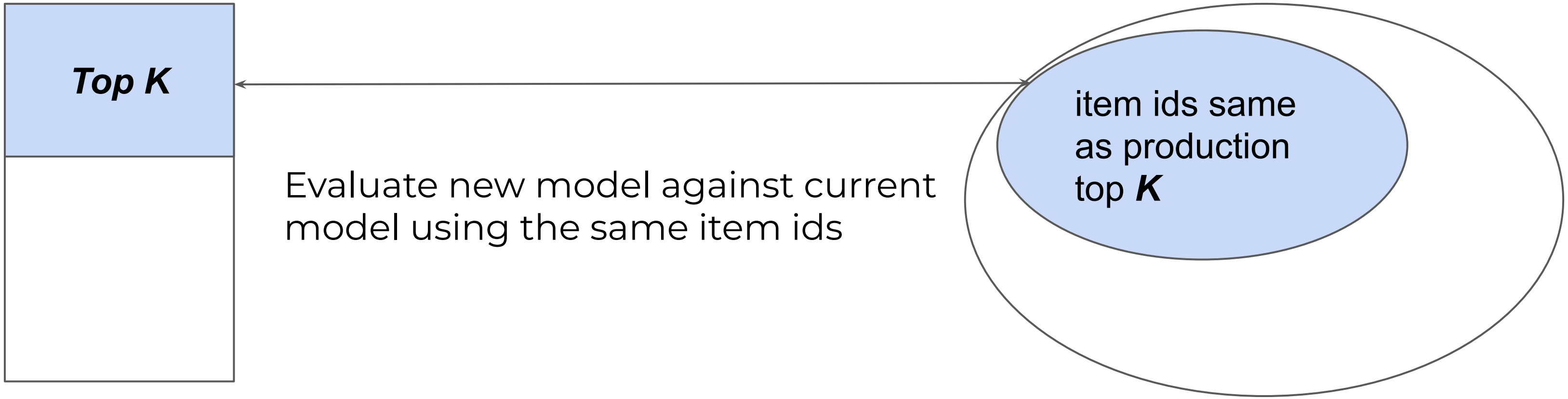
  → Potentially **most accurate** using **multimodal data**

Offline evaluation

e.g.

2020-07-13

2020-07-13



Current model's prediction result In production

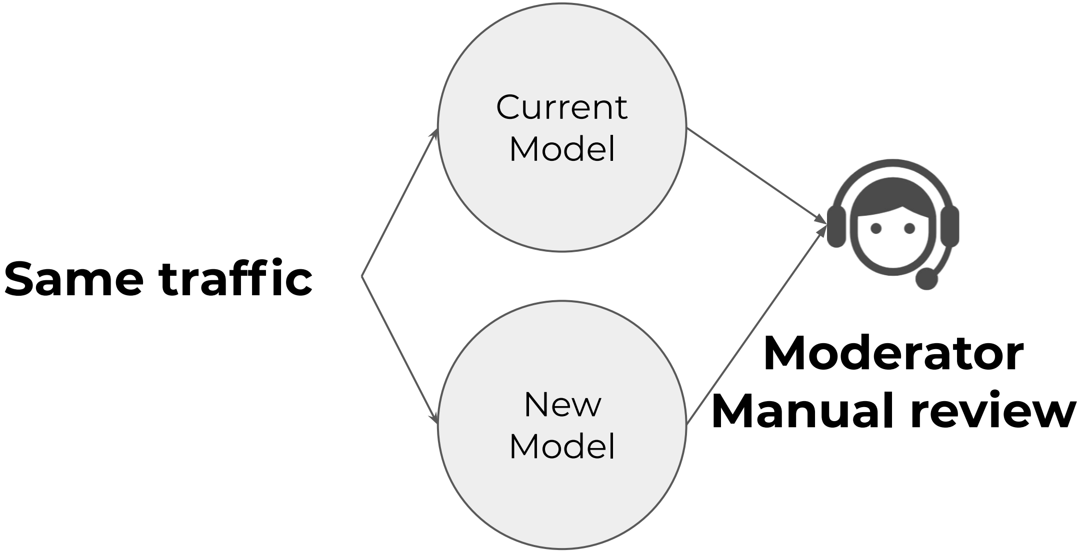
New model's prediction result In test dataset.

Metric is **Precision@*K***: ***K*** is the bound on the daily total number of alerts in each violated topic **decided by Moderators**

mercari

Online evaluation

Classic A/B testing can take several months. It was difficult to collect enough transactions for t-test.



Each model alert number: ***K/2***

Metrics: ***Precision@K/2***

After a certain time after a new model is released, we decide which model should be deprecated based on the above metrics.

→ Faster decision making leads to efficient operation

Offline/online evaluation result

Baseline model is **Logistic regression** that was already released in production

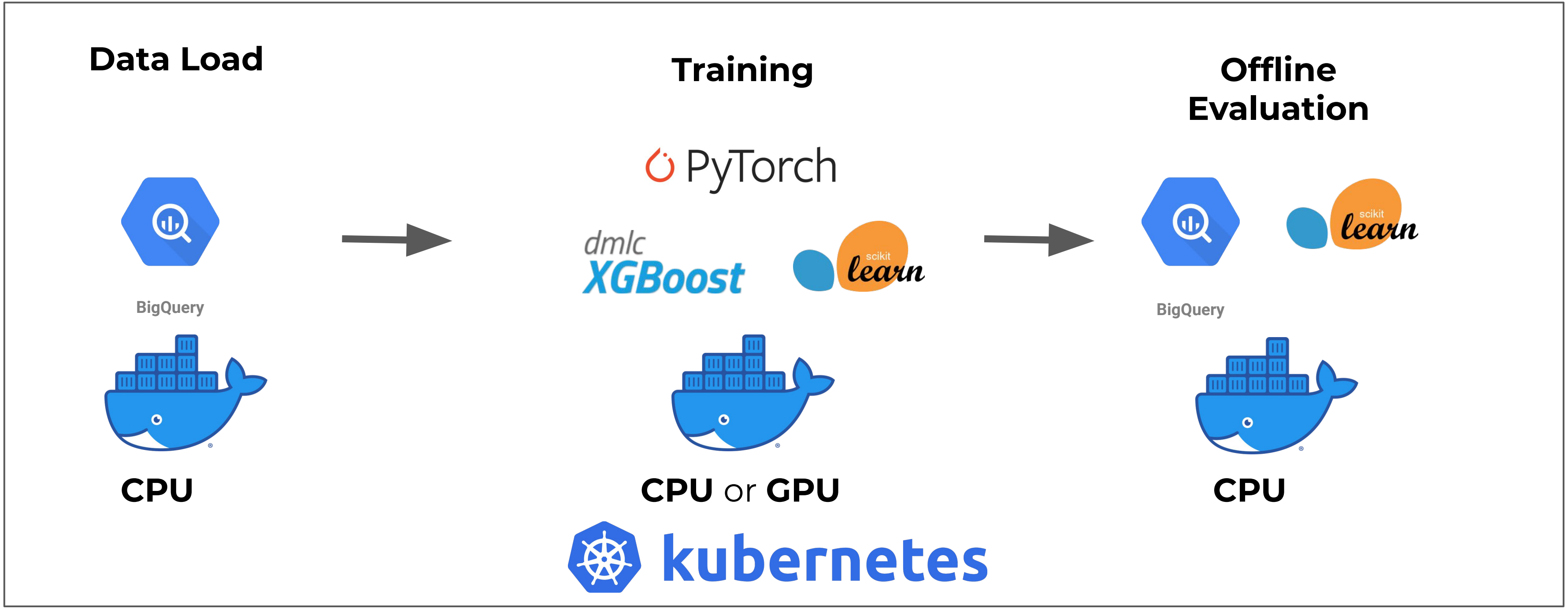
Algorithms	Offline	Online
GBDT	+18.2%	Not Released
GMU	**+21.2%**	**+23.2%**

Table shows the relative performance gain of
offline evaluation metric is ***precision@K*** ,
online evaluation metric is ***precision@K/2***
on one violated topic

mercari

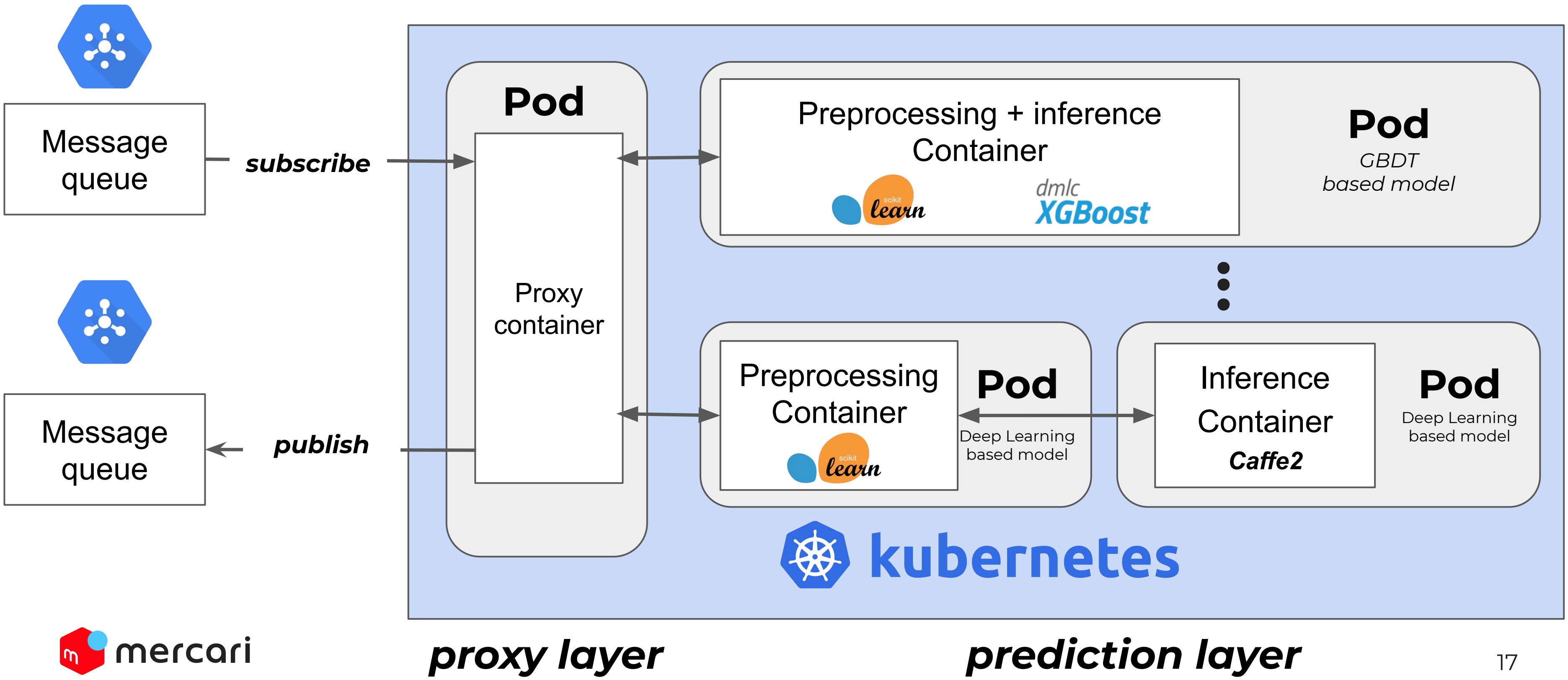
Container based Training Pipeline

Write manifest files containing requirements like CPU, GPU and Storage



Serving system architecture

We manage over 15 Machine Learning models in production



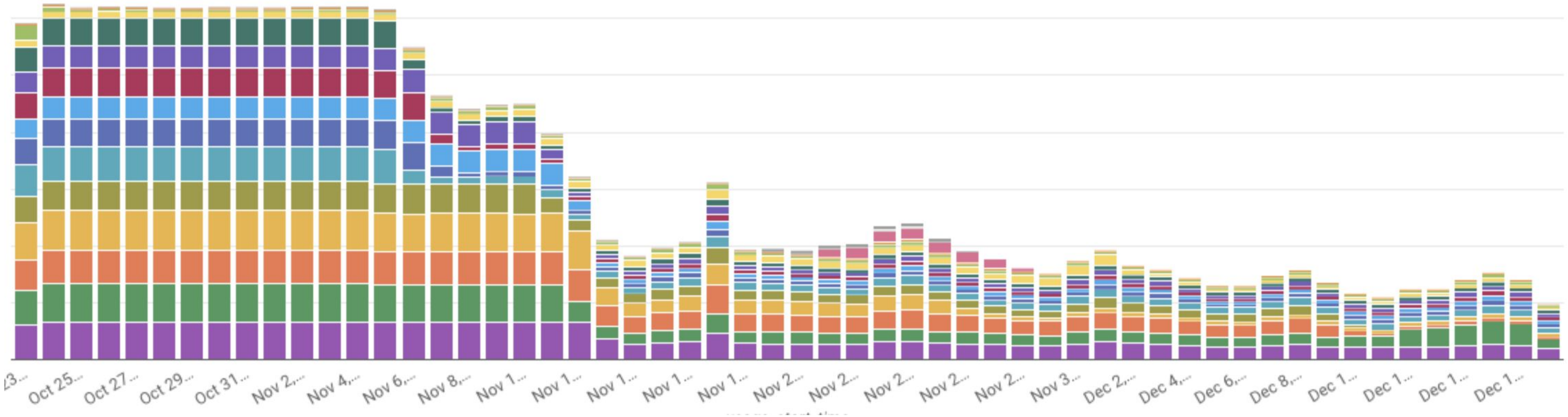
mercari

Horizontal Pod Autoscaler by kubernetes

- Reliable system: Traffic changes with time,
  HPA can adopt to varying traffic
- Cheaper billing cost: Reduce to 1/6 by HPA

Each color is each machine learning model

Billing cost



day

Billing cost transition after applying HPA

mercari

Impact of Machine Learning system

Machine Learning system

has increased coverage by ***554%*** ↑ over rule based approach

e.g.

Hide & Alert

Rule based → Discovered 100 violating items

Machine Learning → **+Discovered 554 violating items**

Moderation Service

Moderator
Manual review

Question and Thanks collaborator

If you have a question to this talk

First author is Shunya UETA, please e-mail: hurutoriya@mercari.com

Acknowledgements

Co-Authors: Suganprabu Nagarajan, Mizuki Sango

Contributter:

- Abhishek Vilas Munagekar, Yusuke Shido, Vamshi Teja Racha, Sumit Verma and Keisuke Umezawa for their contribute to this system
- Dr. Antony for his feedback about the paper
- Yushi Kurita, Yuki Ito as Product Manager, All Trust and Safety project member and all Customer Service as Moderator to success this project.