

# Mercari Shops を支える Google のベクトル検索技術

# スピーカー紹介



野上 和加奈

Souzoh ソフトウェアエンジニア

2019年にメルカリJPに新卒入社。MLOps エンジニアとして、写真検索システムの運用、配送サイズ推定システムの開発を担当。2021年4月からソウゾウに移動し、機械学習チームをリードしている。



佐藤 一憲

Google デベロッパーアドボケイト

Google Cloud のデベロッパー アドボケイトとして、機械学習や AI 系プロダクトの開発者支援を担当。Google Cloud Next、Google I/O、NVIDIA GTC 等の主要イベントでスピーカーを務め、Google Cloud 公式ブログに多数の記事を寄稿。

01

# ベクトル検索技術とは

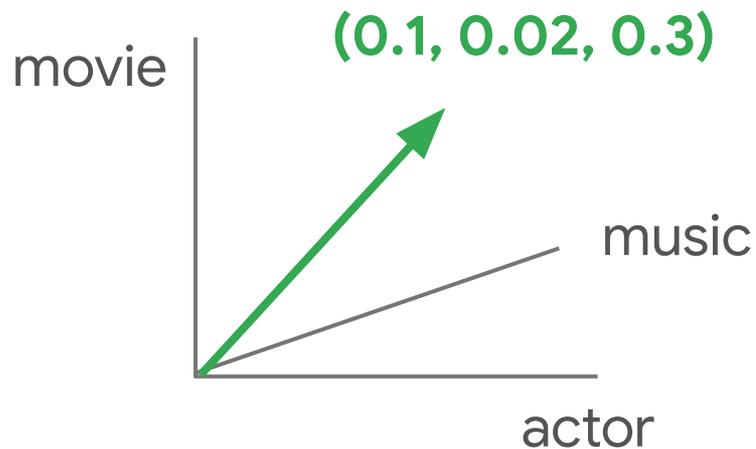
インターネット上の膨大な情報の中から、  
Google はいかにして価値ある情報を見つけているか



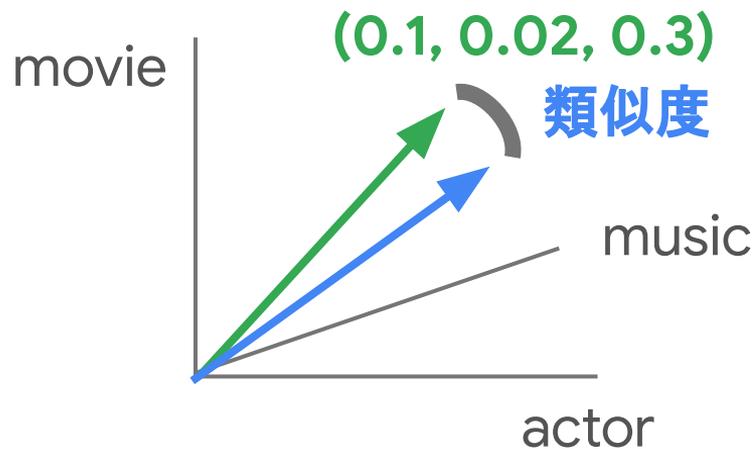
一般的な**キーワード検索**では、文字列やタグ、ラベル等でコンテンツを検索します

```
SELECT id
FROM contents
WHERE tag IN
    ( 'movie', 'music' ... )
```

一方、ベクトル検索では、  
ベクトルの類似度でコンテンツを探します



一方、ベクトル検索では、  
ベクトルの類似度でコンテンツを探します





# ベクトル検索の応用範囲:

ベクトルを定義できるあらゆる用途で利用可能



文書や画像の  
内容で探す



似ている製品を  
探す



似ているユーザーを  
探す



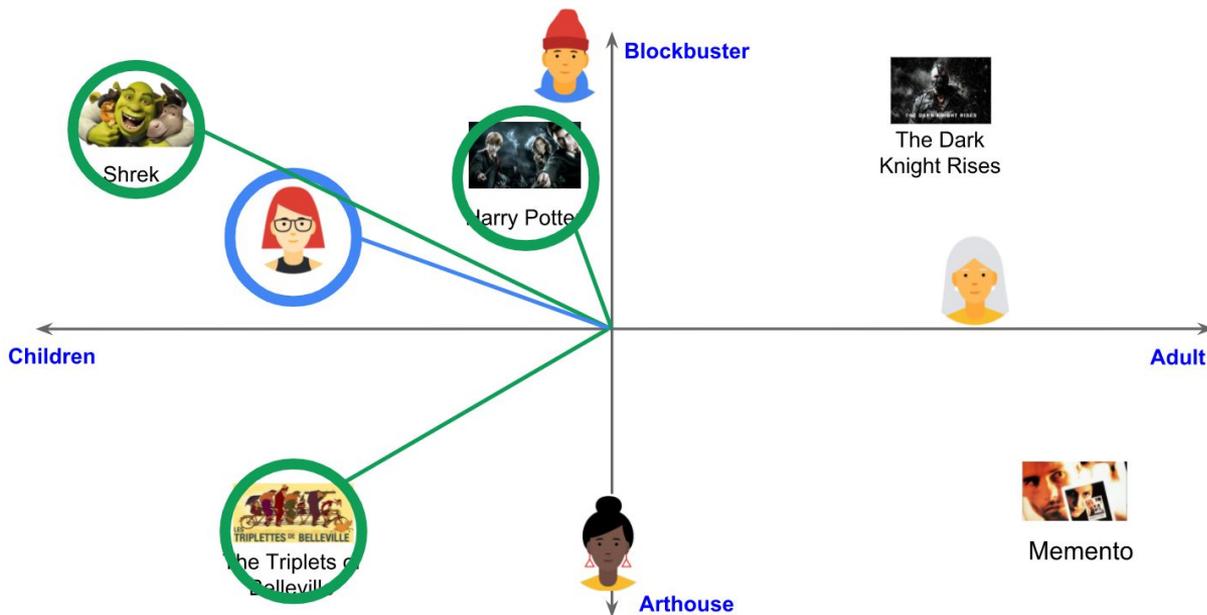
おすすめの音楽  
や動画を探す



故障しそうなIoT デバイスを  
探す

# Embeddings: ML で作る「かしこい」ベクトル

ビジネスやユーザーの要件にフィットするベクトル空間を ML で作る

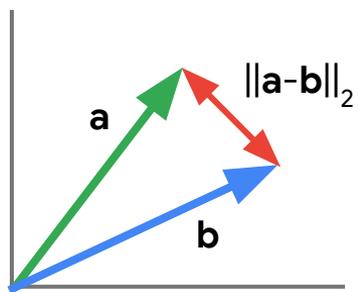


From: [Machine Learning Crash Course](#)

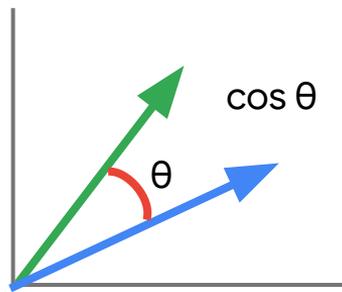
02

# Vertex AI Matching Engine によるベクトル検索

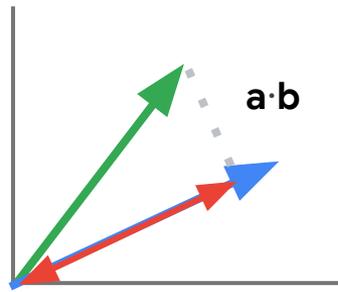
# ベクトル検索の難しさ: 類似度の比較が重い



L2 distance



cosine similarity

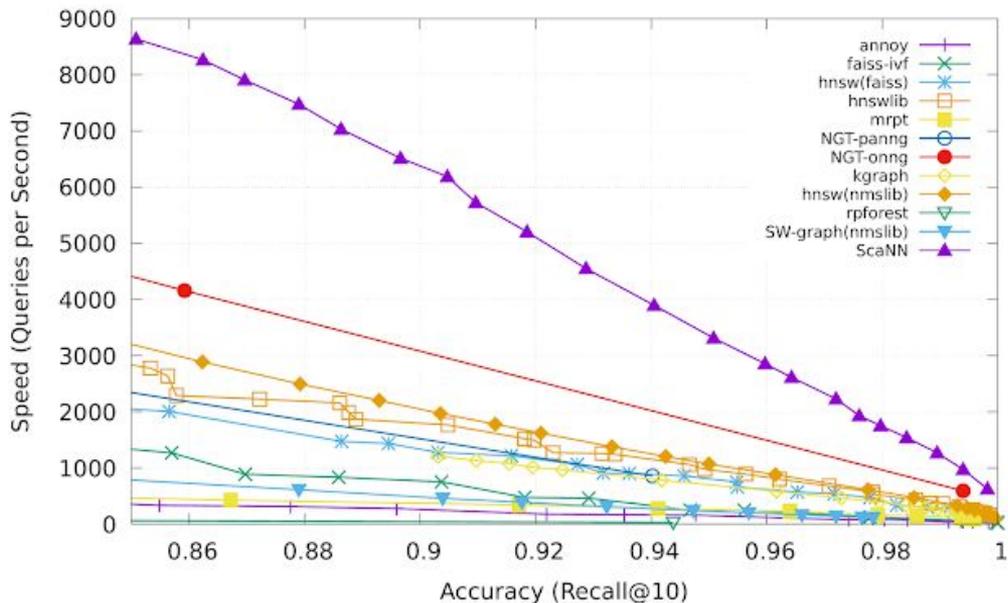
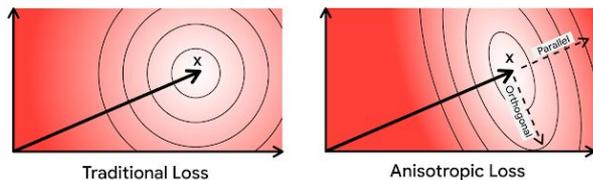


inner product

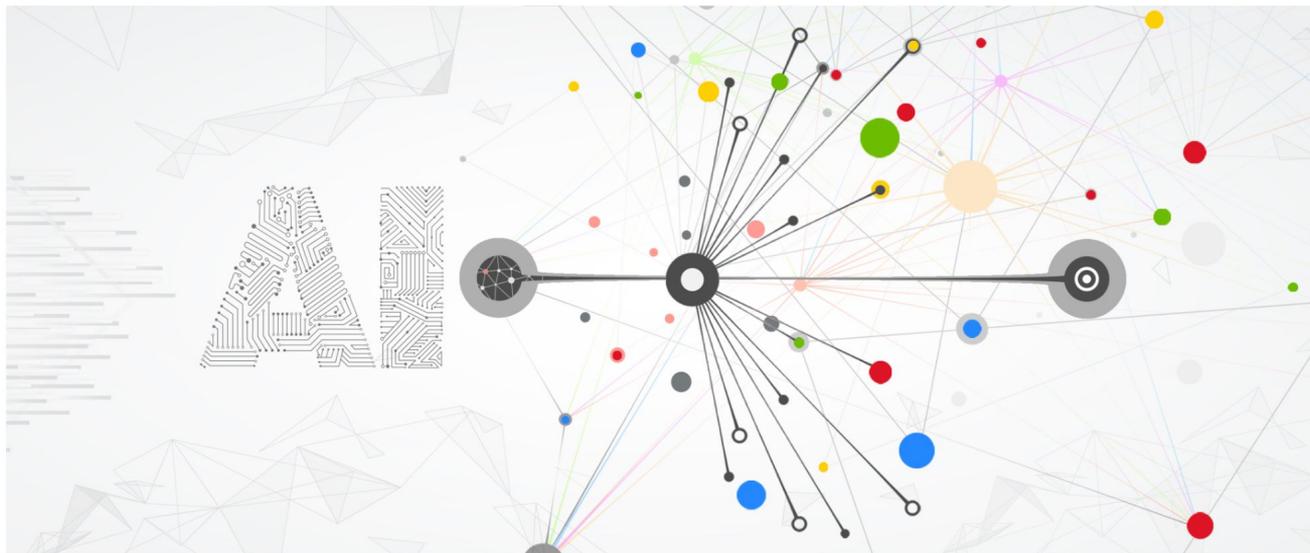
2 dimensions x  
1M items  
=  $O(1M \times 2)$

# ScaNN: Google 画像検索、YouTube、Play 等を支える、 Google Research 開発の高速近似近傍検索

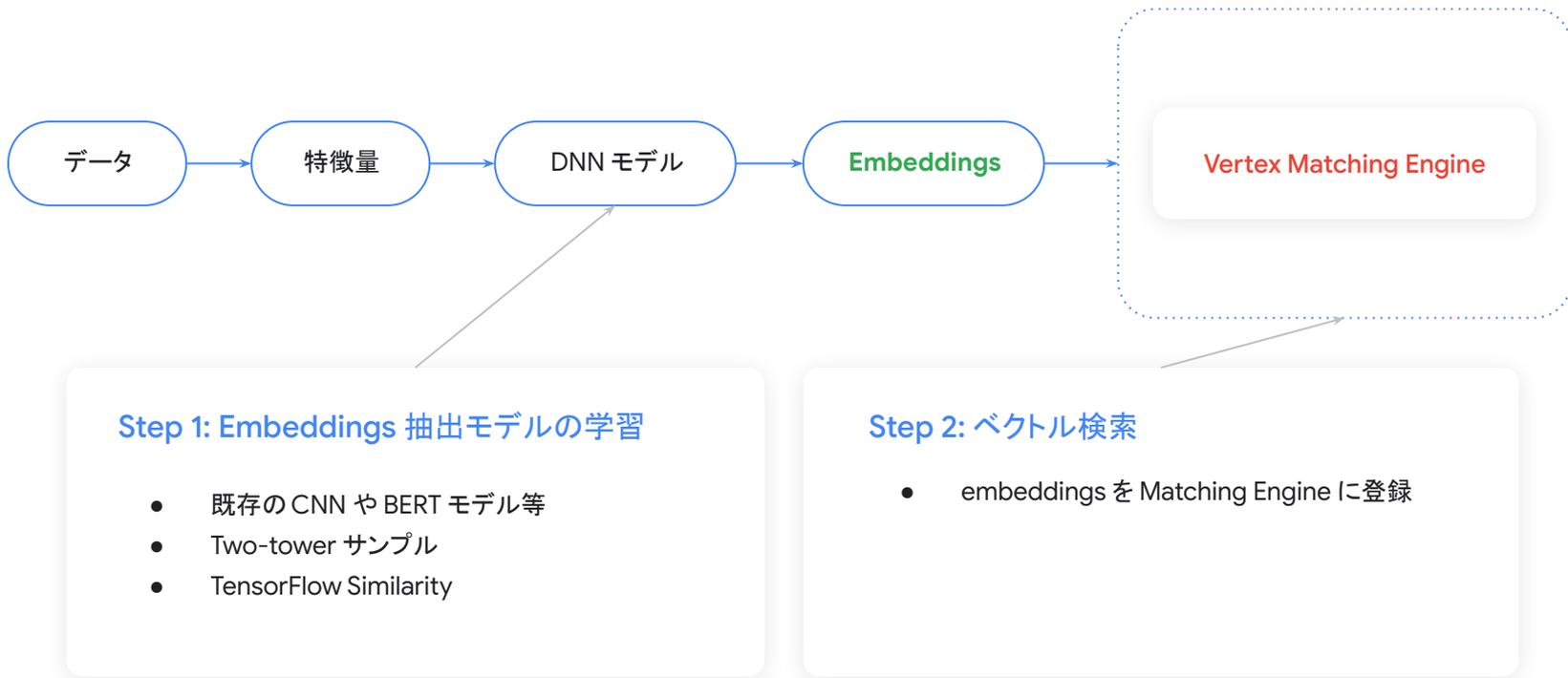
コードブック作成の新技术  
精度と速度のトレードオフを大幅  
に改善



# Vertex Matching Engine: 非常に高速かつ スケーラブルな最近傍探索



# Matching Engine の使い方



03

# メルカリShops が抱えていた課題

# メルカリShopsとは

## PRODUCT



「メルカリShops」は「かんたんで、売れる」をコンセプトに、スマホ1つでネットショップを開設し、ショップ運営者が事業者として商品を直接販売することができる、Eコマースプラットフォームです。



# メルカリShops が抱えていた課題



商品詳細を見たあと....

同じ Shop の商品

or

戻る

買い物体験として  
なめらかでない

# 既存手法の問題点

協調フィルタリングなどのリコメンド手法は大量の閲覧/購入履歴が必要

- サービススタートからまもない
- 日々大量の新しい商品が追加される

メルカリShops の状況には**不向き**

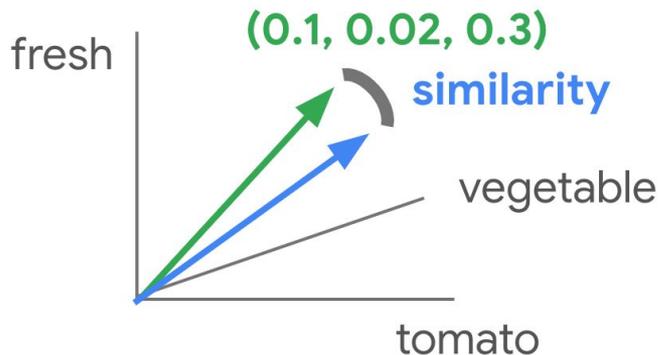
# 「類似性」でつくるマーケットプレイス

商品間の類似性を利用することで、  
行動履歴が蓄積されていないメルカリShopsでも  
新しい商品に対して即座にリコmend可能に

さらに:

ベクトル検索の汎用性の高さも決め手の 1つ

- 類似商品の 카테고리から 카테고리推定
- 類似商品の価格平均から価格提案

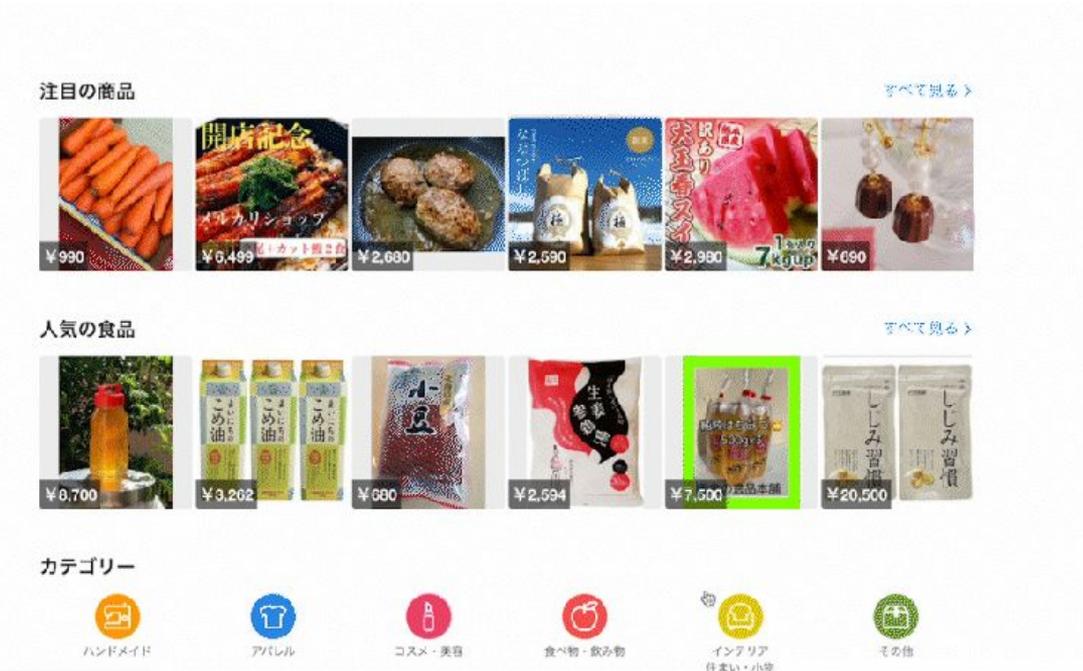


# メルカリShops のマーケットプレイスを ベクトル検索で実装

商品詳細から類似商品に遷移

なめらからショッピング体験

を実現



04

# メルカリShops における Matching Engine の導入

# Matching Engine による実装

## 特徴量の選定

- Shop ごとに画像に特徴があり異なる
- 見た目よりも概念が似ているものを探したい

→ テキストの特徴量を採用

## モデルの選定

- word2vec + TF-IDF で 満足のいく水準
- モデルの軽量さ

→ word2vec + TF-IDF を採用



"Brownie Coffee"

"Sweet and mild, tastes like brownies and nuts."

[0.12, 0.002, 0.98...]

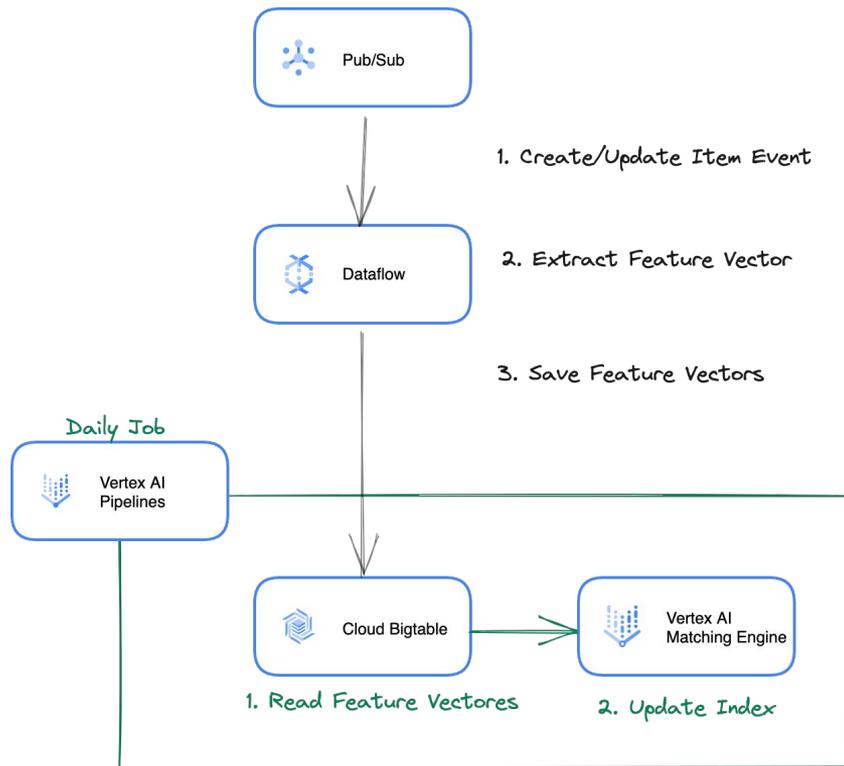
# 特徴抽出パイプライン

## 特徴抽出

1. 商品作成/編集イベント
2. Dataflow で word2vec + TF-IDF で特徴ベクトル生成
3. BigTable に保存

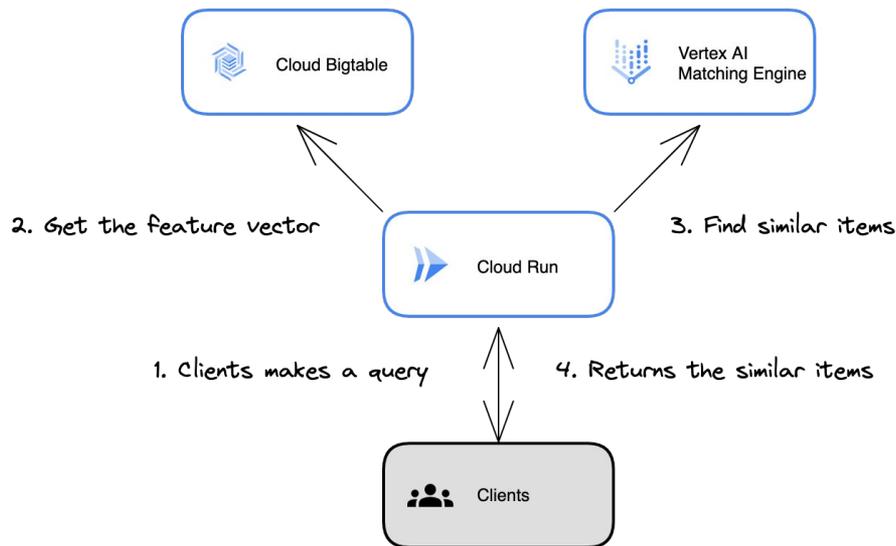
## インデックスの更新

1. VertexAI Pipelines を Daily で trigger
2. BigTable から特徴ベクトル取得
3. Matching Engine の API を呼ぶ



# ベクトル検索サービス

1. CloudRun に product\_id を含んだリクエストを送る
2. Product\_id を key に BigTable から特徴ベクトルを取得
3. 特徴ベクトルをリクエストに Matching Engine API を呼ぶ
4. 結果の類似 product\_id を返す



# メルカリShops の今後の展望

## 類似商品 API の利用の拡大

- 同様の仕組みを用いた Shop のリコメンデーション
- 価格提案

## 類似商品 API 自体の改善

- Matching Engine 検索時に商品をフィルタ
  - 現在は Matching Engine から取得後に売り切れ商品などをフィルタしている
- リアルタイムインデクシング
  - 現在は一日おきにしか検索対象に追加されない



05

まとめ

# Key Takeaways

Google の「虎の子」であるベクトル検索技術で  
ビジネス課題に向けたソリューションを構築



**Thank you.**