

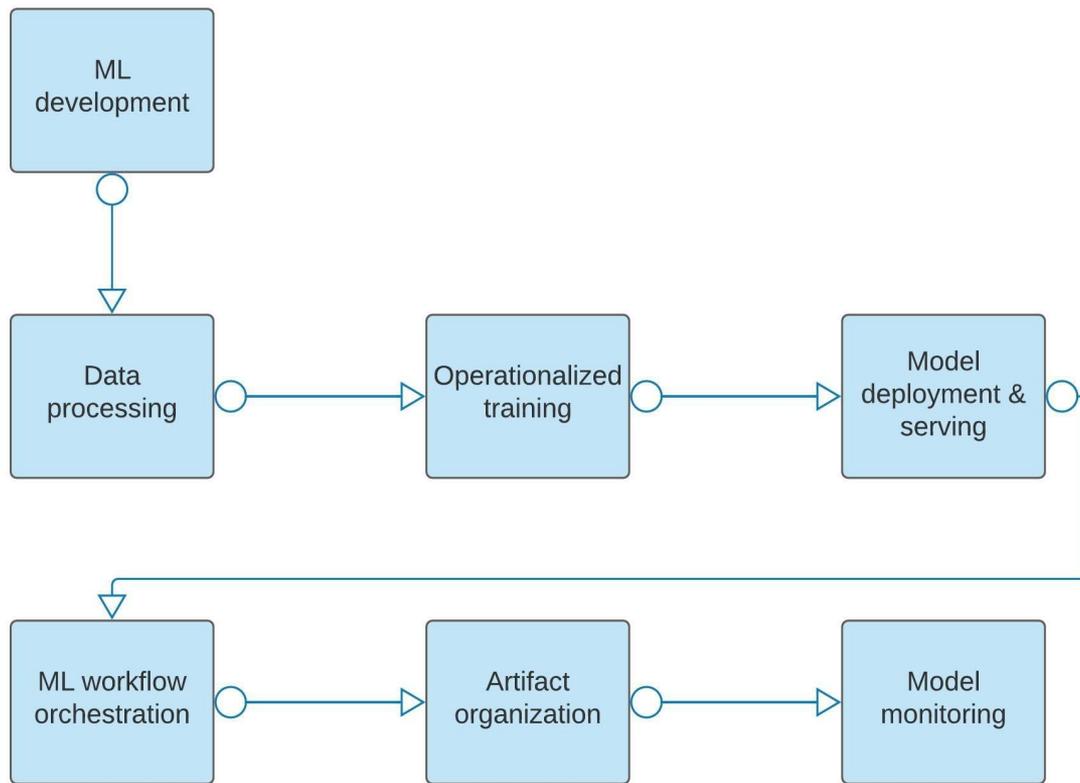
# Google Cloud を活用した 機械学習の実装ベストプラクティス

Google Cloud  
AI / ML スペシャリスト  
児玉 敏男



マネージド サービスを使用してカスタムトレーニング済みモデルを開発および運用するためのベストプラクティス

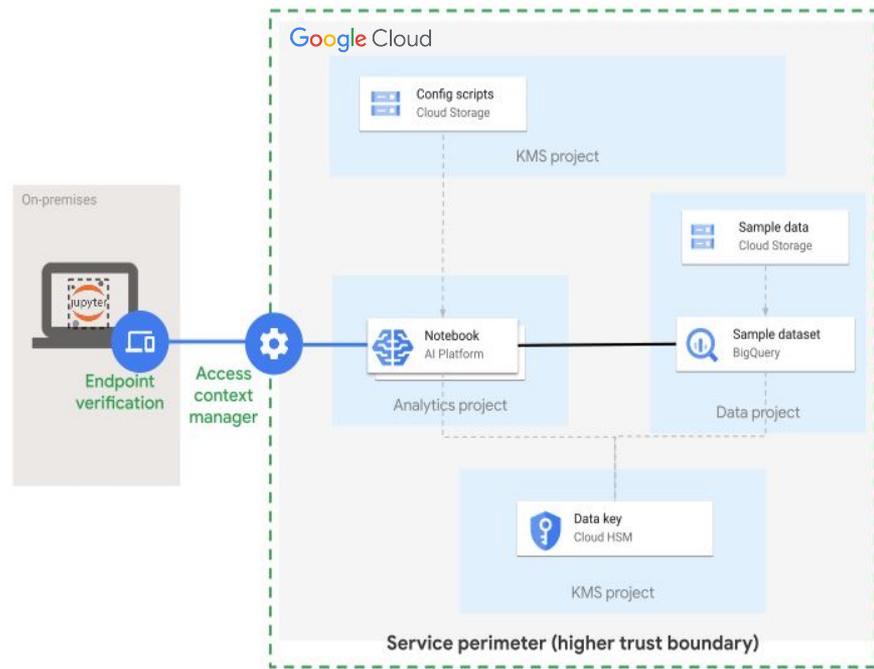
# ML ワークフローの各ステップ



# 機械學習環境

# 機械学習環境のセットアップ

- テストと開発に Notebooks を使用する
- チームメンバーごとに Notebooks インスタンスを作成する
- Notebooks での個人情報の保護
- 準備したデータとモデルを同じプロジェクトに保存する
- Python 用 Vertex SDK を使用する



# 機械学習の開発

# 表形式のデータを BigQuery に保存 画像、動画、音声、非構造化データを Cloud Storage に保存

## データストレージ

BigQuery に表形式データを保存する

利用するフレームワーク	利用するグーグルクラウドのツール
TensorFlow or Keras	<a href="#">tf.Dataset reader for BigQuery</a>
TensorFlow Extended (TFX)	<a href="#">BigQuery client</a>
Dataflow	<a href="#">BigQuery I/O Connector</a>
Any other framework	<a href="#">BigQuery Python client library</a>

非構造化データをクラウドストレージに保存します。

- 画像、ビデオ、オーディオなど。
- 大きなテナファイルにデータを保存します。
- TensorFlowの.tfrecord
- 他のフレームワークの場合は .avro
- 個々のファイルを大きなファイルに結合して、I/Oスループットを向上させます

# 非構造化データに Vertex Data Labeling を使用

The screenshot displays the Vertex AI Data Labeling interface. At the top, there is a navigation bar with a back arrow, a search field containing 've...', a dropdown menu with 've...', and a help icon. Below this, there are three tabs: 'IMPORT', 'BROWSE' (which is selected and underlined), and 'ANALYZE'. On the left side, there is a summary table:

All	5
Labeled	0
Unlabeled	5

Below the table, there are filter options: 'Filter Filter items' with a menu icon, and 'Select all' with an unchecked checkbox. Further down, there is another filter section: 'Filter Filter labels' with a menu icon and a plus sign. Underneath, there is a 'Videos' dropdown menu and a blue button labeled 'ADD NEW LABEL'. A red arrow points to this button. The main area of the console shows a grid of four video thumbnails depicting various golf-related activities: a golfer in a white shirt and cap swinging a club on a green, a golfer in a black shirt and shorts swinging a club on a grassy field, a golfer in a red shirt swinging a club on a golf course with a power line tower in the background, and a golfer in a blue shirt swinging a club on a green with a white ball nearby.

# Vertex Feature Storeを利用し特徴量のデータを再利用

The screenshot displays the Google Cloud Vertex Feature Store interface for a feature named 'promotion'. The interface includes a navigation sidebar, a search bar, and a main content area with tabs for 'FEATURES' and 'ENTITY TYPE INFO'. The 'FEATURES' tab is active, showing a list of features with their respective statistics and distribution charts.

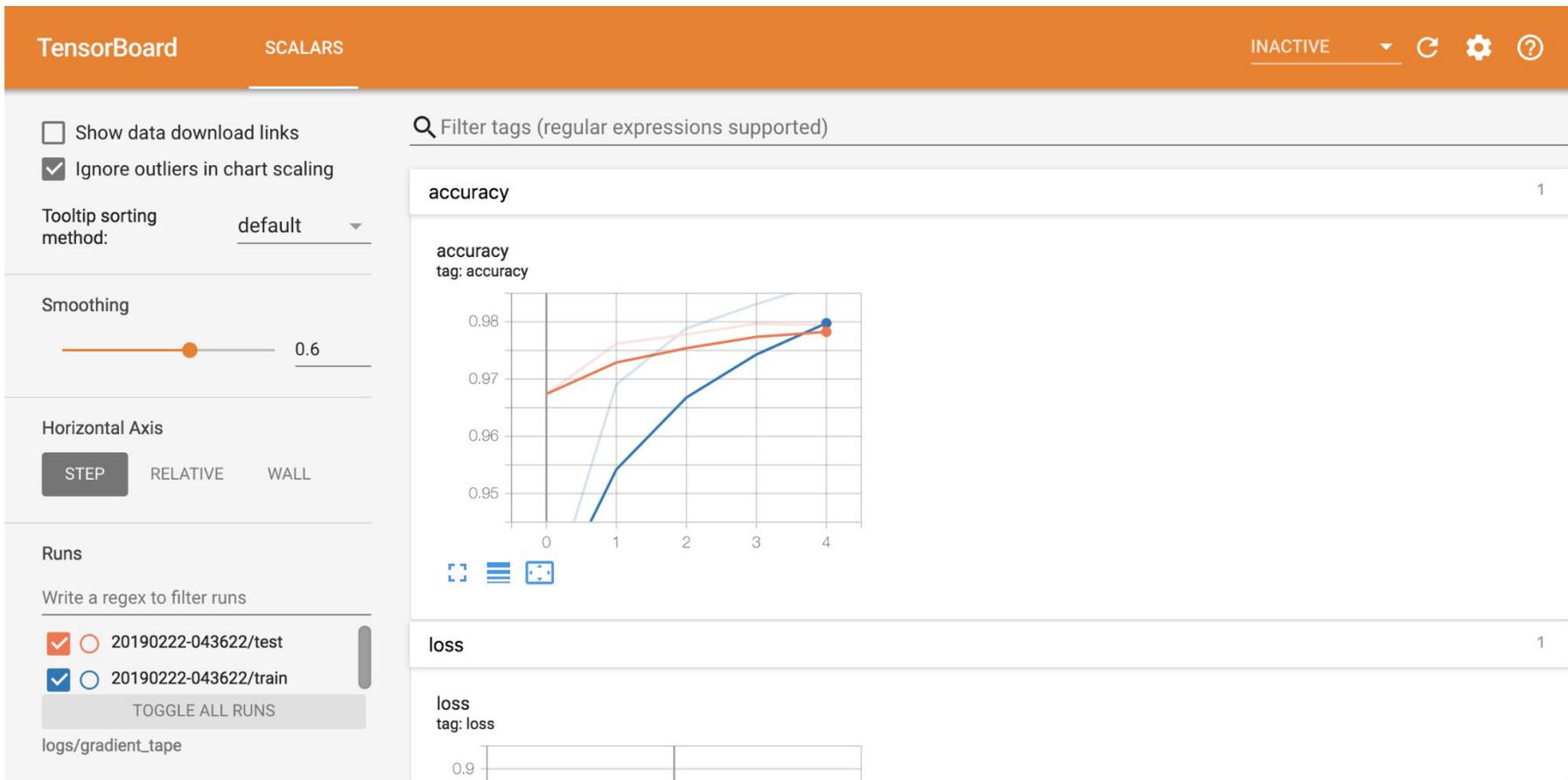
**Description**  
Torquem detraxit hosti et quidem exercitus quid ex ea voluptate ponit, quod maxime. Sed ut calere ignem, nivem esse vult, summumque malum et, quantum possit, a. At vero eos censes aut officis debitis aut fugiat aliquid, praeter voluptatem et.

**Summary**  
Total columns: 6  
Total rows: 4,987  
Size: 789 MB

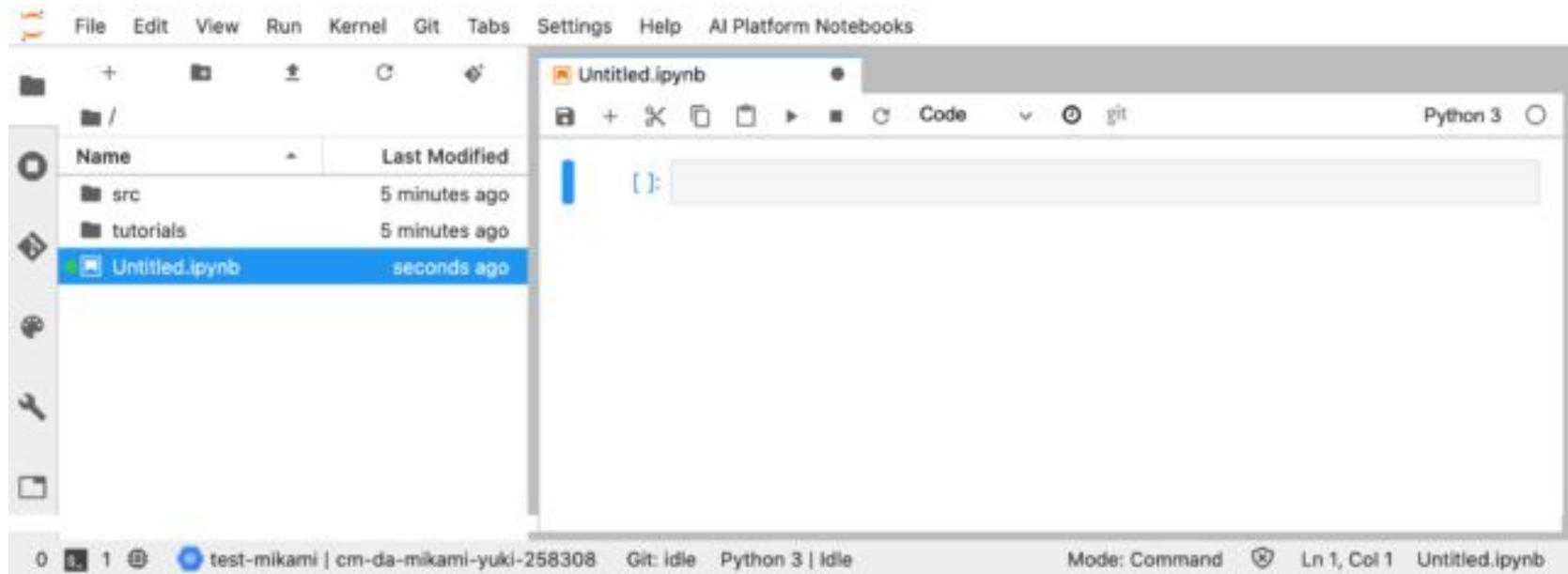
**Features**

Feature name	Type	Count	Missing	Mean	Std. dev.	Zeros	Values	Most common	Avg. string length	Unique	Distribution	Created
<input type="checkbox"/> promo_percentage	INT32	870	0	57	0	16	Min: 0 Median: 52 Max: 80	-	-	779		10:11 AM, Jun 27, 2024
<input type="checkbox"/> promo_zipcode	INT32	106	0	265	0	0	Min: 0 Median: 32 Max: 323	-	-	722		10:11 AM, Jun 27, 2024
<input type="checkbox"/> promo_discount-info	STRING	861	14	57	0	29	Min: 0 Median: 52 Max: 80	"FREE RIDES" Frequency 454	76	632		10:11 AM, Jun 27, 2024
<input type="checkbox"/> promo_offer-description	STRING											10:11 AM, Jun 27, 2024
<input type="checkbox"/> promo_percentage	INT32	342	0	57	0	0	Min: 0 Median: 52 Max: 80	-	-	999		10:11 AM, Jun 27, 2024
<input type="checkbox"/> promo_zipcode	INT32	735	7	265	0	87	Min: 0 Median: 32 Max: 323	"Use this for free rides" Frequency 122	6	687		10:11 AM, Jun 27, 2024

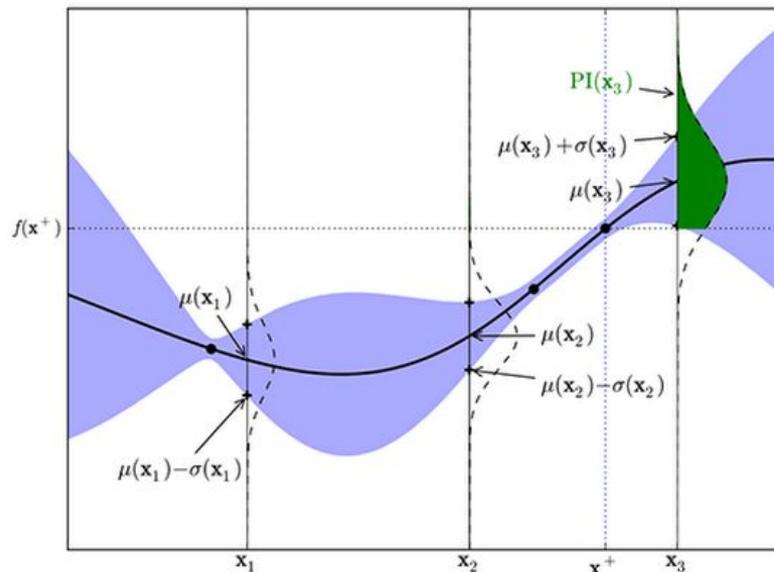
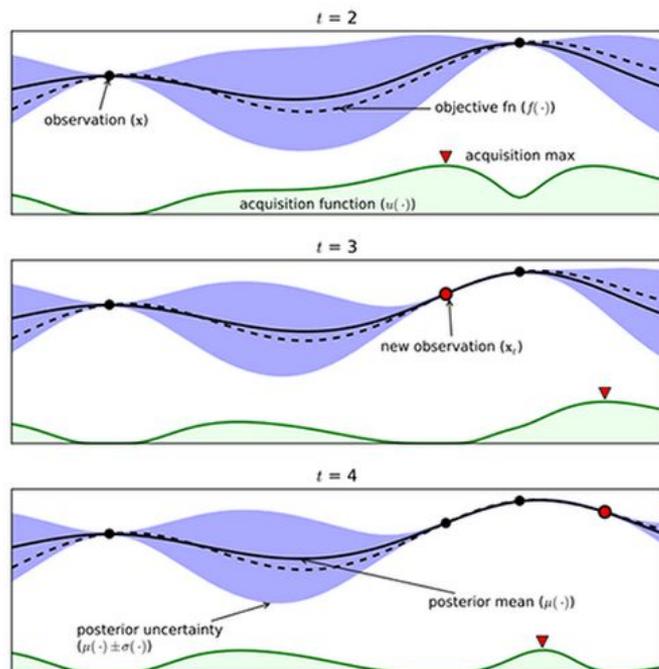
# Vertex TensorBoard を使用してテストを可視化



# Notebooks 内で小さなデータセット用のモデルをトレーニング



# ハイパーパラメータ調整の機能を使用してモデルの予測精度を最大化



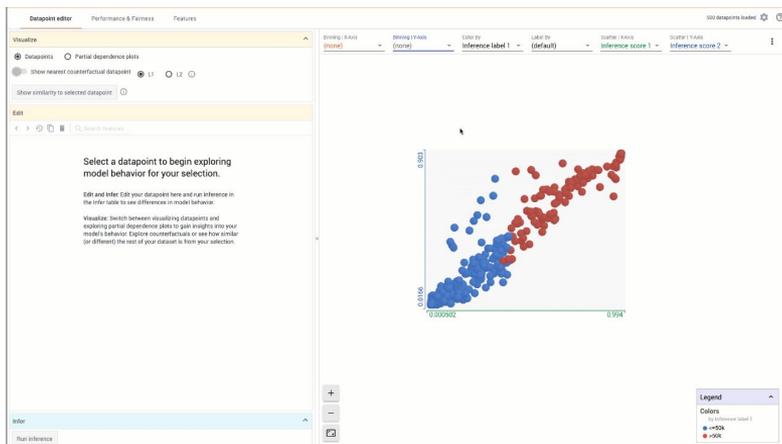
ガウス過程の改善の確率

## 目的関数のガウス過程による近似

'A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning' (<https://arxiv.org/pdf/1012.2599.pdf>)より引用

# Notebooks を使用してモデルを評価、理解

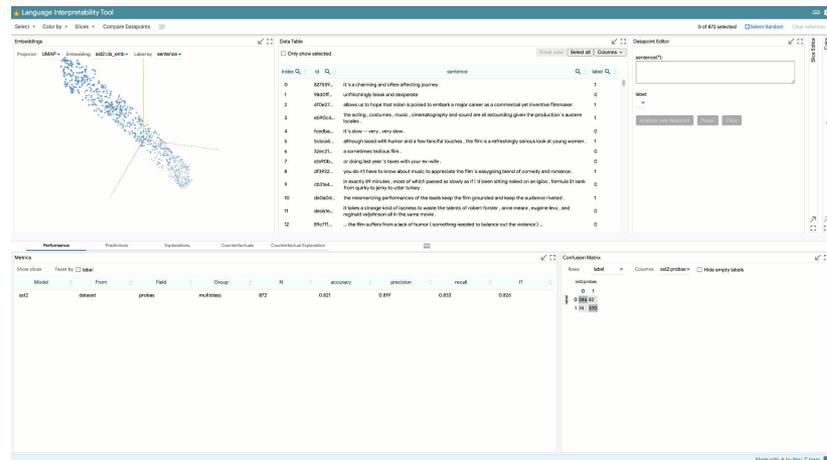
Tabular



## What-If Tool (WIT)

最小限のコーディングで、トレーニングされた機械学習モデルを視覚的に表現します

Text

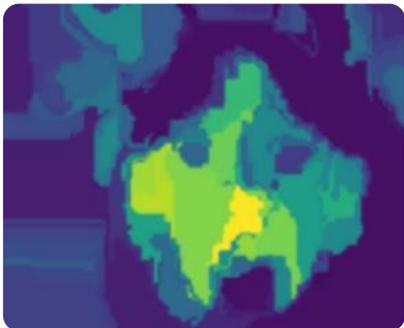


## Language Interpretability Tool (LIT)

NLP モデルの視覚化と理解のためのオープンソースプラットフォームです

# 特徴アトリビューションを使用してモデル予測の分析情報を取得

## Images

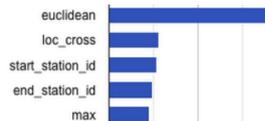


### Explainable AI の説明

モデルの分類に最も貢献した画像のピクセルまたは領域はどれですか？

## Tabular

Name	Feature value	Attribution value
distance	1395.51	-2.44478
start_hr	18	-1.29039
max_temp	20.7239	0.690506
temp	16.168	0.12629
dew_point	7.83396	0.0110318
prcp	0.03	-0.00134132



各特徴列は、単一の予測またはモデル全体にどの程度貢献しましたか？

## Text

The cake tastes  
delicious!

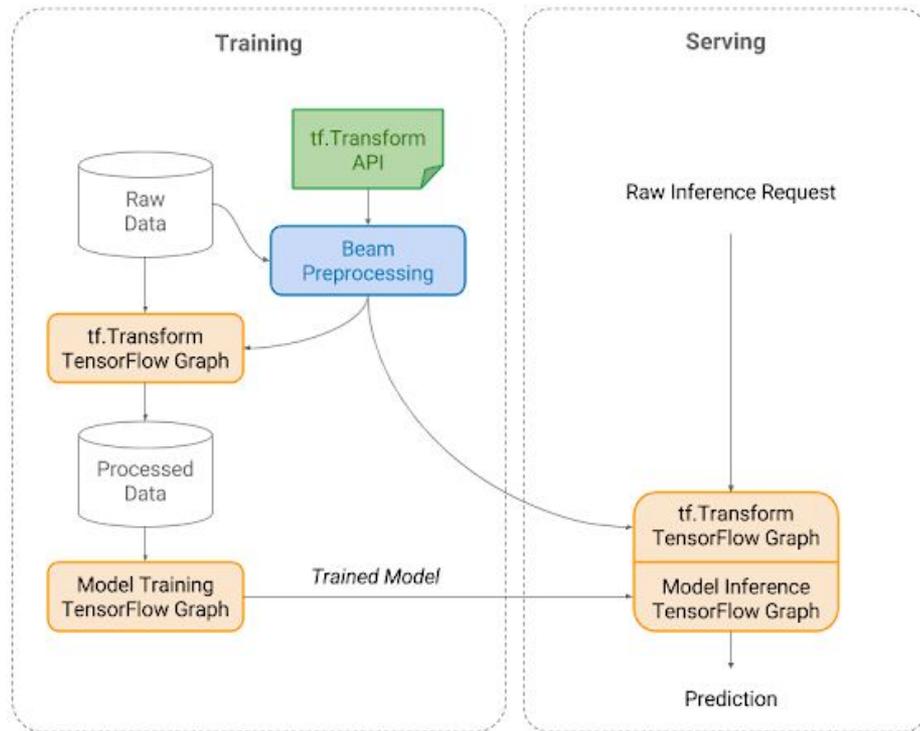
Sentiment score: 0.9

各単語またはトークンは、テキスト分類にどの程度貢献しましたか？

# データ処理

## データ処理のアプローチ

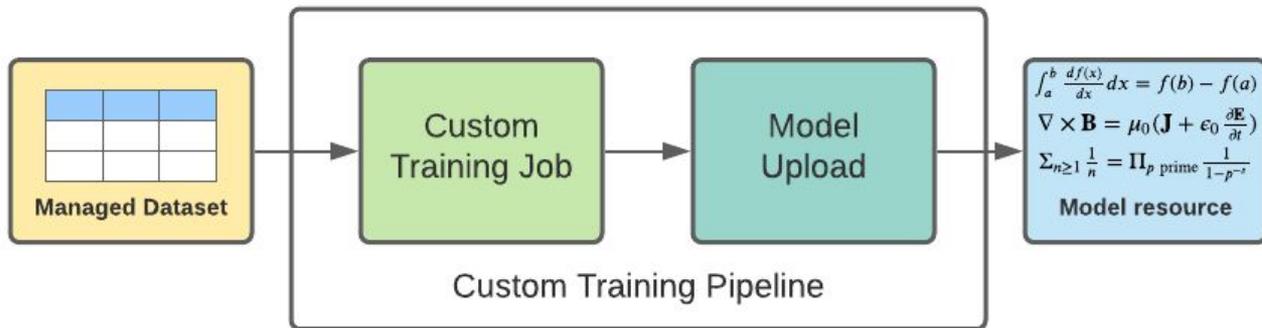
- TensorFlow エコシステムを活用する場合は、TensorFlow Extended を使用する
- BigQuery を使用して表形式のデータを処理する
- データフローを使用して非構造化データを処理する
- マネージドデータセットを使用してデータをモデルにリンクする



運用化されたトレーニング

## 運用時のベストプラクティス

- トレーニングパイプラインを使用してジョブ実行を運用化する
- トレーニングのチェックポイントを使用して、テストの現在の状態を保存する
- Cloud Storage でサービスを提供する本番環境のアーティファクトを準備する
- 新しい特徴量の値を定期的に計算する



# モデルのデプロイとサービング

## 予測入力の計画

大量のデータを予測する必要がありますか？

デシジョン マトリックス

No

Yes

Yes

Vertex Feature Store  
Online Serving API

Vertex Feature Store  
Batch Serving API

No

Operational Data Store  
(e.g. Cloud SQL)

Data Warehouse  
(e.g. BigQuery)

チーム内およびチーム間で特徴量を再利用すると、MLタスクの課題が改善しますか？

## マシンタイプを選択し、自動スケーリングを有効

```
gcloud beta ai endpoints deploy model $MY_ENDPOINT --project=$MY_PROJECT \  
  --region=us-central1 --model=$MY_MODEL --display-name=my_deployed_model \  
  --machine-type=n1-standard-4 \  
  --accelerator=type=nvidia-tesla-k80,count=1 \  
  --min-replica-count=2 \  
  --max-replica-count=4 \  
  --autoscaling-metric-specs=cpu-usage=70
```

← マシンタイプの選定

← アクセラレータのタイプと数の選定

← スケーリングの最大最小ノード数のセット

← 必要に応じて60%の使用率のオーバーライドします

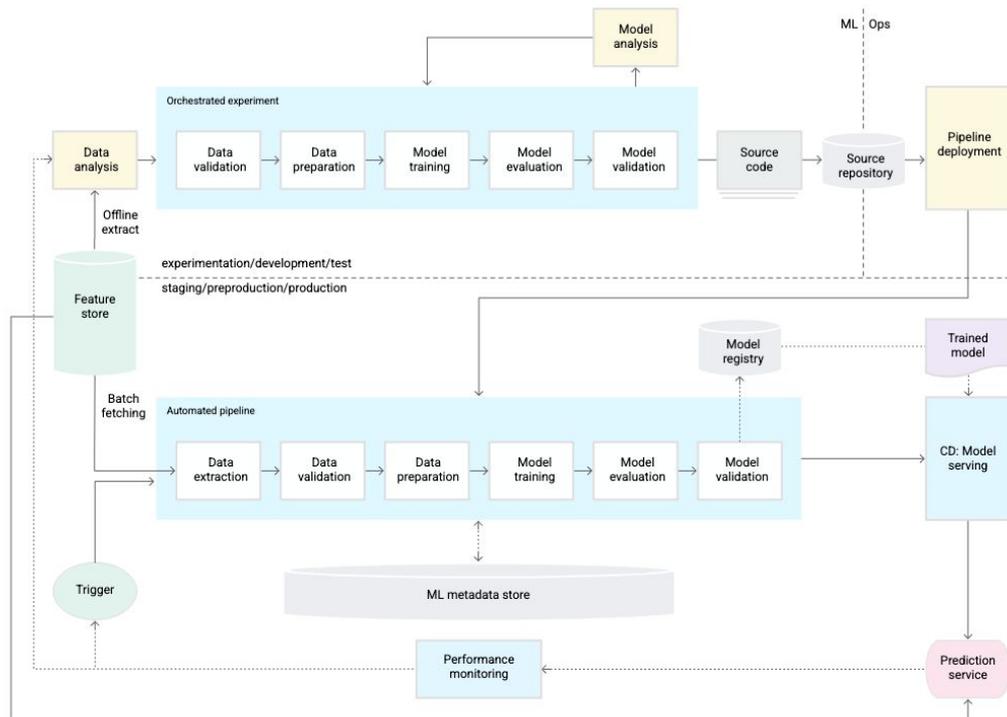
# 機械学習ワークフロー オーケストレーション

# ML パイプラインを使用して ML ワークフローを調整

柔軟なパイプライン構築にKubeflow パイプラインを使用する

- Google Cloud Pipeline Components を使用して、サービスと直接やり取りします
- カスタム機能用の新しいコンポーネントを作成します

TensorFlow Extended SDK を使用して、データの取り込み、データの検証、トレーニングに事前に構築されたコンポーネントを活用します



# アーティファクト構成

## ML モデルのアーティファクトを整理

ストレージの場所	アーティファクト
ソース コントロール リポジトリ	<ul style="list-style-type: none"><li>• ノートブック</li><li>• パイプラインのソースコード</li><li>• 前処理関数</li><li>• モデルのソースコード</li></ul>
テストと ML メタデータ	<ul style="list-style-type: none"><li>• テスト</li><li>• パラメータ</li><li>• 指標</li><li>• データセット (リファレンス)</li><li>• パイプライン メタデータ</li></ul>
Vertex AI	<ul style="list-style-type: none"><li>• トレーニング済みモデル</li></ul>
Artifact Registry	<ul style="list-style-type: none"><li>• パイプライン コンテナ</li><li>• カスタム トレーニング環境</li><li>• カスタム予測環境</li></ul>
Vertex Prediction	<ul style="list-style-type: none"><li>• デプロイするモデルの数</li></ul>

# モデルのモニタリング

# モデルをモニタリングするためのアプローチ

- スキュー検出

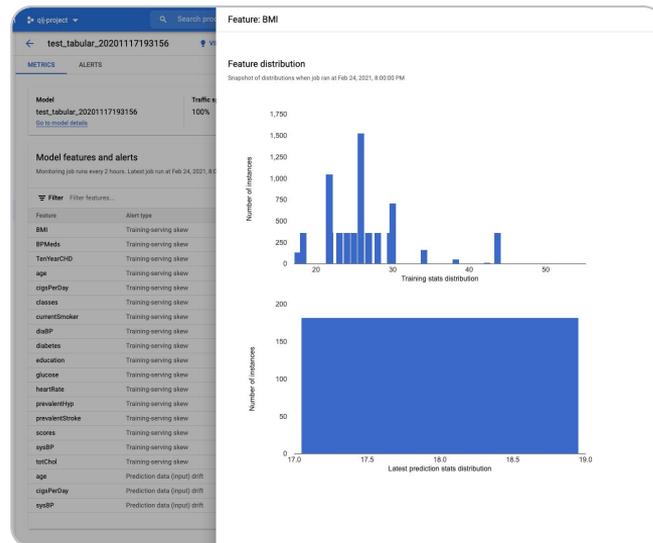
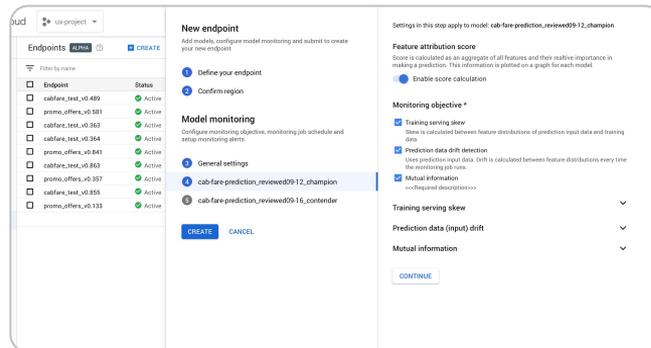
- トレーニングとサービングデータの相違

- ドリフト検出

- サービング時のデータ分布の変化
- トレーニングデータにアクセスできない場合に使用

- 特徴量アトリビュートのドリフト検出

- 予測に使用される特徴量の変更
- 非構造化データでも使用可能



**Thank you**