

# Vertex AI Forecast を用いた、 高度な機械学習モデルによる 最高精度の予測

Google Cloud

カスタマー エンジニアリング

AI/ML スペシャリスト、博士(工学)

児玉 敏男

<b>Vertex AI Forecast の機能</b>	01
<b>ケーススタディ</b>	02
<b>ワークフロー&amp;デモ</b>	03

01

# Vertex AI Forecast の機能

# Google Cloud : みんなのための機械学習ツール



## ビジネスユーザー 洞察と課題



インタラクティブBI  
Looker



スプレッドシートの  
EDW  
Connected  
Sheets



自然言語クエリ  
Data QnA



## データアナリスト クエリと分析



エンドレスEDW  
BigQuery



自己管理型データ  
パイプライン  
Data Fusion,  
Dataflow



データモデル、カ  
タログ  
Looker, Data  
Catalog



SQL での機械学  
習  
BigQuery ML



## データエンジニア 有効なデータの取得



自動運転インフラ  
BigQuery,  
Dataflow,  
Composer



ツール/言語の幅  
広い選択肢  
Dataproc,  
Dataflow



データ品質/系統  
Vertex AI,  
BigQuery, Dataflow



リアルタイム機能  
BigQuery,  
Dataflow



## データサイエンティスト モデルの作成



ポータブルノート  
ブック  
Notebooks



モデルの評価と選  
択  
Explainable AI,  
Tensorboard



ポイントアンドク  
リック開発  
AutoML



コラボレーション  
Feature Store,  
Pipelines



管理対象モデル  
Forecast



## ML 開発者 インテリジェントな APP 作 成



画像、動画  
Vision, Video  
Intelligence



感情分析、エン  
ティティ抽出  
NL, Translation



チャットボット、音  
声コマンド  
Conversation



フリーティアイン  
グ、予測  
Optimization



## ML エンジニア 本番環境でモデルを稼働



スケーラブルなモ  
デルホスティング  
Prediction



ML CI/CD とオー  
ケストレーション  
Pipelines



来歴と血統  
ML Metadata



改善と再トレーニ  
ング  
Model  
Monitoring



# Vertex AI

## Applications

画像

会話

言語

構造化データ

## Core

Workbench

Data Labeling

Deep Learning Env

Experiments

Metadata

AutoML

Training

Explainable AI

Feature Store

Vizier (Optimization)

Prediction

Continuous Monitoring

Pipelines

AI Accelerators

Hybrid AI

# 構造化データの課題のタイプ

## 予測 (Forecasting)

製品が来月いくつ販売されますか？

Sales	Date	Channel	SKU	...	Geo	Domain	Brand
\$1,200	Jun-10-2020	Website	12345	...	US	Shoes	Nike
\$1,350	Jun-11-2020	Email	54221	...	US	Shirts	Adidas
	Jan-10-2021	Website	22345	...	CA	Shoes	Asics

## 分類

この商品は7日内で販売されますか？

Sold	...	ID	Geo	Domain	Title	Description	Tags	Brand
YES	...	104	US	Shoes	"Dark red..."	"Try this soft..."	["A, B, ..."]	Nike
NO	...	204	US	Shoes	"Women's..."	"Medium-size..."	["A, B, ..."]	Adidas
	...	302	CA	Shoes	"Running..."	"All-terrain..."	["A, B, ..."]	Asics

## 回帰

この製品はどの価格で販売されますか？

Price	...	ID	Geo	Domain	Title	Description	Tags	Brand
\$52	...	104	US	Shoes	"Dark red..."	"Try this soft..."	["A, B, ..."]	Nike
\$48	...	204	US	Shoes	"Women's..."	"Medium-size..."	["A, B, ..."]	Adidas
	...	302	CA	Shoes	"Running..."	"All-terrain..."	["A, B, ..."]	Asics

# Forecastingにディープラーニングモデルを使用する理由

- すべての時系列に対して 1つの「グローバル」モデルを作成します
- 時系列全体のパターンを学習します

## 多数のドライバーを使用できます

- ✓ 豊富なメタデータ  
(例: 商品属性、場所属性)
- ✓ 歴史的要因  
(例: 在庫、天気)
- ✓ 将来的に知られている要因  
(例: 計画されたプロモーション/イベント、休日)

## 複雑なシナリオをモデル化できます

- ✓ コールドスタート/新商品
- ✓ 短い製品ライフサイクル
- ✓ 値のスパース性
- ✓ テキストの説明などの非構造化データ

# Vertex AI Forecast

1. データセットを準備する

2. モデル訓練

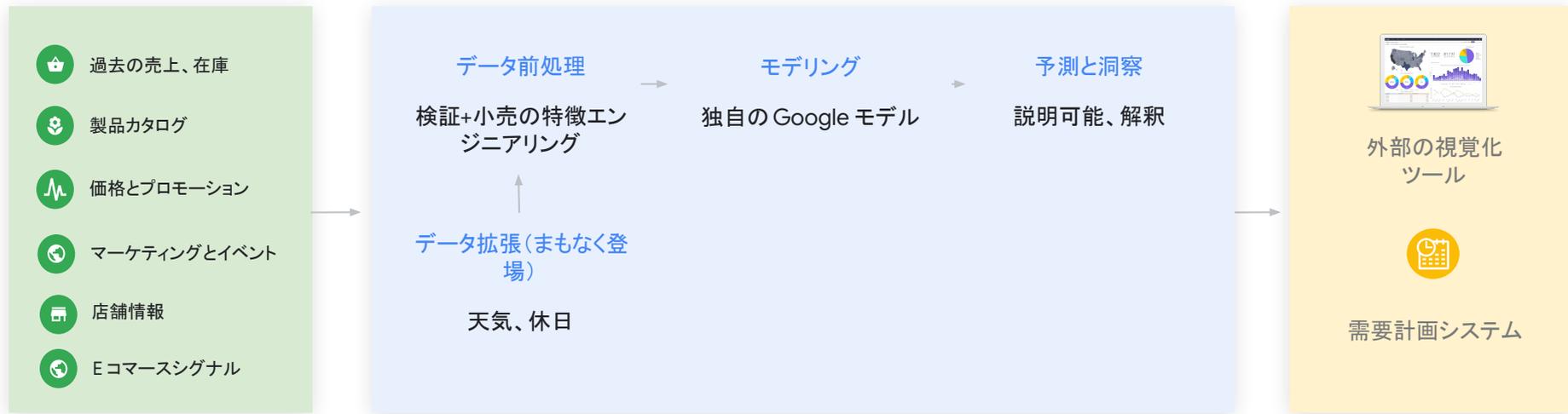
3. 予測

4. 視覚化、統合

## 小売業者データ

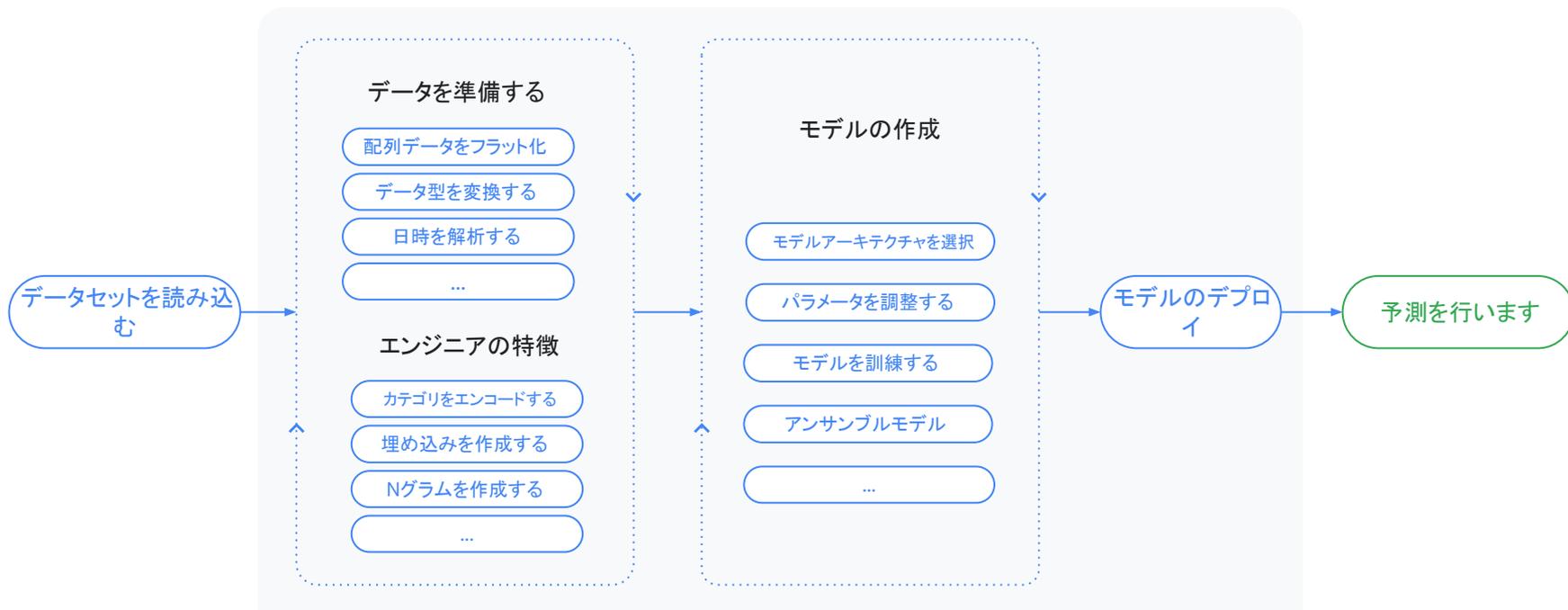
## Vertex AI Forecast

## 統合



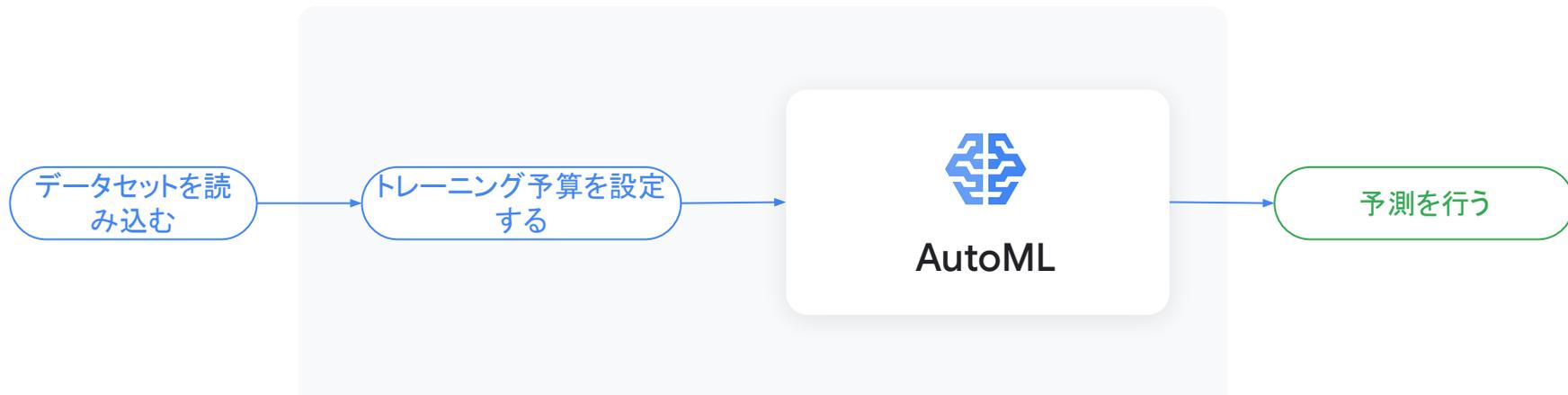
# AutoML - データから値への最速パス

従来の機械学習ワークフロー



# AutoML - データから値への最速パス

AutoML ワークフロー



# Google Brain の最新研究の成果の提供



# 自動特徴エンジニアリング

すべてのデータ型のベストプラクティス変換



**数値:** 分位数、ログ、z\_score 変換を生成します



**日付時間:** 年、月、日、平日を抽出し、分類します



**テキスト:** トークン化、n-gram の生成、埋め込みの作成



**カテゴリ配列:** ルックアップインデックスに変換し、埋め込みを生成します



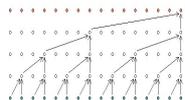
**カテゴリ値:** ワンホットエンコーディング、グループ化、埋め込み



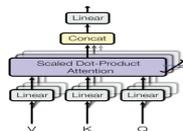
**ネスト型項目:** 平坦化し、型変換を適用します

# 自動モデルアーキテクチャ検索

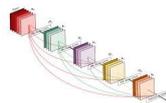
予測のための Google の最適なモデルアーキテクチャの評価



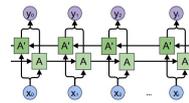
Convolutions



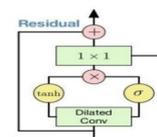
Attention



Skip  
Connections

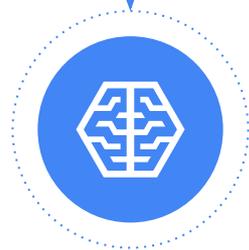


LSTM



Gating

...



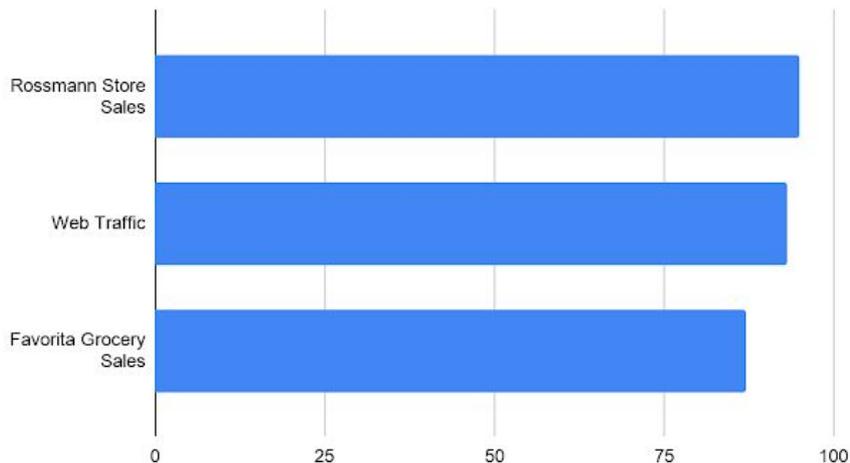
# Kaggle Datasets 上のベンチマーク

世界のトップ予測コンテストで上位 2.5 % (5558 チーム中 138 チーム)に。

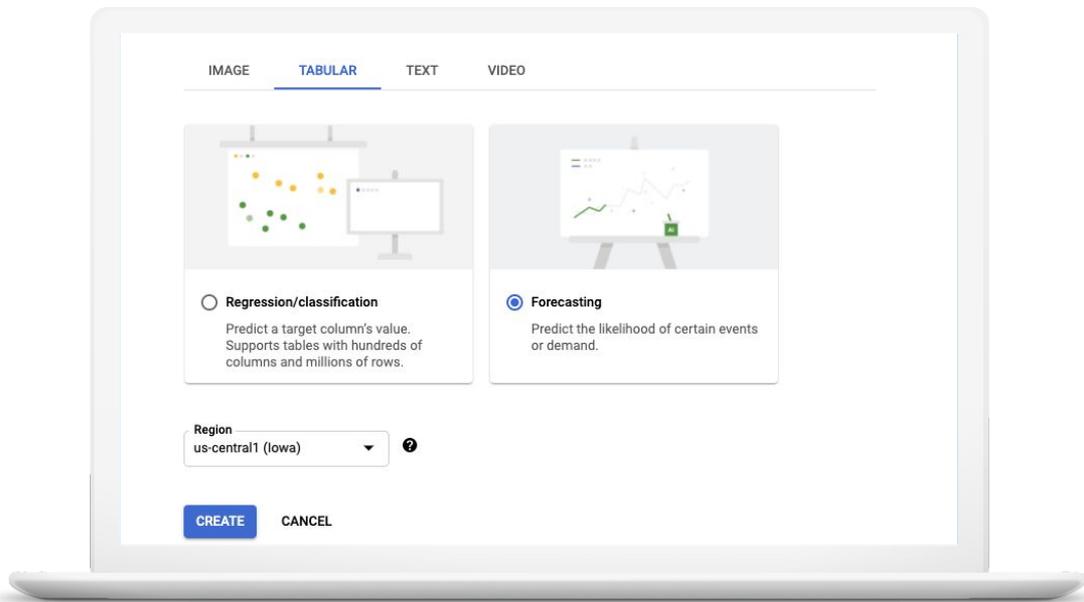
[M5: Estimate the sales of Walmart retail goods.](#)

さまざまな業界のさまざまなデータセットで一貫して上位 20 %にランク付けされています。

Better than % of Kaggle Participants



# さまざまなインターフェース



Intuitive Web UI

## API + SDK

```
In [ ]: # The number of hours to train the model.
        model_train_hours = 1 #@param {type:'integer'}

        create_model_response = tables_client.create_model(
            model_display_name=MODEL_DISPLAY_NAME,
            dataset=dataset,
            train_budget_milli_node_hours=model_train_hours*1000,,
            exclude_column_spec_names=['fnlwgt','income'],
        )

        operation_id = create_model_response.operation.name

        print('Create model operation: {}'.format(create_model_response.operation))
```

## BigQuery SQL\*

```
CREATE OR REPLACE MODEL project_id.mydataset.mymodel
  OPTIONS(model_type='AUTOML_REGRESSOR',
           input_label_cols=['fare_amount'],
           budget_hours=1.0)

AS SELECT
  (tolls_amount + fare_amount) AS fare_amount,
  pickup_longitude,
  pickup_latitude,
  dropoff_longitude,
  dropoff_latitude,
  passenger_count
FROM `nyc-tlc.yellow.trips`
WHERE ABS(MOD(FARM_FINGERPRINT(CAST(pickup_datetime AS STRING)), 100000)) = 1
```

\* currently only supports auto-Arima (no AutoML support) Cloud

02

# ケーススタディ

# 1. 米国を拠点とする大手アパレル小売業者

## 課題

この高品質でオントレンドなアパレル小売業者にとって、需要計画はビジネスの成功の鍵です。同社は、ビッグデータ、予測の粒度、リアルタイムの需要のシグナルに関連する課題に対処するための新しいアプローチが必要であることを認識していました。

## 予測を改善してビジネス上の意思決定を強化する

Vertex AI Forecast を活用することで、同社は、新しいサプライチェーン要件を予測機能に統合しながら、需要と労働力の予測の精度を向上させることができました。



**2 ~ 3 ヶ月**

モデル開発時間の短縮

**年間 500 万ドルから  
1000 万ドル**

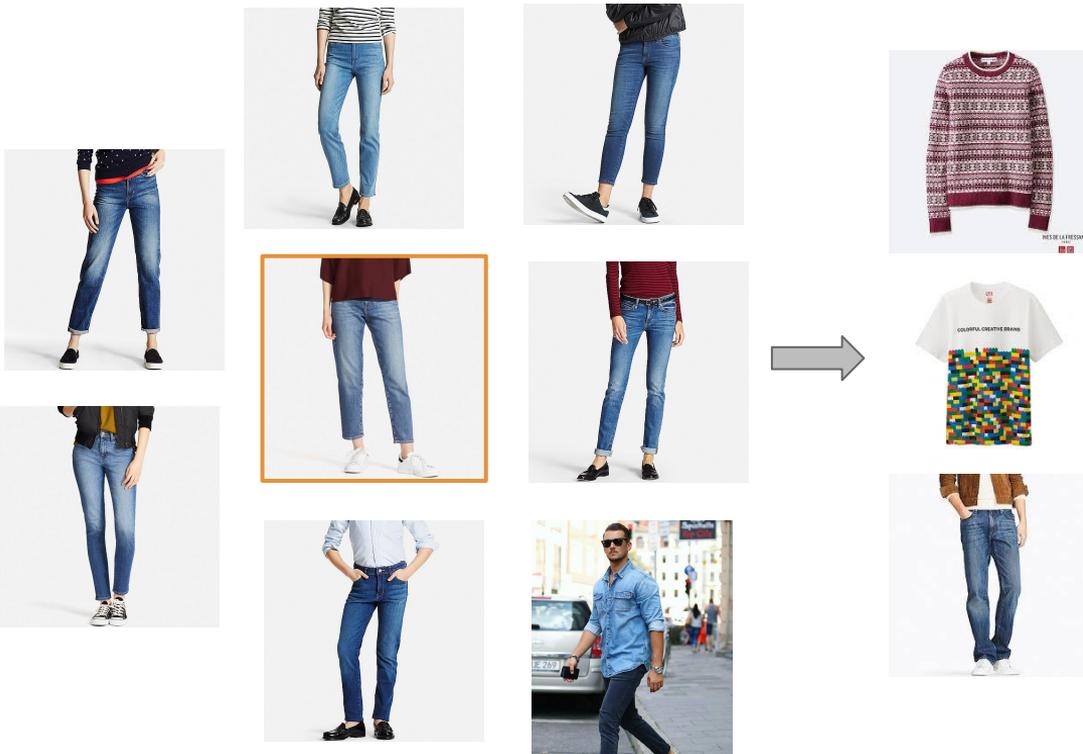
労働効率の向上による節約

**4,000 万ドル以上**

サプライチェーンと製品の割り当ての改善による収益アップ

## アパレルのシーズン前の予測

- **シーズン前とシーズン中の両方:**  
シーズン開始の 3 か月前までのデータに基づいて、1 年先までの予測
- **コールドスタート項目:** 製品間の関係は自動的に学習することが必要
- **細分性:** 商品のバリエーション(色 x サイズなど)から日 xSKULレベルまでの需要予測をブレイクダウン



# 小売予測の入門: 品目別の需要

## 商品メタデータ



- サマードレス
- 軽量。パターン。
- 長袖。Vネック。
- 55%リネン。綿 45%
- 洗濯機洗い

## 需要ドライバー



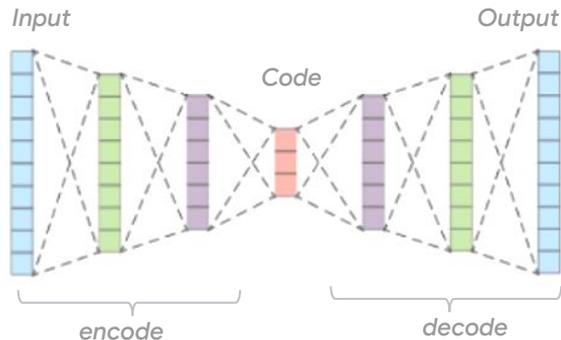
- 売上高
- 価格
- 競合他社の価格
- プロモーション/イベント
- 休日

## 店舗情報



- 店舗の説明: 大、小、専門
- 店の所在地
- 足の交通

## Vertex AI FORECAST



機械学習モデルは、豊富なメタデータ、商品間の関係、および価格設定、製品ライフサイクルなどのシナジーを理解します

## ミディアムホライズン 12 ~ 16 か月

- 例えばシーズン前の計画
- 購入/注文計画
- 新規およびコールドスタートアイテム

## ショートホライズン 0 ~ 8 週間

- シーズン中の計画
- 補充、在庫
- 価格設定、割り当て
- SKU レベル

## 2. グローバルホスピタリティプロバイダー

### 課題

ホスピタリティとパッケージオーナーのリーダーは、COVID 制限の緩和中に需要を予測するのに役立つソリューションを必要としていました。既存のモデルは、絶えず変化する市場のダイナミクスと価格決定を行うために必要な速度で予測結果を出力できませんでした。

### 改善された需要予測の精度とリアルタイムの洞察

Vertex AI Forecast を活用して、同社はBigQuery データを使用して、ワクチン接種後に予期していなかったトラフィックパターンを確認しました。改善された需要予測の精度を利用することで、部屋あたりの収益と占有率の増加につなげました。



**リアルタイム、自動化**  
予測プロセス

**500 万ドル+**  
一部の物件の追加収益

# ホスピタリティ予測の入門: 部屋の占有率

## 部屋のメタデータ



- ラグジュアリースイート
- オーシャンビュー
- 禁煙
- キングサイズベッド

## プロパティデータ



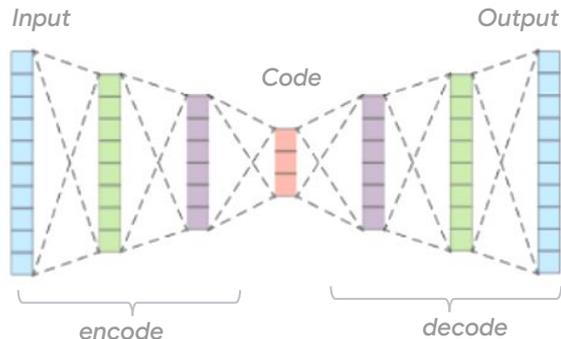
- 位置
- タイプ: ラグジュアリー、レギュラー
- 国
- 空港の近さ
- トリップアドバイザーの星

## 需要ドライバー



- 価格、プロモーション
- 競合他社の在庫
- 地元のイベント、休日
- ウェブサイト検索
- チケット販売

## AUTOML FORECAST



機械学習モデルは、豊富なメタデータ、競合他社の在庫、天気、地域のイベント、物件と種類。

## ミディアムホライズン 6 ~ 12 か月

- 財務計画
- 資本配分
- 価格計画

## ショートホライズン 0 ~ 8 週間

- プロモーション計画
- 労働力の割り当て

03

## ワークフロー&デモ

# Google でのForecastのワークフロー

Google のディープラーニングと統計モデルを使用して、高品質でスケーラブルな予測ソリューションを構築します。



# デモ: データセットの作成

データ取得

JQuants\_API:

- Colab で 10 銘柄データ取得
- csv 出力

データセット登録

BigQuery:

- csv 読み込み

	Code	Close	Date	SectorName
0	69710	5997.0	20170104	電気機器
1	69710	5907.0	20170105	電気機器
2	69710	5904.0	20170106	電気機器
3	69710	5833.0	20170110	電気機器
4	69710	5892.0	20170111	電気機器
...	...	...	...	...
11168	72960	1301.0	20220428	輸送用機器
11169	72960	1310.0	20220502	輸送用機器
11170	72960	1327.0	20220506	輸送用機器
11171	72960	1342.0	20220509	輸送用機器
11172	72960	1360.0	20220510	輸送用機器

クエリ結果

ジョブ情報	結果	JSON	実行の詳細	
行	Code	Close	date	SectorName
11101	49280	4925.0	2022-04-25	化学
11102	49320	927.0	2022-04-25	化学
11103	90290	608.0	2022-04-25	陸運業
11104	72960	1263.0	2022-04-26	輸送用機器
11105	62490	1002.0	2022-04-26	機械
11106	65570	712.0	2022-04-26	サービス業
11107	29220	1981.0	2022-04-26	食料品
11108	69710	6753.0	2022-04-26	電気機器
11109	39260	2006.0	2022-04-26	情報・通信業
11110	90480	2088.0	2022-04-26	陸運業
11111	49280	5000.0	2022-04-26	化学
11112	49320	953.0	2022-04-26	化学

# デモ: データセットの作成

← データセットを作成

データセット名 \*  
jpx\_forecast\_prediction\_10\_companies\_20220512

最大 128 文字を使用できます。

## データタイプと目標の選択

まず、データセットに含めるデータタイプを選択します。次に、目標（トレーニングしたモデルで達成したい結果）を選択します。 [詳細](#)

画像 **表形式** テキスト 動画



回帰 / 分類

ターゲット列の値を予測します。数百の列、数百万の行があるテーブルをサポートします。



予測 **プレビュー**

特定のイベントまたは需要の可能性を予測します。

リージョン  
us-central1 (アイオワ) ▼ ?

作成

キャンセル

← jpx\_forecast\_prediction\_10\_companies\_20220512

ソース 分析

## データセットにデータを追加

開始する前に、[データのガイド](#)を読んでデータの準備方法をご確認ください。次に、データソースを選択します。

予測では、時系列内で観測した履歴値に基づいて、将来の値を予測します。

## データソースを選択

- **CSV file:** Can be uploaded from your computer or on Cloud Storage. [Learn more](#)
- **BigQuery:** Select a table or view from BigQuery. [Learn more](#)

- パソコンから CSV ファイルをアップロード
- Cloud Storage から CSV ファイルを選択
- テーブルまたはビューを BigQuery から選択

テーブルまたはビューを **BigQuery** から選択

BigQuery テーブルまたはビューをデータソースとして使用します。 [データセットへのアクセス権](#)、[データセット ID とテーブル](#)が必要です。 [詳細](#)

BigQuery のパス \*

mythic-flash-340704.jpx.automl\_table

参照 ?

形式は次のようになります: projectId.datasetId.tableId。省略可能なフィールドが空欄の場合は、新しいフィールドが作成されます。

## 次のステップ

選択された BigQuery テーブルは、データセットに関連付けられます。参照された BigQuery テーブルに変更を加えると、トレーニング前のデータセットに影響します。

続行

# デモ: Training

← jpx\_forecast\_prediction\_10\_companies\_20220512



ソース 分析

## プロパティ

作成日	5月 12, 2022 11:28 午前
データセットのフォーマット	BigQuery
Dataset location(s)	<a href="#">bq://mythic-flash-...jpx.automl_table</a>

## 概要

合計行数: 5

合計行数: -

時系列の識別子列

Code

タイムスタンプ列

date

INTEGER	1 (20%)
FLOAT	1 (20%)
DATE	1 (20%)
STRING	2 (40%)

[統計情報を生成](#)

≡ フィルタ プロパティ名または値を入力

列名 ↑	BigQuery の型	BigQuery のモード	欠損 % (カウント)	固有の値
Close	FLOAT	NULLABLE	-	-
Code <a href="#">時系列の識別子</a>	INTEGER	NULLABLE	-	-
date <a href="#">タイムスタンプ列</a>	DATE	NULLABLE	-	-
Section	STRING	NULLABLE	-	-
SectorName	STRING	NULLABLE	-	-

>|

## Training jobs and models

Use this dataset and annotation set to train a new machine learning model with AutoML or custom code

[新しいモデルをトレーニング](#)

# デモ: Training

## Train new model

- 1 トレーニング方法
- 2 モデルの詳細
- 3 トレーニング オプション
- 4 コンピューティングと料金

トレーニングを開始

キャンセル

Dataset

jpx\_forecast\_prediction\_10\_companies\_20220512



各方法の詳細（および利用可能なデプロイのオプション）については、価格ガイドを参照してください。



管理されたデータセットを使用したカスタムトレーニングは現在利用できません。[カスタム モデルトレーニングの詳細をご覧ください](#)

### Model training method

- AutoML  
最小限の労力と機械学習の専門知識で高品質のモデルをトレーニングします。トレーニングする期間を指定するだけです。[詳細](#)
- カスタム トレーニング（上級者向け）  
TensorFlow、scikit-learn、XGBoost トレーニング アプリケーションをクラウドで実行します。Google Cloud のビルド済みコンテナのいずれかを使用してトレーニングするか、独自のコンテナを使用してください。[詳細](#)

続行

# デモ: Training

## Train new model

- トレーニング方法
- 2** モデルの詳細
- 3 トレーニング オプション
- 4 コンピューティングと料金

トレーニングを開始

キャンセル

新しいモデルのトレーニング  
新しいモデルのグループを作成し、トレーニング済みのモデルをバージョン 1 として割り当てます

新しいバージョンのトレーニング  
モデルを既存のモデルの 1 バージョンとしてトレーニングします

名前 \*  
jpx\_forecast\_prediction\_10\_companies\_20220512

説明

Target column \*  
Close (FLOAT) ?

Series identifier column \*  
Code (INTEGER) ?

タイムスタンプ列 \*  
date (DATE) ?

### Forecasting configuration

データの粒度 \*  
Daily

タイムスタンプ列の粒度レベル。粒度はすべての行で同じでなければなりません。たとえば、[日] を選択した場合、タイムスタンプは相互の 1 日以内にする必要があります。データの粒度により、予測ホライズンとコンテキスト期間の期間粒度も設定されます。

Forecast horizon \*  
7

予測を作成する将来の期間の数。将来の期間は、データセット内の最新のタイムスタンプから開始します。

コンテキスト期間 \*  
30

各時系列のモデルに入力ラグを定義します。ほとんどのユースケースでは、コンテキスト期間は予測ホライズン値の 0~5 倍になります。出発点として、コンテキスト期間を予測ホライズン値と同じ値に設定してみます。 [詳細](#)

予測を作成する将来の期間の数。将来の期間は、データセット内の最新のタイムスタンプから開始します。

コンテキスト期間 \*  
30

各時系列のモデルに入力ラグを定義します。ほとんどのユースケースでは、コンテキスト期間は予測ホライズン値の 0~5 倍になります。出発点として、コンテキスト期間を予測ホライズン値と同じ値に設定してみます。 [詳細](#)

Export test dataset to BigQuery

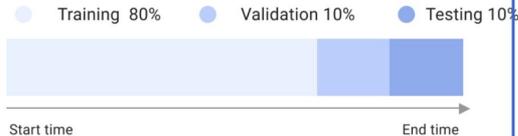
テストセットを新しい BigQuery テーブルにエクスポートします。すでに存在するテーブル ID は使用できません。「ProjectId.DatasetId.tableId」という形式を使用します。

BigQuery のパス  
 mythic-flash-340704.jpx.automl\_table\_forecast\_10\_result\_20220512 参照

形式は次のようになります: projectid.datasetid.tableid。省略可能なフィールドが空欄の場合は、新しいフィールドが作成されます。

### データ分割

時系列  
データセットはタイムスタンプで並べ替えられます。最も時間の古い 80% の行がトレーニングに、次の 10% が検証に、最も時間の新しい 10% がテストに割り当てられます。 [詳細](#)



手動  
ユーザーが、トレーニング用、検証用、テスト用に各データ行を割り当てます。 [詳細](#)

↑ 閉じる

続行

# デモ: Training

## Train new model

- トレーニング方法
- モデルの詳細
- トレーニングオプション**
- コンピューティングと料金

トレーニングを開始

キャンセル

1 続行するには、[変換] 列を使用してデータセット内のデータ型を確認し、指定します。指定しない場合、AutoML は最も関連性の高い変換オプションの適用を試行します。

[統計情報を生成](#)

≡ フィルタ プロパティ名または値を入力 ?

<input type="checkbox"/>	列名 ↑	変換	特徴タイプ <span>?</span>	予測時に利用可能 <span>?</span>	BigQuery の型	BigQuer
<input type="checkbox"/>	Close <span>ターゲット</span>	数値 ▾	共変量 ▾	利用不可 ▾	FLOAT	NULLA
<input type="checkbox"/>	Code <span>時系列の識別子</span>				INTEGER	NULLA
<input type="checkbox"/>	date <span>タイムスタンプ列</span>	タイムスタンプ ▾	共変量 ▾	利用可能 ▾	DATE	NULLA
<input type="checkbox"/>	Section	自動 ▾	— ▾	— ▾	STRING	NULLA ⊕
<input type="checkbox"/>	SectorName	カテゴリ ▾	共変量 ▾	利用可能 ▾	STRING	NULLA ⊖

トレーニングに含まれる特徴列は合計 3 列です

### 重み列

Select a column to specify how to weight each row of the training data. By default, each row of your training data is weighted equally. ?

### 最適化の目標\*

- RMSE (デフォルト)**  
より極端な値を正確に取り込みます
- MAE**  
モデルへの影響を抑えて、極端な値を外れ値として表示します
- RMSLE**  
絶対値ではなく、相対サイズに基づいてエラーにペナルティを適用します。予測値と実際の値の両方がかなり大きくなる可能性がある場合に特に役立ちます。予測または正解が 0 未満の場合は未定義になります。
- RMSPPE**  
RMSE と似ていますが、ターゲットの大きさに比例します。値の範囲が大きい場合に便利です。
- WAPE**  
時間の経過に伴う集計後のターゲットの絶対エラー率を測定します。実際の値が低い場合に便利です。
- 分位点損失**  
分位点回帰を行う際に使用します

[詳細オプション](#)

[続行](#)

# デモ: Evaluate Model

Vertex AI

← jpx\_forecast\_prediction\_10\_companies\_20220512 > バージョン 1 ▾ [データセットを表示](#)

[評価](#)   [バッチ予測](#)   [バージョンの詳細](#)

Target column	MAE <sup>?</sup>	MAPE <sup>?</sup>	RMSE <sup>?</sup>	RMSLE <sup>?</sup>	r <sup>2</sup> <sup>?</sup>
Close numeric	69.054	4.062	107.859	0.059	0.997

- ダッシュボード
- データセット
- 特徴量
- タスクのラベル付け
- ワークベンチ
- パイプライン
- トレーニング
- テスト
- モデル**
- エンドポイント
- バッチ予測
- メタデータ
- マッチングエンジン

# デモ: Create Prediction

The screenshot shows the Google Cloud Platform interface for creating a new batch prediction. The left sidebar contains the navigation menu with 'Vertex AI' selected. The main content area is titled '新しいバッチ予測' (New Batch Prediction) and includes the following elements:

- Batch Prediction Name:** A text input field labeled 'バッチ予測名 \*'.
- Model Name:** A dropdown menu labeled 'モデル名 \*' with the selected value 'jpx\_forecast\_prediction\_10\_companies\_20220512'.
- Region:** A dropdown menu labeled 'リージョン' with the selected value 'us-central1 (アイオワ)'.
- Filters:** A section titled 'フィルタ' with a list of prediction names: '名前', 'req\_test', 'forecast\_test', 'forecast\_predict\_week\_5', 'forecast\_predict\_week\_4', 'forecast\_predict\_week\_3', 'forecast\_predict\_a\_week\_2', 'forecast\_predict\_a\_week', 'forecast\_predict\_3', 'forecast\_predict\_2', 'reg\_prediction', and 'forecast\_prediction'.
- Output Options:** A section titled 'バッチ予測の出力' with the instruction '予測結果の形式と出力場所を選択します'. It includes an '出力形式' (Output Format) field and a 'gs:// 宛先のパス \*' (gs:// Destination Path) field with a '参照' (Reference) button. A note below states: '選択するバケットは、Standard Storage クラス、単一リージョンで、us-central1 内にある必要があります'.
- Actions:** A section titled '詳細オプション' (Advanced Options) with '作成' (Create) and 'キャンセル' (Cancel) buttons.

**Thank you.**