Vertex AI ではじめる時系列分析入門

Google Cloud
カスタマー エンジニア
葛木美紀

Google Cloud

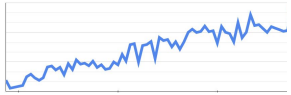
時系列分析の課題

01

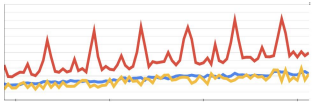
時系列のタイプと要素

データセット

一変量



多変量



一変量の時系列例

- 日付, 売上

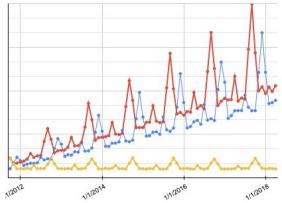
多変量時系列の例

- 日付, 売上, 気温, 株価, 広告費, etc.

時系列を構成する要素

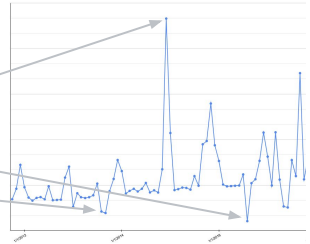
周期性

- 季節性
- トレンド
- その他のランダム性



特徴量に依存

- プロモーション
- 台風
- 競合の価格低下



分析にあたり考慮が必要な時系列データのタイプ

- コールド スタート アイテムの需要予測、ノイズの多いスパースデータセット、さまざまな長さの系列、外れ値、トレンド変化

時系列分析の基本的な考え方

通常の回帰分析データ

X	y
5	1
4	0
5	1
3	1
6	0
3	1

$$Y = \boldsymbol{F}(x)$$

目標：入力データ(X)から出力変数(y)を予測

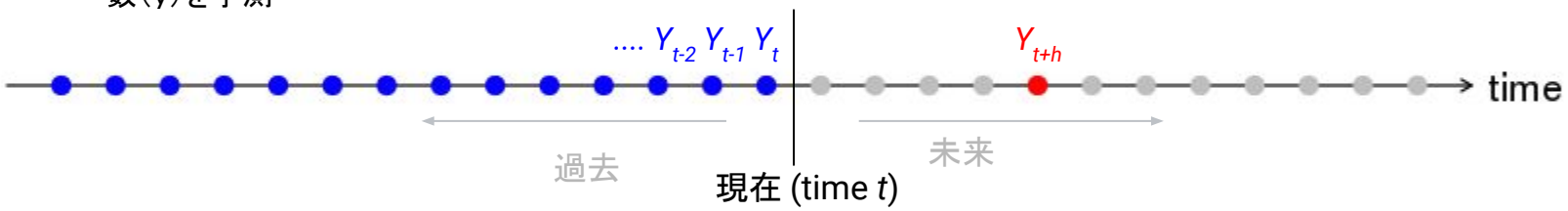
時系列データ

Time	Measure
1	1
2	0
3	1
4	0
5	0
6	?

一時点前のデータから y を予測

Time	Lag_1 or X	y
1	?	1
2	1	0
3	0	1
4	1	0
5	0	0
6	0	?

過去の自身の値から将来の値を予測



ARIMA: **A**uto**R**egressive **I**ntegrated **M**oving **A**verage

ARIMA モデルの 3 つの変数 (*p, d, q*)

p: モデルに含まれるタイムラグの数
（AR モデルの次数）。
ラグオーダーとも呼ばれる

d: 観測値の**差分の**次数。
degree of differencing とも呼ばれる

q: 移動平均ウィンドウのサイズ
（MA モデルの次数）。
移動平均のオーダー とも呼ばれる

$$\underbrace{y'_t}_{\text{differenced time series}} = \overbrace{c}^{\text{intercept}} + \underbrace{\varphi_1 y'_{t-1} + \ldots + \varphi_p y'_{t-p}}_{\text{lagged values}} + \underbrace{\theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}}_{\text{lagged errors}} + \varepsilon_t$$

時系列分析の問題点

例：ARIMA モデル

{p,d,q} のような変数を決定するため以下のような

ヒューリスティックなルールが必要

- 単位根検定
  - 有意かそうでないかで d を決定
- 自己相関関数（ACF）、偏自己相関
  - コレログラムから order(p,q) を決定
- 残差分析
  - どれくらい残差に時系列が残ってるかによって p,d,q の値を変更
- 視覚化
- …

(´・ω・`)

Summary of rules for identifying ARIMA models

Identifying the order of differencing and the constant:

- Rule 1: If the series has positive autocorrelations out to a high number of lags (say, 10 or more), then
- Rule 2: If the lag-1 autocorrelation is zero or negative, or the autocorrelations are all small and patter autocorrelation is -0.5 or more negative, the series may be overdifferenced. **BEWARE OF OVERDIFF**
- Rule 3: The optimal order of differencing is often the order of differencing at which the standard devia differencing can also be corrected with AR or MA terms. See rules 6 and 7.)
- Rule 4: A model with no orders of differencing assumes that the original series is stationary (among ot the original series has a constant average trend (e.g. a random walk or SES-type model, with or witho series has a time-varying trend (e.g. a random trend or LES-type model).
- Rule 5: A model with no orders of differencing normally includes a constant term (which allows for a n not include a constant term. In a model with one order of total differencing, a constant term should be

Identifying the numbers of AR and MA terms:

- Rule 6: If the partial autocorrelation function (PACF) of the differenced series displays a sharp cutoff a "underdifferenced"--then consider adding one or more AR terms to the model. The lag beyond which t
- Rule 7: If the autocorrelation function (ACF) of the differenced series displays a sharp cutoff and/or th "overdifferenced"--then consider adding an MA term to the model. The lag beyond which the ACF cuts
- Rule 8: It is possible for an AR term and an MA term to cancel each other's effects, so if a mixed AR-M fewer MA term--particularly if the parameter estimates in the original model require more than 10 iter **MA TERMS IN THE SAME MODEL.**
- Rule 9: If there is a unit root in the AR part of the model--i.e., if the sum of the AR coefficients is almo order of differencing by one.
- Rule 10: If there is a unit root in the MA part of the model--i.e., if the sum of the MA coefficients is aln order of differencing by one.
- Rule 11: If the long-term forecasts* appear erratic or unstable, there may be a unit root in the AR or

Identifying the seasonal part of the model:

- Rule 12: If the series has a strong and consistent seasonal pattern, then you must use an order of sea fade away over time). However, never use more than one order of seasonal differencing or more than
- Rule 13: If the autocorrelation of the appropriately differenced series is positive at lag s, where s is th the autocorrelation of the differenced series is negative at lag s, consider adding an SMA term to the n which should be done if the data has a stable and logical seasonal pattern. The former is likely to occu seasonal pattern is not stable over time. You should try to avoid using more than one or two seasonal the data and/or problems in estimation.

***A caveat about long-term forecasting in general:** linear time series models such as ARIMA and expone period-ahead forecasts and plugging them in for unknown future values as they look farther ahead. For exar

Duke 大学より引用

Vertex AI AutoML による時系列分析

02

Proprietary + Confidential

Vertex AI

アプリケーション

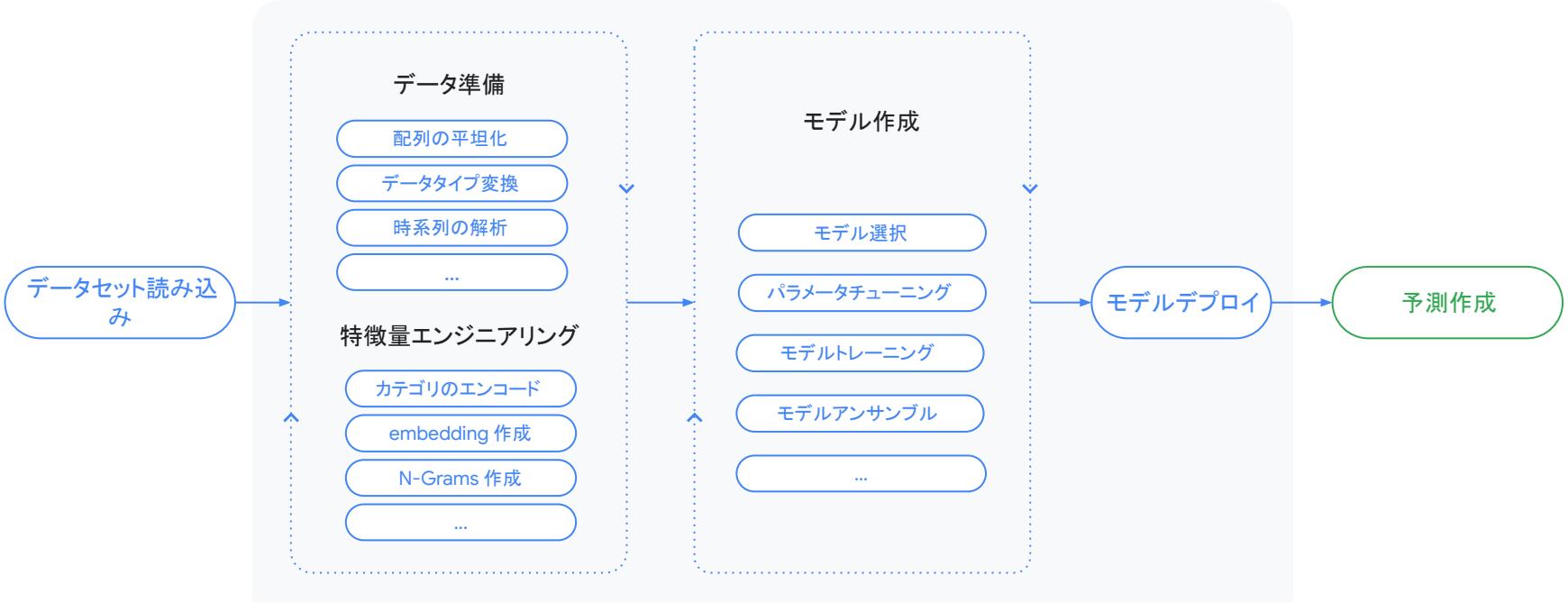
- 画像と動画
- 音声・会話
- 言語
- 構造化データ

カスタムモデル

- Notebooks
- AutoML
- NAS
- Prediction
- ML Metadata
- Data Labeling
- Training
- Explainable AI
- Feature Store
- Model Monitoring
- Experiments
- Vizier (Optimization)
- Pipelines
- Matching Engine

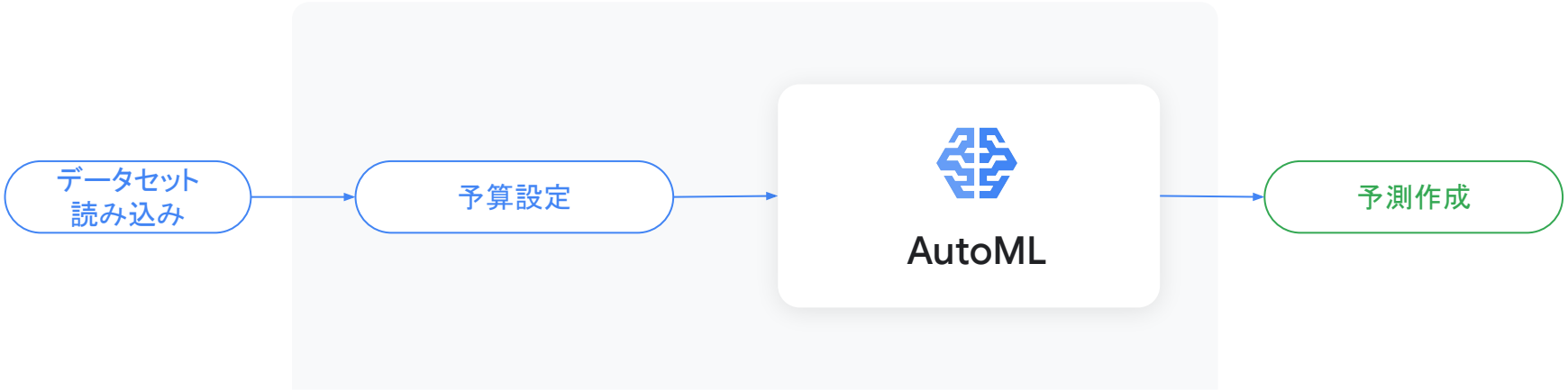
Proprietary + Confidential

従来の ML ワークフロー



Proprietary + Confidential

AutoML - データから最短で価値を得る

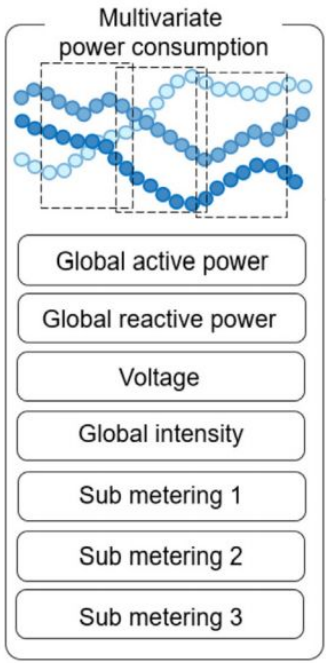


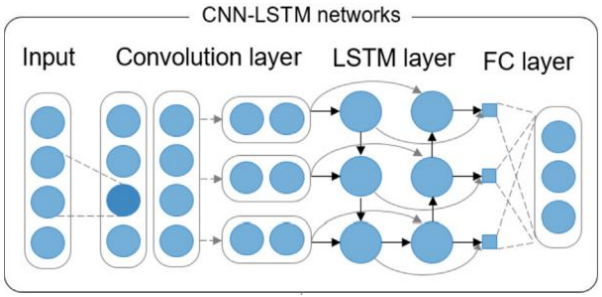
自動的に行われる前処理

あらゆるデータ型をベストプラクティスで変換

- 数値：分位数、対数、z_score変換を生成
- 日付: 年、月、日、平日を抽出し、分類
- テキスト: tokenize、n-gram生成、Embedding 作成
- カテゴリの配列：ルックアップインデックスに変換し、Embedding 生成
- カテゴリ：ワンホットエンコーディング、グループ化、Embedding
- ネストされたフィールド：フラット化し、型変換を適用

Deep Learning による時系列分析





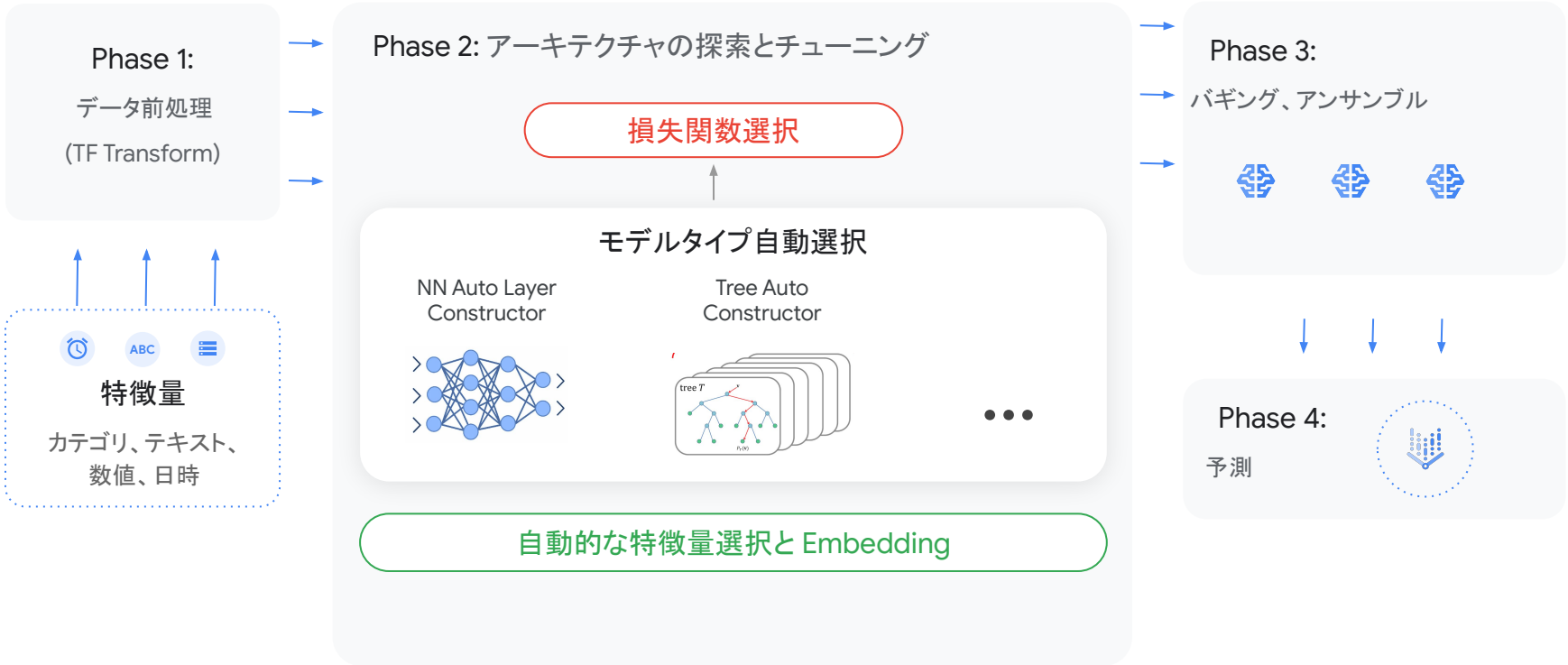
将来の外部要因値より予測

- RNN (Recurrent Neural Network)
- LSTM (Long Short-Term Memory)

Deep learning が向いているケース

- **さまざまな外部要因**を使いたい
  - 商品属性, 位置情報などメタデータ
  - 在庫数、天候など過去履歴
  - 予測時に使用できる情報
    (例: 広告/イベント,休日)
- **複雑なシナリオ**に対応
  - データ収集の頻度が高い
    (例：秒単位のセンサーデータ)
  - コールドスタート、不規則な傾向

Google Brain による最新の研究結果を活用



Vertex Forecast: 特徴量の定義

共変量: 時間の経過にともない変化する外生変数
例：休日、予定されているプロモーション、イベント、天気など

属性: 時系列にともない変化しない変数
例：商品の色、重量、または製品の説明

予測時に使用可能：天気予報

予測時に使用不可: 実際の天気

Time	Timeseries ID	Target	Features				
Date	Item	Item_Sales	Store_1 Foot Traffic	Item Website Clicks	Promo	Color	Price
Jan-01-19	Widget_1	297	437	172	0	Blue	25
Jan-02-19	Widget_1	223	105	116	0	Blue	25
Jan-03-19	Widget_1	222	192	42	0	Blue	25
Jan-04-19	Widget_1	280	143	93	0	Blue	25
Jan-05-19	Widget_1	264	252	92	0	Blue	25
•••							
Feb-02-19	Widget_1	353	121	171	0	Blue	23
Feb-03-19	Widget_1	286	132	85	0	Blue	23
Feb-04-19	Widget_1	215	419	83	0	Blue	23
Feb-05-19	Widget_1	343	232	144	0	Blue	28
Feb-06-19	Widget_1			127	0	Blue	28
Feb-07-19	Widget_1			25	0	Blue	28
Feb-08-19	Widget_1			151	2	Blue	28
Feb-09-19	Widget_1			102	0	Blue	28
Feb-10-19	Widget_1			139	0	Blue	28

過去のデータ

予測リクエスト

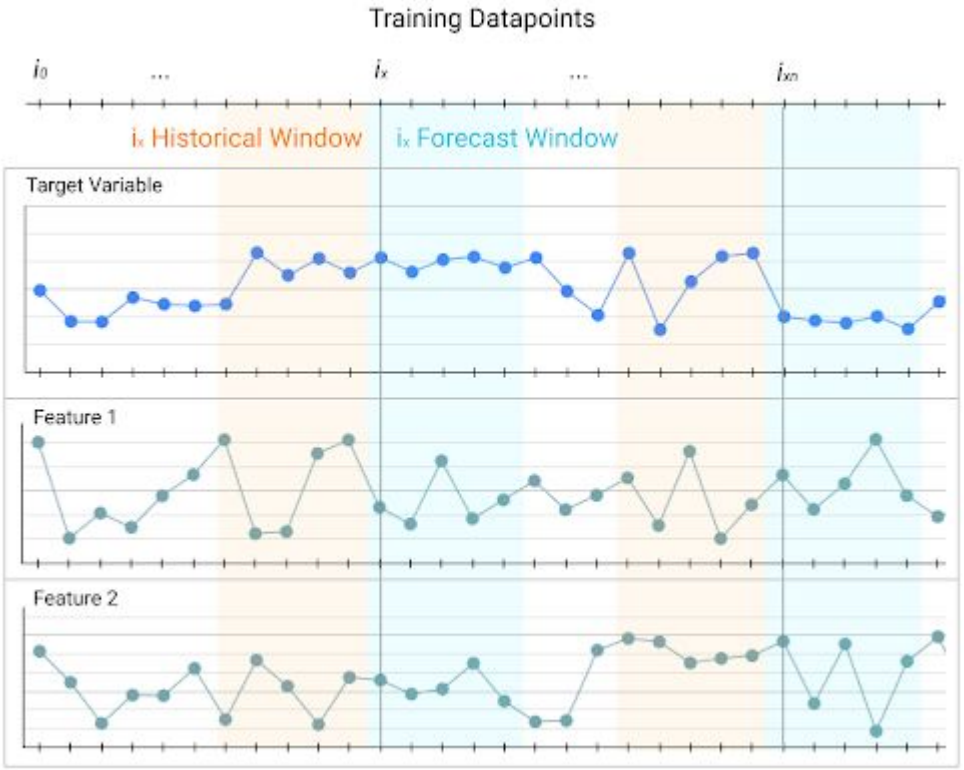
Vertex Forecast: その他のモデル設定

予測ホライズン
予測対象期間

コンテキスト期間
（各トレーニングデータポイントについて）
予測パターンを探すまでの期間

Optimization Objective
最適化に使用する指標
RMSE (default), MAE, RMSPE, Quantile

Budget
トレーニング予算上限



ユースケース: 需要予測

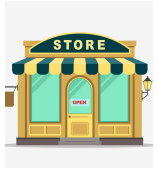
商品メタデータ

- ワンピース
- 模様
- 長袖 V ネック
- 55%リネン、綿 45%
- 洗濯機洗い可能

需要関連要素

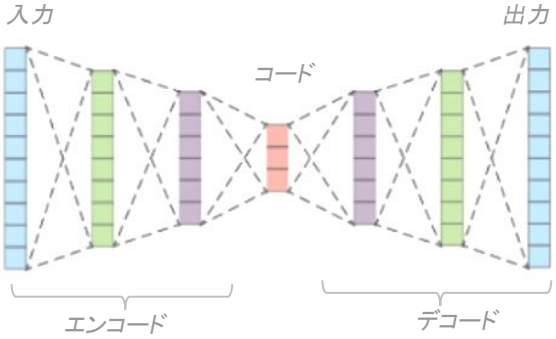
- 販売
- 価格
- 競合他社の価格
- プロモーション/イベント
- 休日

店舗データ



- 店舗の説明:大、小、専門
- 店の所在地
- アクセス

Vertex FORECAST



機械学習モデルが、製品間の関係、価格設定、競争、製品ライフサイクルの相乗効果などを考慮

長期的な予測区間
12-16 ヶ月

- 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

短期的な予測区間
0-8 週

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

ユースケース 2：宿泊施設における部屋の空き状況

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
- キングサイズのベッド

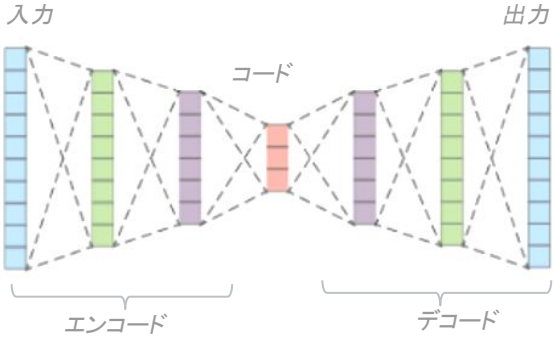
物件データ

- ロケーション
- 部屋のランク
- 国
- 空港の近さ
- サイトでの評価

需要換気

- 価格、プロモーション
- 競合の在庫
- 地域のイベント、祝日
- ウェブサイトの検索
- チケット販売

Vertex FORECAST



機械学習モデルで、競合他社の在庫、天気、地域のイベント、物件とタイプ間の関係など豊富なメタデータを考慮

中期的な予測区間 6-12 ヶ月

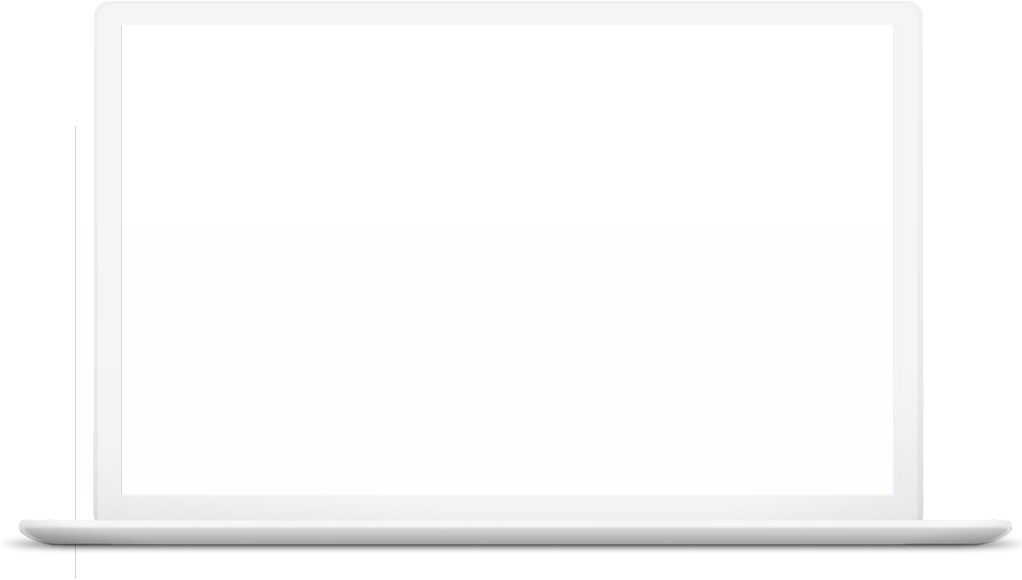
- 財務計画
- 資本配分
- 価格計画

短期的な予測区間 0-8 週

- プロモーション企画
- 人員配置

Demo

参考 : BigQuery ML (BQML)



1. BigQuery からデータを移動せずに機械学習モデル作成
2. 使い慣れた SQL で開発速度を向上
3. 一般的な ML タスクとハイパーパラメータチューニングを自動化

まとめ

03

Proprietary + Confidential

Google Cloud における時系列分析

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

Thank you

Google Cloud