

# Google.Professional-Machine-Learning-Engineer.v2024-06-17.q112

試験コード:	Professional-Machine-Learning-Engineer
試験名称:	Google Professional Machine Learning Engineer
認定資格:	Google
無料問題数:	112
バージョン:	v2024-06-17
アクセス数:	445
ページビュー数:	1120
<a href="https://www.jpnpdf.com/Google.Professional-Machine-Learning-Engineer.v2024-06-17.q112-mondaishu.html">https://www.jpnpdf.com/Google.Professional-Machine-Learning-Engineer.v2024-06-17.q112-mondaishu.html</a>	

## 最新問題: 1

機械学習スペシャリストは、Amazon Athena を使用して Amazon S3 上のデータセットにクエリを実行するプロセスを構築する必要があります。データセットには、プレーンテキストの CSV ファイルとして保存された 800,000 件を超えるレコードが含まれています。各レコードには次の内容が含まれます

200 列で、サイズは約 1.5 MB です。ほとんどのクエリは 5 ~ 10 列のみにまたがります。

機械学習スペシャリストは、クエリの実行時間を最小限に抑えるためにデータセットをどのように変換する必要がありますか？

- A. レコードを Apache Parquet 形式に変換します。
- B. レコードを JSON 形式に変換します。
- C. レコードを GZIP CSV 形式に変換します。
- D. レコードを XML 形式に変換します。

**Answer: A** ([メッセージを残す](#))

圧縮を使用すると、Amazon Athena によってスキャンされるデータの量が減り、S3 バケットのストレージも減ります。AWS の請求額にとっては Win-Win です。サポートされている形式:

GZIP、LZO、SNAPPY (Parquet)、および ZLIB。

参考: <https://www.cloudforecast.io/blog/using-parquet-on-athena-to-save-money-on-aws/>

## 最新問題: 2

You have trained a deep neural network model on Google Cloud. The model has low loss on the training data, but is performing worse on the validation data. You want the model to be resilient to overfitting. Which strategy should you use when retraining the model?

- A. Apply a dropout parameter of 0.2, and decrease the learning rate by a factor of 10
- B. Apply a L2 regularization parameter of 0.4, and decrease the learning rate by a factor of 10.

C. Run a hyperparameter tuning job on AI Platform to optimize for the L2 regularization and dropout parameters

D. AI プラットフォームでハイパーパラメータ調整ジョブを実行して学習率を最適化し、ニューロンの数を 2 倍に増やします。

**Answer: B (メッセージを残す)**

L2 正則化パラメーター 0.4 を適用し、学習率を 10 分の 1 に下げると、過学習を軽減し、モデルの復元力を高めることができます。出典: Google Cloud

最新問題: 3

あなたは、大規模なデータセットを使用して会社のカスタム言語モデルをトレーニングしています。Vertex AI で ReductionServer 戦略を使用する予定です。分散トレーニング ジョブのワーカープールを構成する必要があります。

あなたは何をするべきか？

A. 最初の 2 つのワーカー プールのマシンが GPU を備え、トレーニング コードが実行されるコンテナ イメージを使用するように構成します。3 番目のワーカー プールが GPU を持つように構成し、リダクション サーバー コンテナ イメージを使用します。

B. 最初の 2 つのワーカー プールのマシンが GPU を備え、トレーニング コードが実行されるコンテナ イメージを使用するように構成します。アクセラレータなしで、reductionserver コンテナ イメージを使用するように 3 番目のワーカー プールを構成し、帯域幅を優先するマシン タイプを選択します。

C. 最初の 2 つのワーカー プールのマシンが TPU を持ち、トレーニング コードが実行されるコンテナ イメージを使用するように構成します。アクセラレータなしで 3 番目のワーカー プールを構成し、アクセラレータなしで減らすサーバー コンテナ イメージを使用し、マシン タイプを選択します。帯域幅を優先します。

D. 最初の 2 つのプールのマシンが TPU を持つように構成します。トレーニング コードが実行されるコンテナ イメージを使用するには、TPU を持つように 3 番目のプールを構成し、reductionserver コンテナ イメージを使用します。

**Answer: B (メッセージを残す)**

Web 検索結果によると、Reduction Server は、Google で開発された高速な GPU all-reduce アルゴリズムで、専用のリデューサーのセットを使用してワーカーからの勾配を集約します

12。Reducer は、GPU VM よりも大幅に安価な軽量の CPU VM インスタンスです<sup>2</sup>。したがって、3 番目のワーカー プールにはアクセラレータを持たず、ワーカーとリデューサー間の通信を最適化するために、高いネットワーク帯域幅を持つマシン タイプを使用する必要があります

2。TPU は Reduction Server ではサポートされていないため、最初の 2 つのワーカー プールには GPU があり、トレーニング コード 12 を含むコンテナ イメージを使用する必要があります

す。Reduction-server コンテナ イメージは Google によって提供されており、3 番目のワーカー プールに使用する必要があります<sup>2</sup>。

最新問題: 4

オンライン予測のために Vertex AI にデプロイする予定の XGBoost モデルをトレーニングしました。現在、モデルを Vertex AI Model Registry にアップロードしています。オンライン予測リクエストを最小限の遅延で返す説明メソッドを構成する必要があります。また、モデルの特徴属性が時間の経過とともに大きく変化したときにアラートを受け取ることも必要です。あなたは何をすべきか？

**A.** 1 パス数 5 でサンプリングされた Shapley を説明方法として指定します。

2 モデルを Vertex AI Endpoints に展開します。

3. 監視目標として予測ドリフトを使用するモデル監視ジョブを作成します。

**B.** 1 パス数 5 で説明方法として統合勾配を指定します。

2 モデルを Vertex AI Endpoints に展開します。

3. 監視目標として予測ドリフトを使用するモデル監視ジョブを作成します。

**C.** 1. パス数 50 の説明メソッドとしてサンプリングされた Shapley を指定します。

2. モデルを Vertex AI Endpoints にデプロイします。

3. トレーニングとサービングのスキューを監視目標として使用するモデル監視ジョブを作成します。

**D.** 1 パス数 50 の説明方法として統合勾配を指定します。

2. モデルを Vertex AI Endpoints にデプロイします。

3 トレーニングとサービングのスキューを監視目標として使用するモデル監視ジョブを作成します。

**Answer: A (メッセージを残す)**

サンプリングされた Shapley は、Shapley 値の高速かつスケーラブルな近似値です。Shapley 値は、モデル予測に対する各特徴の寄与を測定するゲーム理論的な概念です。サンプリングされた Shapley は、最小限の遅延で特徴属性を返すことができるため、オンライン予測リクエストに適しています。パス カウント パラメータは、Shapley 値の推定に使用されるサンプルの数を制御します。値が低いほど、計算が高速になります。統合勾配は、ベースライン入力から実際の入力までのパスに沿った平均勾配を計算するもう 1 つの説明方法です。Integrated Gradients は、Sampled Shapley よりも正確ですが、計算量も多くなります。したがって、特にパス数が多いオンライン予測リクエストには推奨されません。予測ドリフトは、時間の経過に伴う特徴値またはラベルの分布の変化です。これはモデルのパフォーマンスと精度に影響を与える可能性があり、モデルの再トレーニングまたは再デプロイが必要になる場合があります。Vertex AI Model Monitoring を使用すると、デプロイされたモデルとエンドポイントの予測ドリフトを監視し、ドリフトが特定のしきい値を超えたときにアラートと通知を設定できます。通知を受信する電子メール アドレスを指定し、その情報を使用してトレーニング パイプラインを再トリガーし、モデルの更新バージョンをデプロイできます。これは、目標を達成するための最も直接的で便利な方法です。

トレーニングと提供のスキューは、モデルのトレーニングに使用されるデータとモデルの提供に使用されるデータの差です。また、モデルのパフォーマンスと精度に影響を与える可能性があり、データ品質の問題やモデルの古さを示している可能性があります。Vertex AI Model Monitoring を使用すると、デプロイされたモデルとエンドポイントのトレーニングと提供のスキューを監視し、スキューが特定のしきい値を超えた場合のアラートと通知を設定できます。

ただし、質問はデータ分布ではなくモデルの特徴属性に関するものであるため、これは質問とは関係ありません。参考文献:

※Vertex AI : 説明方法

\* Vertex AI: 設定の説明

\* Vertex AI: 予測ドリフトのモニタリング

\* Vertex AI: トレーニングとサービスのスキューを監視

#### 最新問題: 5

あなたは最近、自動運転車用の画像セグメンテーション モデルの最初のバージョンを構築しました。モデルをデプロイした後、曲線下面積 (AUC) メトリックの減少が観察されます。ビデオ録画を分析すると、トラフィックが非常に混雑している場合はモデルが失敗するが、トラフィックが少ない場合は期待どおりに機能することもわかります。この結果の最も考えられる理由は何ですか？

A. モデルは、交通量が少ないエリアでは過剰適合し、交通量が多いエリアでは過小適合します。

B. 出力ノードから入力ノードへの逆伝播中に勾配が小さくなり、消滅します。

C. AUC は、この分類モデルを評価するための正しい指標ではありません。

D. モデルのトレーニングに使用された混雑したエリアを表すデータが多すぎます。

**Answer: B** ([メッセージを残す](#))

#### 最新問題: 6

あなたの組織のコールセンターから、各通話における顧客の感情を分析するモデルを開発するよう依頼されました。コールセンターには毎日 100 万件を超える電話がかかり、データは Cloud Storage に保存されます。収集されたデータは通話が発信された地域から出てはならず、個人を特定できる情報 (PII) を保存または分析することはできません。データサイエンスチームは、SQL ANSI-2011 準拠のインターフェイスを必要とする視覚化とアクセス用のサードパーティ ツールを備えています。データ処理および分析用のコンポーネントを選択する必要があります。データパイプラインはどのように設計すべきでしょうか？

A. 1 = データフロー、2 = BigQuery

B. 1 = Pub/Sub、2 = データストア

C. 1 = Cloud Function、2 = Cloud SQL

D. 1 = データフロー、2 = Cloud SQL

**Answer: B** ([メッセージを残す](#))

#### 最新問題: 7

あなたは、世界中の何千もの企業に企業向け電子製品を販売する会社で働いています。あなたの会社は、過去の顧客データを BigQuery に保存しています。今後 3 年間の顧客生涯価値を予測するモデルを構築する必要があります。最も単純なアプローチを使用してモデルを構築し、視覚化ツールにアクセスしたいと考えています。あなたは何をするべきか？

**A.** Vertex AI Workbench ノートブックを作成して、探索的データ分析を実行します。IPython マジックを使用して、入力機能を備えた新しい BigQuery テーブルを作成します。BigQuery コンソールを使用して、モデル作成ステートメントを実行します。ml を使用して結果を検証します。評価して ml します。発言を予測する。

**B.** BigQuery コンソールから create model ステートメントを実行して AutoML モデルを作成します。ml を使用して結果を検証します。評価して ml します。発言を予測する。

**C.** Vertex AI Workbench ノートブックを作成して、探索的データ分析を実行し、入力フィーチャを作成します。フィーチャを CSV ファイルとして Cloud Storage に保存します。CSV ファイルを新しい BigQuery テーブルとしてインポートします。BigQuery コンソールを使用して、モデル作成ステートメントを実行します。ml を使用して結果を検証します。評価して ml します。発言を予測する。

**D.** Vertex AI Workbench ノートブックを作成して探索的データ分析を実行します。IPython マジックを使用して、入力特徴を持つ新しい BigQuery テーブルを作成し、モデルを作成し、モデル作成 ml を使用して結果を検証します。と ml を評価します。発言を予測する。

**Answer: B** ([メッセージを残す](#))

BigQuery は、スケーラブルでコスト効率の高い方法で大量のデータの保存とクエリを可能にするサービスです。BigQuery を使用すると、create model ステートメントを使用して、今後 3 年間の顧客生涯価値を予測するモデルを構築できます。create model ステートメントは、BigQuery のデータを使用して ML モデルを作成およびトレーニングできる SQL コマンドです。create model ステートメントを使用して、データに最適な機能とアーキテクチャを自動的に選択するモデルの一種である AutoML モデルを作成できます。AutoML モデルを使用すると、コードを記述したり特徴量エンジニアリングを実行したりすることなく、最も簡単なアプローチでモデルを構築できます。ml.evaluate ステートメントと ml.predict ステートメントを使用してモデルの結果を検証することもできます。ml.evaluate ステートメントは、さまざまなメトリックを使用してモデルのパフォーマンスと品質を評価できる SQL コマンドです。ml.predict ステートメントは、モデルと新しいデータを使用して予測を行うことができる SQL コマンドです。BigQuery コンソールを使用してチャートやグラフなどの視覚化ツールにアクセスし、データやモデルの結果を調査および分析することもできます。BigQuery コンソール、create model ステートメント、ml.evaluate ステートメントと ml.predict ステートメントを使用すると、今後 3 年間の顧客ライフタイム バリューを予測するモデルを構築して検証し、可視化ツールにアクセスできます。参考文献:

- \* BigQuery ドキュメント
- \* モデルステートメントのドキュメントを作成する
- \* ml.evaluate ステートメントのドキュメント
- \* ml.predict ステートメントのドキュメント
- \* Google Cloud 認定資格の準備: 機械学習エンジニア プロフェッショナル認定資格

**最新問題: 8**

あなたは配送会社に勤めています。配達された荷物やトラックの位置などの機能を長期にわたって保存および管理するシステムを設計する必要があります。システムは、低遅延で特徴を取得し、

それらの特徴をオンライン予測用のモデルにフィードする必要があります。データサイエンスチームは、モデルのトレーニングのために特定の時点の履歴データを取得します。最小限の労力で特徴を保存したいと考えています。あなたは何をするべきか？

- A. 特徴量をキー/値データとして Bigtable に保存します。
- B. フィーチャを Vertex AI フィーチャストアに保存します。
- C. フィーチャを Vertex AI データセットとして保存し、それらのフィーチャを使用して Vertex AI エンドポイントでホストされているモデルをトラムします。
- D. BigQuery タイムスタンプパーティション分割テーブルに機能を保存し、BigQuery Storage Read API を使用して機能を提供します。

**Answer: B** ([メッセージを残す](#))

Vertex AI Feature Store は、ML 機能を Google Cloud に保存して管理できるサービスです。

Vertex AI Feature Store を使用すると、配達された荷物やトラックの位置などのフィーチャを長期にわたって保存し、オンライン予測のために低レイテンシで取得できます。オンライン予測は、入力データの個別または小さなバッチに対して低遅延の応答を提供する予測の一種です。Vertex AI Feature Store を使用して、モデルのトレーニングのために特定の時点の履歴データを取得することもできます。モデルのトレーニングは、データから ML モデルのパラメーターを学習するプロセスです。Vertex AI Feature Store を使用すると、最小限の労力でフィーチャを保存でき、独自のデータストレージとサービスシステムの管理の複雑さを回避できます。参考文献:

\* Vertex AI フィーチャーストアのドキュメント

\* Google Cloud 認定資格の準備: 機械学習エンジニア プロフェッショナル認定資格

#### 最新問題: 9

あなたは、会社の ML 実験とモデルの再トレーニングを自動化する MLOps プラットフォームを構築しています。数十のパイプラインのアーティファクトを整理する必要があるパイプラインのアーティファクトはどのように保存すればよいですか？

- A. パラメータを Cloud SQL に保存し、モデルのソースコードとバイナリを GitHub に保存します。
- B. パラメータを Cloud SQL に保存し、モデルのソースコードを GitHub に保存し、モデルのバイナリを Cloud Storage に保存します。
- C. Vertex ML メタデータにパラメータを保存します。モデルのソースコードは GitHub に保存され、モデルのバイナリは Cloud Storage に保存されます。
- D. パラメーターを Vertex ML メタデータに保存し、モデルのソースコードとバイナリを GitHub に保存します。

**Answer: (解答を表示する)**

数十のパイプラインのアーティファクトを整理するには、パラメータを Vertex ML Metadata に保存し、モデルのソースコードを GitHub に保存し、モデルのバイナリを Cloud Storage に保存する必要があります。このオプションには次の利点があります。

\* Vertex ML Metadata は、データセット、モデル、メトリクス、パラメーターなどの ML ワークフローのメタデータの追跡と管理に役立つサービスです<sup>1</sup>。また、データリネージ、モデルのバージョン管理、モデルのパフォーマンスの監視にも役立ちます<sup>2</sup>。

\* GitHub は、コードリポジトリをホスティングおよび共同作業するための人気のあるプラットフォームです。これは、モデルのソースコードだけでなく、ML パイプラインの一部である構成ファイル、スクリプト、ノートブックの管理にも役立ちます<sup>3</sup>。

\* Cloud Storage は、モデルバイナリ<sup>4</sup>を含むあらゆる種類のデータを保存できる、スケーラブルで耐久性のあるオブジェクトストレージサービスです。また、Vertex AI、Cloud Functions、Cloud Run などの他のサービスと統合して、モデルのデプロイと提供を容易にすることもできます<sup>5</sup>。

参考文献:

\*1: Vertex ML メタデータの概要 | 頂点 AI | グーグルクラウド

\* 2: ML ワークフローのメタデータを管理 | 頂点 AI | グーグルクラウド

\*3: GitHub - 世界中がソフトウェアを構築する場所

※4 :クラウドストレージ | グーグルクラウド

\* 5: モデルの展開 | 頂点 AI | グーグルクラウド

#### 最新問題: 10

あなたは、本番環境でカスタム モデルをトレーニングおよび実行するためのプロセスを開発しています。モデルと予測のシステムを表示できる必要があります。あなたは何をするべきか？

**A.** 1 Vertex AI 管理データセットを作成します

2 Vertex AI トレーニング パイプラインを使用してモデルをトレーニングします

3 Vertex AI でバッチ予測を生成する

**B.** 1 Vertex AI Pipelines カスタム トレーニング ジョブ コンポーネントを使用してモデルをトレーニングします

2. Vertex AI Pipelines モデルのバッチ予測コンポーネントを使用して予測を生成する

**C.** 1 データセットを BigQuery にアップロードします

2. Vertex AI カスタム トレーニング ジョブを使用してモデルをトレーニングします

3 Vertex AI SDK カスタム予測ルーチンを使用して予測を生成する

**D.** 1 Vertex AI Experiments を使用してモデルをトレーニングします。

2 Vertex AI Model Registry にモデルを登録します。

3. Vertex AI でバッチ予測を生成する

**Answer: D (メッセージを残す)**

公式試験ガイド<sup>1</sup>によると、試験で評価されるスキルの 1 つは「パイプラインアーティファクトのシステムを追跡する」ことです。Vertex AI Experiments<sup>2</sup> は、モデル トレーニングの実行結果を追跡および比較できるサービスです。Vertex AI Experiments は、トレーニング実行ごとにハイパーパラメータ、メトリクス、アーティファクトなどのメタデータを自動的に記録します。Vertex AI Experiments を使用すると、TensorFlow、PyTorch、XGBoost、または scikit-learn を使用してカスタム モデルをトレーニングできます。Vertex AI Model Registry<sup>3</sup> は、トレーニングされたモデルを一元管理できるサービスです。Vertex AI Model Registry を使用して、モデルを登録し、ラベル

と説明を追加し、モデルの系統グラフを表示できます。リネージグラフには、データセット、トレーニングパイプライン、評価メトリクスなど、モデル作成の一部であるアーティファクトと実行が表示されます。他のオプションは、このシナリオには関係がなく、最適でもありません。参考文献:

- \* Professional ML Engineer 試験ガイド
- \* Vertex AI 実験
- \* Vertex AI モデル レジストリ
- \* Google プロフェッショナル機械学習認定試験 2023
- \* 最新の Google Professional Machine Learning Engineer の実際の無料試験問題

#### 最新問題: 11

あなたは、Google Kubernetes Engine で Kubeflow パイプラインを開発しています。パイプラインの最初のステップは、BigQuery に対してクエリを発行することです。そのクエリの結果をパイプラインの次のステップへの入力として使用する予定です。これをできるだけ簡単な方法で達成したいと考えています。あなたは何をするべきか？

- A. BigQuery コンソールを使用してクエリを実行し、クエリ結果を新しい BigQuery テーブルに保存します。
- B. BigQuery API を使用して BigQuery に対してクエリを実行する Python スクリプトを作成します。このスクリプトを Kubeflow パイプラインの最初のステップとして実行します。
- C. Kubeflow Pipelines ドメイン固有言語を使用して、Python BigQuery クライアント ライブラリを使用してクエリを実行するカスタム コンポーネントを作成します。
- D. GitHub で Kubeflow Pipelines リポジトリを見つけます。BigQuery クエリ コンポーネントを見つけて、そのコンポーネントの URL をコピーし、それを使用してコンポーネントをパイプラインに読み込みます。コンポーネントを使用して BigQuery に対してクエリを実行する

**Answer: D (メッセージを残す)**

Kubeflow は、Kubernetes 上でスケーラブルでポータブルな機械学習ワークフローを開発、オーケストレーション、デプロイ、実行するためのオープンソース プラットフォームです。Kubeflow Pipelines は、グラフィカル ユーザー インターフェイスまたは Python ベースのドメイン固有言語 (DSL) を使用してエンドツーエンドの機械学習パイプラインを構築および管理できる Kubeflow のコンポーネントです。Kubeflow Pipelines は、機械学習ワークフローの自動化とオーケストレーションを支援し、さまざまな Google Cloud サービスやツールと統合できます<sup>1</sup>。Kubeflow Pipelines で使用できる Google Cloud サービスの 1 つは、サーバーレスでスケーラブルでコスト効率の高いデータである BigQuery です。大規模なデータに対して高速かつ複雑なクエリを実行できるウェアハウス。BigQuery は、機械学習用のデータの分析と準備、機械学習モデルの保存と管理に役立ちます<sup>2</sup>。Kubeflow パイプラインの最初のステップとして BigQuery に対してクエリを実行し、そのクエリの結果を次のステップへの入力として使用します。パイプラインで行う最も簡単な方法は、BigQuery クエリ コンポーネントを使用することです。これは、GitHub の Kubeflow Pipelines リポジトリにある事前に構築されたコンポーネントです。

The BigQuery Query Component allows you to run a SQL query on BigQuery, and output the results as a table or a file. You can use the component's URL to load the component into your pipeline, and specify the query and the output parameters. You can then use the output of the component as the input to the next step in your pipeline, such as a data processing or a model training step<sup>3</sup> The other options are not as easy or feasible. Using the BigQuery console to execute your query and then save the query results into a new BigQuery table is not a good idea, as it does not integrate with your Kubeflow pipeline, and requires manual intervention and duplication of data. Writing a Python script that uses the BigQuery API to execute queries against BigQuery is not ideal, as it requires writing custom code and handling authentication and error handling. Using the Kubeflow Pipelines DSL to create a custom component that uses the Python BigQuery client library to execute queries is not optimal, as it requires creating and packaging a Docker container image for the component, and testing and debugging the component.

References: 1: Kubeflow Pipelines overview 2: BigQuery overview 3: BigQuery Query Component

### 最新問題: 12

あなたは銀行に勤めており、不正行為検出のためのランダム フォレスト モデルを構築しています。トランザクションを含むデータセットがあり、そのうち 1% が不正であると特定されています。どのデータ変換戦略が分類器のパフォーマンスを向上させる可能性がありますか？

- A. データを TFRecord に書き込みます。
- B. すべての数値特徴を Z 正規化します。
- C. 不正なトランザクションを 10 回オーバーサンプリングします。
- D. すべてのカテゴリ特徴量でワンホット エンコーディングを使用します。

**Answer: C** ([メッセージを残す](#))

オーバーサンプリングは、多数派のクラスが少数派のクラスを支配する不均衡なデータセットを処理するための手法です。少数派クラスのサンプル数を増やすことで、クラスの分布のバランスをとります。

オーバーサンプリングにより、多数派のクラスに対するバイアスが軽減され、少数派のクラスに対する感度が高まるため、分類器のパフォーマンスを向上させることができます。

この場合、データセットにはトランザクションが含まれており、そのうち 1% が不正であると特定されています。これは、不正取引が少数派であり、非不正取引が多数派であることを意味します。このデータセットでトレーニングされたランダム フォレスト モデルは、不正なトランザクションの再現率が低い可能性があります。つまり、トランザクションの多くを見逃して、不正を検出できない可能性があります。これは銀行とその顧客にとって高額なコストとなる可能性があります。

この問題を克服する 1 つの方法は、不正なトランザクションを 10 回オーバーサンプリングすることです。これは、各不正なトランザクションがトレーニング データセット内で 10 回複製されることを意味します。これにより、不正取引の割合が 1% から約 10% に増加し、データセットのバランスがより整います。これにより、ランダム フォレスト モデルは、不正なトランザクションと不正でないトランザクションを区別するパターンと特徴をより認識できるようになり、少数派のクラスの精度と再現率が向上します。

オーバーサンプリングや不均衡なデータに対するその他の手法の詳細については、次の参考資料を参照してください。

- \* 不均衡な分類のためのランダムなオーバーサンプリングとアンダーサンプリング
- \* 不均衡なデータセットに対するオーバーサンプリング手法の探求

### 最新問題: 13

あなたは、会社がよりターゲットを絞ったオンライン広告キャンペーンを作成できるようにするモデルを開発しています。モデルのトレーニングに使用するデータセットを作成する必要があります。モデル内に不当なバイアスが生じたり、強化されたりすることは避けたいと考えています。あなたは何をすべきか？

2つの答えを選択してください

- A. 人口統計上の特徴の包括的なセットが含まれます。
- B. 広告と最も頻繁にやり取りする人口統計グループのみを含めます。
- C. 実稼働トラフィックのランダムなサンプルを収集して、トレーニング データセットを構築します。
- D. 実稼働トラフィックの層別サンプルを収集して、トレーニング データセットを構築します。
- E. トレーニングされたモデルで、機密性の高いカテゴリと人口統計にわたる公平性テストを実施します。

**Answer:** ([解答を表示する](#))

モデル内で不公平なバイアスが生じたり強化されたりすることを回避するには、運用トラフィックの代表的なサンプルを収集してトレーニング データセットを構築し、トレーニングされたモデルで機密性の高いカテゴリと人口統計全体にわたる公平性テストを実施する必要があります。代表的なサンプルとは、母集団の真の分布を反映するサンプルであり、どのグループも過大または過小代表するものではありません。ランダム サンプルは、すべてのデータ ポイントが等しい確率で選択されるようにするため、代表的なサンプルを取得する簡単な方法です。層別サンプルは、すべてのサブグループがサンプル内で比例代表を持つことを保証するため、代表的なサンプルを取得するもう 1 つの方法です。ただし、層別サンプルにはサブグループとそのサイズに関する事前の知識が必要ですが、それが利用できないか、簡単に入手できない場合があります。したがって、この場合はランダムサンプルの方がより現実的な選択肢となります。公平性テストは、年齢、性別、人種などのさまざまなカテゴリや人口統計に基づいて、モデルの潜在的な偏見と差別を測定および評価する方法です。公平性テストは、不公平な結果や影響を特定し、軽減するのに役立ちます。モデルがすべてのグループを公正かつ公平に扱うことを保証します。公平性テストは、混同行列、ROC 曲線、公平性指標などのさまざまな方法やツールを使用して実行できます。参考: 答えは、Google Cloud の公式ドキュメントと、データ サンプリングと公平性テストに関連するリソースから確認できます。

- \* サンプリングデータ | BigQuery
- \* 公平性指標 | TensorFlow
- \* What-if ツール | TensorFlow

### 最新問題: 14

あなたは、カスタマー サポート チケットをサポート エージェントにルーティングする前に有益なメタデータで強化するサーバーレス ML システムを使用したアーキテクチャを設計しています。チケットの優先順位を予測し、チケットの解決時間を予測し、エージェントがサポート リクエストを処理する際に戦略的な決定を下せるようにセンチメント分析を実行するには、一連のモデルが必要です。チケットにはドメイン固有の用語や専門用語が含まれることは期待されません。提案されたアーキテクチャには次のようなフローがあります。

Enrichment Cloud Functions はどのエンドポイントを呼び出す必要がありますか？

- A. 1 = AI プラットフォーム、2 = AI プラットフォーム、3 = AutoML ビジョン
- B. 1 = AI プラットフォーム、2 = AI プラットフォーム、3 = AutoML Natural Language
- C. 1 = AI プラットフォーム、2 = AI プラットフォーム、3 = クラウド自然言語 API
- D. 1 = クラウド自然言語 API、2 = AI プラットフォーム、3 = クラウド ビジョン API

**Answer: C (メッセージを残す)**

<https://cloud.google.com/architecture/architecture-of-a-serverless-ml-model#architecture> アーキテクチャには次のようなフローがあります。

ユーザーが Firebase にチケットを書き込むと、Cloud Function がトリガーされます。

- Cloud Function は、チケットを強化するために 3 つの異なるエンドポイントを呼び出します。
- AI Platform エンドポイント。関数は優先度を予測できます。
- AI Platform エンドポイント。この関数は解決時間を予測できます。
- 感情分析と単語の顕著性を行うための Natural Language API。
- For each reply, the Cloud Function updates the Firebase real-time database.
- The Cloud Function then creates a ticket into the helpdesk platform using the RESTful API.

#### 最新問題: 15

会社のモバイル アプリケーションからのユーザー アクティビティ データを分析する必要があります。チームは、データ分析、変換、ML アルゴリズムの実験に BigQuery を使用します。ユーザー アクティビティ データを確実にリアルタイムで BigQuery に取り込む必要があります。あなたは何をするべきか？

- A. BigQuery にデータをストリーミングするように Pub/Sub を構成します。
- B. Dataproc で Apache Spark ストリーミング ジョブを実行して、データを BigQuery に取り込みます。
- C. Dataflow ストリーミング ジョブを実行して、データを BigQuery に取り込みます。
- D. データを BigQuery に取り込むように Pub/Sub と Dataflow ストリーミング ジョブを構成します。

**Answer: C (メッセージを残す)**

ユーザー アクティビティ データを確実にリアルタイムで BigQuery に取り込むための最良のオプションは、Dataflow ストリーミング ジョブを実行してデータを BigQuery に取り込むことです。Dataflow は、データのバッチ処理とストリーム処理の両方を処理できるフルマネージド サービスで、BigQuery や他の Google Cloud サービスとシームレスに統合できます。

Dataflow では、Apache Beam をプログラミング モデルとして使用することもできます。これにより、データ パイプラインを開発するための統合された移植可能な API が提供されます。Dataflow を使用すると、独自のインフラストラクチャ管理の複雑さとオーバーヘッドを回避し、データのロジックと変換に集中できます。データフローは、構造化データ、非構造化データ、バイナリ データなどのさまざまな種類のデータを処理することもでき、データ ストリームに対してウィンドウ処理、集計、その他の操作を適用することもできます。

他のオプションは、次の理由により最適ではありません。

\* A. BigQuery にデータをストリーミングするように Pub/Sub を構成することは、良い選択肢ではありません。Pub/Sub は、データ ストリームのパブリッシュとサブスクライブができるメッセージング サービスですが、データの変換や処理を実行できないからです。Pub/Sub は Dataflow のソースまたはシンクとして使用できますが、BigQuery にデータを取り込むためのスタンドアロン ソリューションとしては使用できません。

\* B. Dataproc で Apache Spark ストリーミング ジョブを実行して BigQuery にデータを取り込むことは、適切なオプションではありません。これには、仮想マシンの独自のクラスタを設定して管理する必要があり、ソリューションのコストと複雑さが増大する可能性があります。さらに、Apache Spark は BigQuery とネイティブに統合されていないため、BigQuery にデータを書き込むためにコネクタまたは中間ストレージを使用する必要があるため、レイテンシや非効率が発生する可能性があります。

\* D. データを BigQuery に取り込むように Pub/Sub と Dataflow ストリーミング ジョブを構成することは悪い選択肢ではありませんが、Dataflow は Pub/Sub を仲介者として使用せずにモバイルアプリケーションから直接データを読み取ることができるため、必須ではありません。Pub/Sub を使用すると、抽象化と信頼性の追加レイヤーを追加できますが、ソリューションのコストと複雑さが増加し、データの取り込みに多少の遅延が生じる可能性もあります。

参考文献:

\* Professional ML Engineer 試験ガイド

\* Google Cloud 認定資格の準備: 機械学習エンジニア プロフェッショナル認定資格

\* Google Cloud が機械学習エンジニア認定資格を開始

\* データフローのドキュメント

\* BigQuery ドキュメント

### 最新問題: 16

あなたは、コンタクト センターを最新化したいと考えている大規模なテクノロジー企業に勤めています。あなたは、リクエストをより迅速に適切なサポート チームに転送できるように、着信コールを製品ごとに分類するソリューションを開発するように依頼されました。Speech-to-Text API を使用して通話をすでに文字に起こしています。データの前処理と開発時間を最小限に抑えたいと考えています。どのようにモデルを構築すればよいでしょうか？

A. カスタム モデルを構築して、文字起こしされたコールから製品キーワードを特定し、分類アルゴリズムを通じてキーワードを実行します。

B. Cloud Natural Language API を使用して、分類用のカスタム エンティティを抽出します

- C. AI Platform Training 組み込みアルゴリズムを使用してカスタム モデルを作成します  
D. AutoML Natural Language を使用して、分類用のカスタム エンティティを抽出します  
**Answer:** ([解答を表示する](#))

有効な **Professional-Machine-Learning-Engineer** 問題集は GoShiken.com が提供された合格しやすい Professional-Machine-Learning-Engineer 試験問題集！ GoShiken.com が最新の **Professional-Machine-Learning-Engineer** 試験問題集を提供しています。GoShiken.com Professional-Machine-Learning-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Professional-Machine-Learning-Engineer 問題集をゲットする人はこちら：  
<https://www.goshiken.com/Google/Professional-Machine-Learning-Engineer-mondaishu.html>  
(**29030%OFF**問題集溶と正解付きで 30%w 特別割引コード: **Freepdfdumps**)

**最新問題: 17**

複数の Google Cloud プロジェクトにわたるデータを使用して、Vertex AI でモデルをトレーニングしています。モデルのさまざまなバージョンのパフォーマンスを追跡し、比較する必要があります。ML ワークフローにどの Google Cloud サービスを含めるべきですか？

- A. Vertex AI パイプライン、Vertex AI Experiments および Vertex AI メタデータ
- B. Vertex AI Pipelines、Vertex AI Feature Store、および Vertex AI Experiments
- C. データプレックス、Vertex AI Feature Store と Vertex AI TensorBoard
- D. データプレックス、Vertex AI 実験と Vertex AI ML メタデータ

**Answer:** A ([メッセージを残す](#))

**最新問題: 18**

Vertex AI でモデル トレーニング パイプラインを実行しているときに、メモリ不足エラーにより評価ステップが失敗していることがわかりました。現在、評価ステップでは TensorFlow Model Analysis (TFMA) を標準の Evaluator TensorFlow Extended (TFX) パイプライン コンポーネントとともに使用しています。インフラストラクチャのオーバーヘッドを最小限に抑えながら、評価品質を低下させることなくパイプラインを安定させたいと考えています。あなたは何をするべきか？

- A. tfma.MetricsSpec () を追加して、評価ステップのメトリクス数を制限します。
- B. パイプラインを Google Kubernetes Engine でホストされている Kubeflow に移行し、評価ステップに適切なノード パラメーターを指定します。
- C. Dataflow で評価ステップを実行するには、beam\_pipeline\_args にフラグ -runner=DataflowRunner を含めます。
- D. 評価ステップをパイプラインから移動し、十分なメモリを備えたカスタム Compute Engine VM で実行します。

**Answer:** ([解答を表示する](#))

インフラストラクチャのオーバーヘッドを最小限に抑えながら、評価の品質を低下させることなくパイプラインを安定させるための最良のオプションは、評価ステップのランナーとして Dataflow を使用することです。Dataflow は、ワークロードに応じてスケールアップおよびスケールダウンできる Apache Beam パイプラインを実行するためのフルマネージド サービスです。Dataflow は、モデル評価などの大規模な分散データ処理タスクを処理でき、Vertex AI Pipelines や TensorFlow Extended (TFX) と統合することもできます。beam\_pipeline\_args でフラグ -runner=DataflowRunner を使用すると、ローカルで実行されメモリ不足エラーが発生する可能性があるデフォルトの DirectRunner を使用する代わりに、Dataflow で評価ステップを実行するように Evaluator コンポーネントに指示できます。オプション A は不正解です。評価ステップでのメトリクス数を制限するために tfma.MetricsSpec() を追加すると、一部の重要なメトリクスが省略される可能性があるため、評価の品質が低下する可能性があります。

さらに、データとモデルのサイズと複雑さによっては、評価ステップで依然として大量のメモリが消費される可能性があるため、メトリクス数を減らしてもメモリ不足エラーが解決されない可能性があります。オプション B は不正解です。Google Kubernetes Engine (GKE) でホストされている Kubeflow にパイプラインを移行すると、GKE クラスタを自分でプロビジョニング、管理、監視する必要があるため、インフラストラクチャのオーバーヘッドが増加する可能性があります。

さらに、評価ステップでは適切なノードパラメーターを指定する必要があり、最適な構成を見つけるには試行錯誤が必要になる場合があります。オプション D は不正解です。評価ステップをパイプラインの外に移動してカスタム Compute Engine VM 上で実行すると、VM を自分で作成、構成、削除する必要があるため、インフラストラクチャのオーバーヘッドが増加する可能性があります。さらに、評価ステップのために十分なメモリが VM にあることを確認する必要があり、最適なマシンタイプを見つけるには試行錯誤が必要になる場合があります。参考文献:

- \* データフローのドキュメント
- \* DataflowRunner の使用
- \* 評価コンポーネントのドキュメント
- \* Evaluator コンポーネントの構成

### 最新問題: 19

You work for a company that manages a ticketing platform for a large chain of cinemas. Customers use a mobile app to search for movies they're interested in and purchase tickets in the app. Ticket purchase requests are sent to Pub/Sub and are processed with a Dataflow streaming pipeline configured to conduct the following steps:

1. Check for availability of the movie tickets at the selected cinema.
2. Assign the ticket price and accept payment.
3. Reserve the tickets at the selected cinema.
4. Send successful purchases to your database.

Each step in this process has low latency requirements (less than 50 milliseconds). You have developed a logistic regression model with BigQuery ML that predicts whether offering a promo code for free popcorn increases the chance of a ticket purchase, and this prediction should be

added to the ticket purchase process. You want to identify the simplest way to deploy this model to production while adding minimal latency. What should you do?

- A. Run batch inference with BigQuery ML every five minutes on each new set of tickets issued.
- B. Export your model in TensorFlow format, deploy it on Vertex AI, and query the prediction endpoint from your streaming pipeline.
- C. Export your model in TensorFlow format, and add a `tfx_bsl.public.beam.RunInference` step to the Dataflow pipeline.
- D. Convert your model with TensorFlow Lite (TFLite), and add it to the mobile app so that the promo code and the incoming request arrive together in Pub/Sub.

**Answer: A** ([メッセージを残す](#))

#### 最新問題: 20

機械学習スペシャリストは、企業 VPC のプライベートサブネットに Amazon SageMaker ノートブック インスタンスを使用しています。ML スペシャリストは、Amazon SageMaker ノートブック インスタンスの Amazon EBS ボリュームに重要なデータを保存しているため、その EBS ボリュームのスナップショットを取得する必要があります。ただし、ML スペシャリストは、VPC 内で Amazon SageMaker ノートブック インスタンスの EBS ボリュームまたは Amazon EC2 インスタンスを見つけることができません。

ML スペシャリストが VPC にインスタンスを表示できないのはなぜですか？

- A. Amazon SageMaker ノートブック インスタンスは顧客アカウント内の EC2 インスタンスに基づいていますが、VPC の外部で実行されます。
- B. Amazon SageMaker ノートブック インスタンスは、顧客アカウント内の Amazon ECS サービスに基づいています。
- C. Amazon SageMaker ノートブック インスタンスは、AWS サービス アカウント内で実行される EC2 インスタンスに基づいています。
- D. Amazon SageMaker ノートブック インスタンスは、AWS サービス アカウント内で実行される AWS ECS インスタンスに基づいています。

**Answer: C** ([メッセージを残す](#))

説明/参照: <https://docs.aws.amazon.com/sagemaker/latest/dg/gs-setup-working-env.html>

#### 最新問題: 21

あなたは、ビデオ フィードからスライスされたフレームを使用し、特定のオブジェクトの周囲に境界ボックスを作成する ML モデルを開発しています。トレーニング パイプラインの次のステップを自動化したいと考えています。Cloud Storage でのデータの取り込みと前処理、続いて Vertex AI ジョブを使用したオブジェクト モデルのトレーニングとハイパーパラメータ調整、最後にモデルのエンドポイントへのデプロイです。最小限のクラスター管理でパイプライン全体を調整したいと考えています。どのようなアプローチを使用する必要がありますか？

- A. Google Kubernetes Engine で Kubeflow Pipelines を使用します。
- B. TensorFlow Extended (TFX) SDK で Vertex AI Pipelines を使用します。
- C. Kubeflow Pipelines SDK で Vertex AI Pipelines を使用します。

D. オーケストレーションに Cloud Composer を使用します。

**Answer: B (メッセージを残す)**

\* オプション A は不正解です。Google Kubernetes Engine で Kubeflow Pipelines を使用することは、最小限のクラスター管理でパイプライン全体をオーケストレーションする最も便利な方法ではありません。Kubeflow Pipelines は、コンテナを使用して ML パイプラインを構築、実行、管理できるオープンソース プラットフォームです<sup>1</sup>。Google Kubernetes Engine は、コンテナ化されたアプリケーションをオーケストレーションするためのオープンソース システムである Kubernetes を実行する仮想マシンのクラスターを作成および管理できるサービスです<sup>2</sup>。ただし、このオプションにはクラスターの作成と構成、Kubeflow Pipelines のインストールと保守、パイプラインコードの作成と実行が含まれるため、オプション B よりも多くの労力とリソースが必要になります。

\* オプション B は正解です。TensorFlow Extended (TFX) SDK で Vertex AI Pipelines を使用することが、最小限のクラスター管理でパイプライン全体をオーケストレーションする最良の方法であるためです。Vertex AI Pipelines は、Google Cloud<sup>3</sup> 上でスケラブルでポータブルな ML パイプラインを作成して実行できるサービスです。TensorFlow Extended (TFX) は、TensorFlow<sup>4</sup> を使用して本番環境に対応した ML パイプラインを構築するためのコンポーネントとライブラリのセットを提供するフレームワークです。Vertex AI Pipelines と TFX SDK を使用すると、Cloud Storage 内のデータの取り込みと前処理、Vertex AI ジョブを使用したオブジェクトモデルのトレーニングと調整、事前定義またはカスタム コンポーネントを使用したモデルのエンドポイントへのデプロイが可能になります。Vertex AI Pipelines は基盤となるインフラストラクチャとオーケストレーションを処理するため、クラスターの管理やスケラビリティについて心配する必要はありません。

\* オプション C は不正解です。Vertex AI Pipelines を Kubeflow Pipelines SDK とともに使用することは、最小限のクラスター管理でパイプライン全体をオーケストレーションする最適な方法ではないためです。Kubeflow Pipelines SDK は、Kubeflow Pipelines<sup>5</sup> を使用して ML パイプラインを構築および実行できるようにするライブラリです。Vertex AI Pipelines と Kubeflow Pipelines SDK を使用すると、コンテナを使用して Google Cloud 上で ML パイプラインを作成して実行できます。ただし、このオプションでは、トレーニングとデプロイには Vertex AI SDK、取り込みと前処理には Kubeflow Pipelines SDK など、パイプラインのさまざまなステップで異なる API とツールを使用する必要があるため、オプション B よりも利便性と一貫性が劣ります。さらに、このオプションは、標準コンポーネント、メタデータ ストア、ML メタデータ ライブラリなどの TFX の利点を活用しません。

\* オーケストレーションに Cloud Composer を使用することは、最小限のクラスター管理でパイプライン全体をオーケストレーションする最も効率的な方法ではないため、オプション D は不正解です。Cloud Composer は、オープンソース プラットフォームである Apache Airflow を使用してワークフローを作成および実行できるサービスです。

\* 複雑なタスクを調整します。Cloud Composer を使用すると、タスクの依存関係と順序を定義する DAG (有向非循環グラフ) を作成および管理することで、パイプライン全体をオーケストレーションできます。

ただし、このオプションには環境の作成と構成、Airflow のインストールと保守、DAG の作成と実行が含まれるため、オプション B よりも複雑でコストがかかります。

参考文献:

- \* Kubeflow パイプラインのドキュメント
- \* Google Kubernetes Engine ドキュメント
- \* Vertex AI パイプラインのドキュメント
- \* TensorFlow Extended ドキュメント
- \* Kubeflow Pipelines SDK ドキュメント
- \* [Cloud Composer ドキュメント]
- \* [Vertex AI ドキュメント]
- \* [クラウドストレージのドキュメント]
- \* [TensorFlow ドキュメント]

#### 最新問題: 22

ある企業は、Amazon Polly を使用して平文ドキュメントを音声に変換し、社内アナウンスを自動化しています。ただし、現在の文書では会社の頭字語が誤って発音されています。

機械学習スペシャリストは今後の文書でこの問題にどのように対処すべきでしょうか？

- A. 現在のドキュメントを発音タグを含む SSML に変換します。
- B. 適切な発音辞書を作成します。
- C. 発音のガイドとなる音声記号を出力します。
- D. Amazon Lex を使用してテキストファイルを発音用に前処理します。

**Answer: A** ([メッセージを残す](#))

説明/参照: <https://docs.aws.amazon.com/polly/latest/dg/ssml.html>

#### 最新問題: 23

最近、新しい Google Cloud プロジェクトを作成しました。Cloud Shell から Vertex AI Pipeline ジョブを送信できることをテストした後、Vertex AI Workbench のユーザー管理のノートブックインスタンスを使用して、そのインスタンスからコードを実行したいと考えています。インスタンスを作成して実行しました。コードを記述しましたが、今回は権限が不十分なエラーでジョブが失敗します。あなたは何をするべきか？

- A. 作成した Workbench インスタンスが、使用する Vertex AI Pipelines リソースの同じリージョンにあることを確認してください。
- B. Vertex AI Workbench インスタンスが、使用する Vertex AI Pipeline リソースの同じサブネットワーク上にあることを確認します。
- C. Vertex AI Workbench インスタンスに Identity and Access Management (IAM) Vertex AI User role が割り当てられていることを確認します。
- D. Vertex AI Workbench インスタンスに Identity and Access Management (IAM) Notebooks Runner ロールが割り当てられていることを確認します。

**Answer: C** ([メッセージを残す](#))

#### 最新問題: 24

最近モデルを Vertex AI エンドポイントにデプロイしました。データが頻繁にドリフトするため、リクエストとレスポンスのログを有効にし、Vertex AI モデル監視ジョブを作成しました。モデルが予想よりも多くのトラフィックを受信していることがわかりました。ドリフトを迅速に検出し続けながら、モデルの監視コストを削減する必要があります。あなたは何をするべきか？

- A. モニタリング ジョブを TensorFlow Data Validation (TFDV) を使用する DataFlow パイプラインに置き換えます。
- B. モニタリング ジョブをカスタム SQL スクリプトに置き換えて、BigQuery の機能と予測に関する統計を計算します。
- C. モニタリング ジョブの Randomsampleconfig の sample\_rate パラメータを減らします。
- D. モニタリング ジョブの scheduleconfig の monitor\_interval パラメータを増やします。

**Answer: C** ([メッセージを残す](#))

公式試験ガイド1によると、試験で評価されるスキルの 1 つは「モデル監視ジョブの構成と最適化」です。Vertex AI モデル監視のドキュメントには、「モデル監視のコストを削減するために、モデル監視によってログに記録され分析されるリクエストのサンプル レートを構成できる」と記載されています。したがって、モニタリング ジョブの Randomsampleconfig のサンプルレート パラメータを減らすと、ドリフトを迅速に検出し続けながら、モデルのモニタリング コストが削減されます。他のオプションは、このシナリオには関係がなく、最適でもありません。参考文献:

- \* Google Cloud 認定資格の準備: 機械学習エンジニア プロフェッショナル認定資格
- \* Professional ML Engineer 試験ガイド
- \* Google プロフェッショナル機械学習認定試験 2023
- \* 最新の Google Professional Machine Learning Engineer の実際の無料試験問題
- \* [Vertex AI モデルのモニタリング]

#### 最新問題: 25

コードを記述せずに、カスタム カテゴリを使用して顧客レビューのセンチメントを予測するモデルを迅速に構築してトレーニングする必要があります。モデルを最初からトレーニングするのに十分なデータがありません。結果として得られるモデルは、高い予測パフォーマンスを備えている必要があります。どのサービスを使用する必要がありますか？

- A. AutoML 自然言語
- B. AI Hub の事前作成された Jupyter Notebook
- C. クラウド自然言語 API
- D. AI Platform Training 組み込みアルゴリズム

**Answer: A** ([メッセージを残す](#))

#### 最新問題: 26

あなたは、さまざまなオンプレミス データ マートにわたる統合分析環境を構築する責任を負います。あなたの会社は、サーバー間でデータを統合するときに、さまざまな非接続ツールや一時的なソリューションの使用が原因で、データ品質とセキュリティの問題に直面しています。総作業コ

ストを削減し、反復作業を削減する、フルマネージドのクラウドネイティブ データ統合サービスが必要です。チームのメンバーの中には、抽出、変換、読み込み (ETL) プロセスを構築するためのコードレス インターフェイスを好む人もいます。どのサービスを使用する必要がありますか？

- A. クラウド データ フェージョン
- B. データ準備
- C. Apache フリンク
- D. データフロー

**Answer:** ([解答を表示する](#))

#### 最新問題: 27

あなたは、会社の ML 実験とモデルの再トレーニングを自動化する MLOps プラットフォームを構築しています。数十のパイプラインのアーティファクトを整理する必要があるパイプラインのアーティファクトはどのように保存すればよいですか？

- A. パラメーターを Vertex ML メタデータに保存し、モデルのソース コードとバイナリを GitHub に保存します。
- B. パラメータを Cloud SQL に保存し、モデルのソースコードとバイナリを GitHub に保存します。
- C. Vertex ML メタデータにパラメータを保存します。モデルのソースコードは GitHub に保存され、モデルのバイナリは Cloud Storage に保存されます。
- D. パラメータを Cloud SQL に保存し、モデルのソースコードを GitHub に保存し、モデルのバイナリを Cloud Storage に保存します。

**Answer: C** ([メッセージを残す](#))

#### 最新問題: 28

会社のモバイルアプリケーションからのユーザー アクティビティ データを分析する必要があります。チームは、データ分析、変換、ML アルゴリズムの実験に BigQuery を使用します。ユーザー アクティビティ データを確実にリアルタイムで BigQuery に取り込む必要があります。あなたは何をすべきか？

- A. BigQuery にデータをストリーミングするように Pub/Sub を構成します。
- B. データを BigQuery に取り込むように Pub/Sub と Dataflow ストリーミング ジョブを構成します。
- C. Dataflow ストリーミング ジョブを実行して、データを BigQuery に取り込みます。
- D. Dataproc で Apache Spark ストリーミング ジョブを実行して、データを BigQuery に取り込みます。

**Answer: A** ([メッセージを残す](#))

#### 最新問題: 29

あなたは、規制対象の保険会社の ML エンジニアです。潜在顧客からの保険申請を承認または拒否する保険承認モデルを開発するように求められます。モデルを構築する前にどのような要素を考慮する必要がありますか？

- A. 編集、再現性、説明可能性
- B. トレーサビリティ、再現性、説明可能性
- C. フェデレーテッド ラーニング、再現性、説明可能性
- D. 差分プライバシー連合学習と説明可能性

**Answer: B** ([メッセージを残す](#))

<https://www.oecd.org/finance/Impact-Big-Data-AI-in-the-Insurance-Sector.pdf>

<https://medium.com/artefact-engineering-and-data-science/include-ethics-best-practices-in-your-data-science-project-from-day-one-c15b26c2bf99>

#### 最新問題: 30

Cloud TPU v2 を使用して物体検出モデルをトレーニングしています。トレーニング時間が予想よりも長くかかっています。Cloud TPU プロファイルで取得したこの簡略化されたトレースに基づいて、コスト効率の高い方法でトレーニング時間を短縮するにはどのようなアクションを実行する必要がありますか？

- A. Cloud TPU v2 から 8 つの NVIDIA V100 GPU に移行し、バッチ サイズを増やします。
- B. 入力関数を書き換えて、入力画像のサイズと形状を変更します。
- C. Cloud TPU v2 から Cloud TPU v3 に移行し、バッチ サイズを増やします。
- D. 並列読み取り、並列処理、プリフェッチを使用して入力関数を書き換えます。

**Answer: C** ([メッセージを残す](#))

#### 最新問題: 31

あなたは、TensorFlow フレームワークを使用して最先端の深層学習モデルを構築するチームで働いています。

チームは毎週複数の ML 実験を実行しているため、実験の実行を追跡することが困難になっています。

オーバーヘッド コードを最小限に抑えながら、Google Cloud 上で実行される ML 実験を効果的に追跡、可視化、デバッグするためのシンプルなアプローチが必要です。どのように進めるべきでしょうか？

- A. メトリクスとパラメーターを追跡するために Vertex AI Experiments をセットアップします。視覚化のために Vertex AI TensorBoard を構成します。
- B. メトリクス ファイルを Cloud Storage バケットに書き込んで保存するための Cloud Function を設定します。可視化のために TensorBoard をローカルでホストするように Google Cloud VM を構成します。
- C. Vertex AI Workbench ノートブック インスタンスをセットアップします。インスタンスを使用して指標データを Cloud Storage バケットに保存し、可視化のために TensorBoard をローカルにホストします。
- D. 指標ファイルを BigQuery テーブルに書き込んで保存するように Cloud Function を設定します。可視化のために TensorBoard をローカルでホストするように Google Cloud VM を構成します。

**Answer: A** ([メッセージを残す](#))

Vertex AI Experiments は、Google Cloud での ML 実験を追跡、比較、最適化できるサービスです。Vertex AI Experiments を使用して、TensorFlow モデルからのメトリクスとパラメーターをログに記録し、それらを Vertex AI TensorBoard で視覚化できます。Vertex AI TensorBoard は、ML 実験を表示およびデバッグするための Web インターフェイスを提供するマネージド サービスです。Vertex AI TensorBoard を使用して、さまざまな実行の比較、モデル グラフの検査、スカラー、ヒストグラム、画像などの分析を行うことができます。

Vertex AI Experiments と Vertex AI TensorBoard を使用すると、ML 実験の追跡と可視化のワークフローを簡素化し、独自の Cloud Functions、Cloud Storage バケツ、または VM のセットアップと保守のオーバーヘッドを回避できます。参考文献:

\* [Vertex AI Experiments ドキュメント]

\* [Vertex AI TensorBoard ドキュメント]

\* Google Cloud 認定資格の準備: 機械学習エンジニア プロフェッショナル認定資格

有効な **Professional-Machine-Learning-Engineer** 問題集は GoShiken.com が提供された合格しやすい Professional-Machine-Learning-Engineer 試験問題集！ GoShiken.com が最新の **Professional-Machine-Learning-Engineer** 試験問題集を提供しています。GoShiken.com Professional-Machine-Learning-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Professional-Machine-Learning-Engineer 問題集をゲットする人はこちら:  
<https://www.goshiken.com/Google/Professional-Machine-Learning-Engineer-mondaishu.html>  
(**29030%OFF**問題集溶と正解付きで **30%w**特別割引コード: **Freepdfdumps**)

最新問題: 32

チームは、画像に運転免許証、パスポート、クレジットカードが含まれているかどうかを予測するモデルを構築する必要があります。データ エンジニアリング チームはすでにパイプラインを構築し、運転免許証の画像 10,000 枚、パスポートの画像 1,000 枚、クレジットカードの画像 1,000 枚で構成されるデータセットを生成しました。次に、ラベル マップ ['driverslicense', 'passport', 'credit\_card'] を使用してモデルをトレーニングする必要があります。どの損失関数を使用する必要がありますか？

- A. カテゴリカル ヒンジ
- B. バイナリクロスエントロピー
- C. カテゴリカルクロスエントロピー
- D. スパースのカテゴリカルクロスエントロピー

**Answer: C (メッセージを残す)**

- **\*\*カテゴリエントロピー\*\***は、**\*\*モデルが特定のクラスをより重要視しないようにしたい\*\***場合に使用することをお勧めします。または、**\*\*クラスが非常に不均衡である\*\***場合は、**カテゴリエントロピー**を使用すると、より良い結果が得られます。

-ただし、メモリ使用量が多くなるほど大量のクラスがある場合は、\*\*疎カテゴリーエントロピー\*\*の方がより最適です。そのため、スパースカテゴリーエントロピーは使用する列が少ないため、\*\*メモリ使用量が少なくなります\*\*。

<https://stats.stackexchange.com/questions/326065/cross-entropy-vs-sparse-cross-entropy-when-to-use-one-over-the-other>

### 最新問題: 33

あなたは、生物有機体の特性に基づいた深層学習 ML モデルを実験しているバイオテクノロジーのスタートアップ企業で働いています。あなたのチームは、ML モデルの新しいアーキテクチャを使用した初期段階の実験に頻繁に取り組み、カスタム TensorFlow 演算を C++ で作成しています。大規模なデータセットと大きなバッチ サイズでモデルをトレーニングします。一般的なバッチ サイズには 1024 個のサンプルがあり、各サンプルのサイズは約 1 MB です。すべての重みと埋め込みを含むネットワークの平均サイズは 20 GB です。モデルにはどのハードウェアを選択する必要がありますか？

- A. それぞれ 8 個の NVIDIA Tesla V100 GPU (合計 128 GB GPU メモリ) を備えた 2 台の n1-highcpu-64 マシンと、64 個の vCPU と 58 GB RAM を備えた n1-highcpu-64 マシンで構成されるクラスター
- B. v2-8 TPU と 64 GB RAM を搭載した n1-highcpu-64 マシンを備えたクラスター
- C. それぞれ 96 個の vCPU と 86 GB RAM を備えた 4 台の n1-highcpu-96 マシンで構成されるクラスター
- D. 2 台の a2-megagpu-16g マシンで構成されるクラスター。各マシンには 16 個の NVIDIA Tesla A100 GPU (合計 640 GB GPU メモリ)、96 個の vCPU、および 1.4 TB RAM が搭載されています。

**Answer: D** ([メッセージを残す](#))

### 最新問題: 34

あなたは広告会社に勤めており、会社の最新の広告キャンペーンの効果を理解したいと考えています。500 MB のキャンペーン データを BigQuery にストリーミングしました。テーブルにクエリを実行し、AI Platform ノートブックの pandas データフレームを使用してそのクエリの結果を操作したいと考えています。あなたは何をするべきか？

- A. AI Platform Notebooks の BigQuery セルマジックを使用してデータをクエリし、結果を pandas データフレームとして取り込みます
- B. テーブルを CSV ファイルとして BigQuery から Google ドライブにエクスポートし、Google Drive API を使用してファイルをノートブック インスタンスに取り込みます。
- C. BigQuery からテーブルをローカル CSV ファイルとしてダウンロードし、AI Platform ノートブック インスタンスにアップロードします。パンダを使用します。ファイルをパンダデータフレームとして取り込むための read\_csv
- D. AI Platform ノートブックの bash セルから、bq extract コマンドを使用してテーブルを CSV ファイルとして Cloud Storage にエクスポートし、gsutil cp を使用してデータをノートブックに

コピーします。パンダを使用します。ファイルをパンダデータフレームとして取り込むための read\_csv

**Answer:** ([解答を表示する](#))

詳細については、このリンクを参照してください:

<https://cloud.google.com/bigquery/docs/bigquery-storage-python-pandas> 最初の 2 つのポイントは、データのクエリについて説明しています。

Jupyter ノートブックで BigQuery 用の IPython マジックから BigQuery Storage API を使用して、クエリ結果を pandas DataFrame にダウンロードします。

Python 用 BigQuery クライアント ライブラリを使用して、クエリ結果を pandas DataFrame にダウンロードします。

Python 用 BigQuery クライアント ライブラリを使用して、BigQuery テーブル データを pandas DataFrame にダウンロードします。

Python 用 BigQuery Storage API クライアント ライブラリを使用して、BigQuery テーブル データを pandas DataFrame にダウンロードします。

<https://googleapis.dev/python/bigquery/latest/magics.html#python-magics-for-bigquery>

<https://cloud.google.com/bigquery/docs/bigquery-storage-python-pandas>

#### 最新問題: 35

あなたは、大規模な BigQuery テーブルに保存されているデータに基づいて分類モデルをトレーニングする Vertex AI パイプラインを開発しました。パイプラインには 4 つのステップがあり、各ステップは KubeFlow v2 API を使用する Python 関数によって作成されます。コンポーネントの名前は次のとおりです。

次のように Vertex AI パイプラインを起動します。

トレーニング ステップのコードとパラメーターを調整することで、モデルの反復を多数実行します。開発、特にデータのエクスポートと前処理のステップに関連するコストが高額であることがわかります。モデル開発コストを削減する必要があります。あなたは何をするべきか？

- A.
- B.
- C.
- D.

**Answer:** B ([メッセージを残す](#))

#### 最新問題: 36

あなたは、ビジュアル検索エンジンを開発しているオンライン小売会社で働いています。画像に会社の製品が含まれているかどうかを分類するために、Google Cloud 上にエンドツーエンドの ML パイプラインを設定しました。近い将来の新製品のリリースを予期して、新しいデータを ML モデルにフィードできるようにパイプラインに再トレーニング機能を構成しました。また、AI Platform の継続的評価サービスを使用して、テスト データ セットに対してモデルの精度が高いことを確認したいと考えています。あなたは何をするべきか？

- A. 新しい製品が再トレーニングに組み込まれた場合でも、元のテスト データセットを変更しないでください。
  - B. 評価指標が事前に決められたしきい値を下回った場合、新しい製品の画像でテスト データセットを更新します。
  - C. 再トレーニングに導入される新しい製品の画像を使用してテスト データセットを拡張します。
  - D. 再トレーニングに導入される新しい製品のイメージでテスト データセットを置き換えます。
- Answer: C** ([メッセージを残す](#))

**最新問題: 37**

あなたのチームは、何百万もの顧客が使用する世界的な銀行用のアプリケーションを構築しています。3 日後の顧客 1 の口座残高を予測する予測モデルを構築しました。あなたのチームはその結果を、アカウント残高が 25 ドルを下回る可能性があるときにユーザーに通知する新機能で使します。予測をどのように提供すべきでしょうか？

- A. 1. Firebase 上に通知システムを構築する  
2. 各ユーザーを Firebase Cloud Messaging サーバー上のユーザー ID に登録します。すべてのアカウント残高予測の平均が 25 ドルのしきい値を下回ると、通知が送信されます。
- B. 1. ユーザーごとに Pub/Sub トピックを作成します。  
2 ユーザーのアカウント残高が 25 ドルのしきい値を下回るとモデルが予測したときに通知を送信する Cloud Function をデプロイします。
- C. 1. ユーザーごとに Pub/Sub トピックを作成します。  
2. ユーザーのアカウント残高が 25 ドルのしきい値を下回るとモデルが予測したときに通知を送信するアプリケーションを App Engine スタンダード環境にデプロイします。
- D. 1 Firebase 上に通知システムを構築する  
2. 各ユーザーを Firebase Cloud Messaging サーバー上のユーザー ID に登録します。ユーザーのアカウント残高が 25 ドルのしきい値を下回るとモデルが予測すると、通知が送信されます。

**Answer:** ([解答を表示する](#))

**最新問題: 38**

あなたは大企業のデータサイエンス部門のディレクターで、データサイエンスチームは最近、Kubeflow Pipelines SDK を使用してトレーニングパイプラインを調整し始めました。あなたのチームは、カスタム Python コードを Kubeflow Pipelines SDK に統合するのに苦労しています。コードを Kubeflow Pipelines SDK と迅速に統合するには、続行するようにどのように指示すればよいでしょうか？

- A. `func_to_container_op` 関数を使用して、Python コードからカスタムコンポーネントを作成します。
- B. Kubeflow Pipelines SDK で利用可能な事前定義コンポーネントを使用して Dataproc にアクセスし、そこでカスタムコードを実行します。
- C. カスタム Python コードを Docker コンテナにパッケージ化し、`load_component_from_file` 関数を使用してコンテナをパイプラインにインポートします。

D. カスタム Python コードを Cloud Functions にデプロイし、Kubeflow Pipelines を使用して Cloud Functions をトリガーします。

**Answer:** ([解答を表示する](#))

カスタム Python コードを Kubeflow Pipelines SDK に統合する最も簡単な方法は、Python 関数をパイプライン コンポーネントに変換する `func_to_container_op` 関数を使用することです。この関数は、Python 関数を実行する Docker イメージを自動的に構築し、パイプラインの `kfp.dsl.ContainerOp` インスタンスの作成に使用できるファクトリ関数を返します。このオプションには次の利点があります。

- \* これにより、データサイエンスチームは、手動で書き直したりコンテナにパッケージ化したりすることなく、既存の Python コードを再利用できます。
- \* 関数シグネチャがコンポーネントインターフェイスを定義し、関数本体がコンポーネントロジックを定義するため、コンポーネントの仕様と実装が簡素化されます。
- \* プリミティブ型、ファイル、ディレクトリ、辞書など、さまざまなタイプの入出力をサポートします。

他のオプションは、次の理由からあまり最適ではありません。

\* オプション B: Kubeflow Pipelines SDK で利用可能な事前定義コンポーネントを使用して Dataproc にアクセスし、そこでカスタムコードを実行すると、さらに複雑さとコストが発生します。このオプションでは、Dataproc クラスタの作成と管理が必要です。Dataproc クラスタは、Apache Spark と Apache Hadoop を実行する Compute Engine インスタンスの一時的でスケラブルなクラスタです。さらに、このオプションでは、PySpark または Hadoop MapReduce でカスタムコードを記述する必要がありますが、既存の Python コードと互換性がない可能性があります。

\* オプション C: カスタム Python コードを Docker コンテナにパッケージ化し、`load_component_from_file` 関数を使用してコンテナをパイプラインにインポートすると、追加の手順とオーバーヘッドが発生します。このオプションでは、Dockerfile の作成と保守、Docker イメージの構築とプッシュ、YAML ファイルへのコンポーネント仕様の記述が必要です。さらに、このオプションでは、Python コードと Docker イメージの依存関係とバージョンを管理する必要があります。

\* オプション D: カスタム Python コードを Cloud Functions にデプロイし、Kubeflow Pipelines を使用して Cloud Functions をトリガーすると、追加のレイテンシと制限が発生します。このオプションでは、イベントにตอบสนองして実行されるサーバーレス関数である Cloud Functions を作成してデプロイする必要があります。

さらに、このオプションでは、HTTP リクエストを使用して Kubeflow Pipelines から Cloud Functions を呼び出す必要があるため、ネットワークのオーバーヘッドと遅延が発生する可能性があります。さらに、このオプションは、最大実行時間やメモリ使用量など、Cloud Functions の割り当てと制限の影響を受けます。

参考文献:

- \* Python 関数ベースのコンポーネントの構築 | キューブフロー
- \* Python 関数ベースのコンポーネントの構築 | キューブフロー

### 最新問題: 39

You work for a global footwear retailer and need to predict when an item will be out of stock based on historical inventory data. Customer behavior is highly dynamic since footwear demand is influenced by many different factors. You want to serve models that are trained on all available data, but track your performance on specific subsets of data before pushing to production. What is the most streamlined and reliable way to perform this validation?

- A.** Use the TFX ModelValidator tools to specify performance metrics for production readiness
- B.** Use k-fold cross-validation as a validation strategy to ensure that your model is ready for production.
- C.** Use the last relevant week of data as a validation set to ensure that your model is performing accurately on current data
- D.** Use the entire dataset and treat the area under the receiver operating characteristics curve (AUC ROC) as the main metric.

**Answer: A** ([メッセージを残す](#))

\* TFX ModelValidator is a tool that allows you to compare new models against a baseline model and evaluate their performance on different metrics and data slices<sup>1</sup>. You can use this tool to validate your models before deploying them to production and ensure that they meet your expectations and requirements.

\* k-fold cross-validation is a technique that splits the data into k subsets and trains the model on k-1 subsets while testing it on the remaining subset. This is repeated k times and the average performance is reported<sup>2</sup>. This technique is useful for estimating the generalization error of a model, but it does not account for the dynamic nature of customer behavior or the potential changes in data distribution over time.

\* Using the last relevant week of data as a validation set is a simple way to check the model's performance on recent data, but it may not be representative of the entire data or capture the long-term trends and patterns. It also does not allow you to compare the model with a baseline or evaluate it on different data slices.

\* データセット全体を使用し、AUC ROC を主要な指標として扱うことは、検証やテストのためのデータが残らないため、良い方法ではありません。また、AUC ROC が重要な唯一の指標であると想定していますが、ビジネス上の問題には当てはまらない可能性があります。精度、再現率、収益などの他の指標を考慮することもできます。

### 最新問題: 40

You want to train an AutoML model to predict house prices by using a small public dataset stored in BigQuery. You need to prepare the data and want to use the simplest most efficient approach. What should you do?

- A.** Use Dataflow to preprocess the data Write the output in TFRecord format to a Cloud Storage bucket.
- B.** Write a query that preprocesses the data by using BigQuery and creates a new table Create a Vertex AI managed dataset with the new table as the data source.

C. Write a query that preprocesses the data by using BigQuery Export the query results as CSV files and use those files to create a Vertex AI managed dataset.

D. Use a Vertex AI Workbench notebook instance to preprocess the data by using the pandas library Export the data as CSV files, and use those files to create a Vertex AI managed dataset.

**Answer: B (メッセージを残す)**

#### 最新問題: 41

トラフィックを処理している実稼働 Vertex AI エンドポイントにモデルの新しいバージョンをデプロイしています。すべてのユーザーがトラフィックを新しいモデルに誘導する予定です。アプリケーションへの中断を最小限に抑えてモデルをデプロイする必要があります。どうすればよいですか？

A. 1 新しいエンドポイントを作成します。

2 新しいモデルを作成します。デフォルトバージョンとして設定します。モデルを Vertex AI Model Registry にアップロードします。

3. 新しいモデルを新しいエンドポイントにデプロイします。

4 新しいエンドポイントを指すように Cloud DNS を更新します

B. 1. 新しいエンドポイントを作成します。

2. 新しいモデルを作成します。parentModel パラメーターを現在デプロイされているモデルのモデル ID に設定し、デフォルトバージョンとして設定します。モデルを Vertex AI Model Registry にアップロードします。

3. 新しいモデルを新しいエンドポイントにデプロイし、新しいモデルをトラフィックの 100% に設定します。

C. 1 新しいモデルを作成します。parentModel パラメーターを現在デプロイされているモデルのモデル ID に設定します。モデルを Vertex AI Model Registry にアップロードします。

2 新しいモデルを既存のエンドポイントに展開し、新しいモデルをトラフィックの 100% に設定します。

D. 1、新しいモデルを作成します。デフォルトのバージョンとして設定します。モデルを Vertex AI Model Registry にアップロードします。

2 新しいモデルを既存のエンドポイントにデプロイします

**Answer: C (メッセージを残す)**

トラフィックを処理し、すべてのユーザーがトラフィックを新しいモデルに誘導し、アプリケーションへの中断を最小限に抑えてモデルをデプロイする本番環境の Vertex AI エンドポイントにモデルの新しいバージョンをデプロイするための最良のオプションは、新しいモデルを作成し、設定することです。parentModel パラメーターを現在デプロイされているモデルのモデル ID に設定し、モデルを Vertex AI Model Registry にアップロードし、新しいモデルを既存のエンドポイントにデプロイして、新しいモデルをトラフィックの 100% に設定します。このオプションを使用すると、Vertex AI のパワーとシンプルさを活用してモデルのバージョンを更新し、低遅延でオンライン予測を提供できます。Vertex AI は、Google Cloud 上で機械学習ソリューションを構築およびデプロイするための統合プラットフォームです。Vertex AI は、トレーニングされたモデルをオンライン予測エンドポイントにデプロイし、個々のインスタンスに対して低レイテンシの予測

を提供できます。モデルは、予測に使用できる機械学習モデルを表すリソースです。モデルには、同じモデルの異なる実装である 1 つ以上のバージョンを含めることができます。モデルバージョンには、同じモデルの別のバージョンとは異なるパラメータ、コード、またはデータを含めることができます。モデルバージョンは、モデルの実験と反復を行い、モデルのパフォーマンスと精度を向上させるのに役立ちます。parentModel パラメータは、新しいモデルバージョンのベースとなるモデルのモデル ID を指定するパラメータです。parentModel パラメータは、既存のモデルの設定とメタデータを継承し、モデル構成の重複を避けるのに役立ちます。Vertex AI Model Registry は、機械学習モデルを Google Cloud に保存して管理できるサービスです。Vertex AI Model Registry は、モデルのアップロードと整理、モデルのバージョンとメタデータの追跡に役立ちます。エンドポイントは、予測のリクエストに使用するサービス エンドポイント (URL) を提供するリソースです。エンドポイントには、物理リソースに関連付けられたモデルバージョンのインスタンスである 1 つ以上のデプロイされたモデルを含めることができます。デプロイされたモデルは、低遅延でオンライン予測を提供し、トラフィックに基づいてスケールアップまたはスケールダウンするのに役立ちます。新しいモデルを作成し、parentModel パラメータを現在デプロイされているモデルのモデル ID に設定し、モデルを Vertex AI Model Registry にアップロードし、新しいモデルを既存のエンドポイントにデプロイし、新しいモデルをトラフィックの 100% に設定することで、トラフィックを処理している運用環境の Vertex AI エンドポイントにモデルの新しいバージョンをデプロイし、すべてのユーザー トラフィックを新しいモデルに誘導し、アプリケーションへの中断を最小限に抑えてモデルをデプロイできます<sup>1</sup>。

他のオプションは、次の理由により、オプション C ほど優れていません。

\* オプション A: 新しいエンドポイントを作成し、新しいモデルを作成してデフォルトバージョンとして設定し、モデルを Vertex AI Model Registry にアップロードし、新しいモデルを新しいエンドポイントにデプロイし、新しいエンドポイントを指すように Cloud DNS を更新します。新しいモデルを作成し、parentModel パラメータを現在デプロイされているモデルのモデル ID に設定し、モデルを Vertex AI モデル レジストリにアップロードし、新しいモデルを既存のエンドポイントにデプロイし、新しいモデルを 100 に設定するよりも多くのスキルと手順が必要です。トラフィックの %。Cloud DNS は、信頼性が高くスケーラブルなドメイン ネーム システム (DNS) サービスを Google Cloud 上で提供できるサービスです。Cloud DNS は、DNS レコードを管理し、ドメイン名を IP アドレスに解決するのに役立ちます。新しいエンドポイントを指すように Cloud DNS を更新すると、ユーザー トラフィックを新しいエンドポイントにリダイレクトし、既存のアプリケーションの破損を回避できます。ただし、新しいエンドポイントの作成、新しいモデルの作成、デフォルトバージョンとしての設定、Vertex AI Model Registry へのモデルのアップロード、新しいエンドポイントへの新しいモデルのデプロイ、新しいエンドポイントを指すように Cloud DNS の更新を行うには、さらに多くの作業が必要になります。新しいモデルの作成、parentModel パラメータを現在デプロイされているモデルのモデル ID に設定、モデルを Vertex AI モデル レジストリにアップロード、新しいモデルを既存のエンドポイントにデプロイ、新しいモデルを

トラフィックの 100%。コードを記述し、新しいエンドポイントを作成して構成し、新しいモデルを作成して構成し、モデルを Vertex AI Model Registry にアップロードし、モデルを新しいエンド

ポイントにデプロイし、新しいエンドポイントを指すように Cloud DNS を更新する必要があります。さらに、このオプションでは新しいエンドポイントが作成されるため、メンテナンスと管理のコストが増加する可能性があります<sup>2</sup>。

\* オプション B: 新しいエンドポイントを作成し、新しいモデルを作成し、parentModel パラメーターを現在デプロイされているモデルのモデル ID に設定し、それをデフォルトバージョンとして設定し、モデルを Vertex AI モデル レジストリにアップロードし、新しいモデルを新しいエンドポイントを設定し、新しいモデルをトラフィックの 100% に設定するには、新しいモデルを作成し、parentModel パラメーターを現在デプロイされているモデルのモデル ID に設定し、モデルを Vertex AI Model Registry にアップロードし、デプロイするよりも多くのスキルと手順が必要になります。新しいモデルを既存のエンドポイントに設定し、新しいモデルを

トラフィックの 100%。parentModel パラメーターは、新しいモデルバージョンのベースとなるモデルのモデル ID を指定するパラメーターです。parentModel パラメーターは、既存のモデルの設定とメタデータを継承し、モデル構成の重複を避けるのに役立ちます。デフォルトバージョンは、他のバージョンが指定されていない場合に予測に使用されるモデルバージョンです。デフォルトバージョンを使用すると、予測リクエストが簡素化され、毎回モデルバージョンを指定する必要がなくなります。parentModel パラメーターを現在デプロイされているモデルのモデル ID に設定し、それをデフォルトバージョンとして設定すると、モデルバージョンを指定せずに、既存のモデルに基づいて新しいモデルを作成し、それを予測に使用できます。ただし、新しいエンドポイントを作成し、新しいモデルを作成し、parentModel パラメーターを現在デプロイされているモデルのモデル ID に設定し、それをデフォルトバージョンとして設定し、モデルを Vertex AI モデル レジストリにアップロードし、新しいモデルを新しいエンドポイントを作成し、新しいモデルをトラフィックの 100% に設定するには、新しいモデルを作成して、parentModel パラメーターを現在デプロイされているモデル ID に設定するよりも多くのスキルと手順が必要になります。

\* モデル、モデルを Vertex AI Model Registry にアップロードし、新しいモデルを既存のエンドポイントにデプロイし、新しいモデルをトラフィックの 100% に設定します。コードを記述し、新しいエンドポイントを作成して構成し、新しいモデルを作成して構成し、モデルを Vertex AI Model Registry にアップロードして、モデルを新しいエンドポイントにデプロイする必要があります。さらに、このオプションでは新しいエンドポイントが作成されるため、メンテナンスと管理のコストが増加する可能性があります<sup>2</sup>。

\* オプション D: 新しいモデルを作成し、それをデフォルトバージョンとして設定し、モデルを Vertex AI Model Registry にアップロードし、新しいモデルを既存のエンドポイントにデプロイすると、既存のモデルの設定とメタデータを継承できなくなります。エラーやパフォーマンスの低下を引き起こす可能性があります。デフォルトバージョンは、他のバージョンが指定されていない場合に予測に使用されるモデルバージョンです。デフォルトバージョンを使用すると、予測リクエストが簡素化され、毎回モデルバージョンを指定する必要がなくなります。

新しいモデルをデフォルトバージョンとして設定すると、モデルのバージョンを指定せずに新しいモデルを予測に使用できます。ただし、新しいモデルを作成し、それをデフォルトバージョンとして設定し、モデルを Vertex AI Model Registry にアップロードし、新しいモデルを既存のエンドポイントにデプロイすると、既存のモデルの設定とメタデータを継承できなくなり、次のような

問題が発生する可能性があります。エラーやパフォーマンスの低下。コードを記述し、新しいモデルを作成して構成し、モデルを Vertex AI Model Registry にアップロードして、モデルを既存のエンドポイントにデプロイする必要があります。さらに、このオプションでは、parentModel パラメーターが現在デプロイされているモデルのモデル ID に設定されないため、既存のモデルの設定とメタデータを継承できなくなり、モデルバージョン間で不整合や競合が発生する可能性があります2。

参考文献:

\* Google Cloud 認定資格の準備: 機械学習エンジニア、コース 3: 本番環境 ML システム、第 2 週: ML 予測の提供

\* Google Cloud Professional Machine Learning Engineer 試験ガイド、セクション 3: 本番環境での ML モデルのスケーリング、3.1 本番環境への ML モデルのデプロイ

\* 公式 Google Cloud Certified Professional Machine Learning Engineer 学習ガイド、第 6 章: 実稼働 ML システム、セクション 6.2: ML 予測の提供

\* 頂点AI

\* クラウドDNS

#### 最新問題: 42

あなたは、Google Kubernetes Engine で Kubeflow パイプラインを開発しています。パイプラインの最初のステップは、BigQuery に対してクエリを発行することです。そのクエリの結果をパイプラインの次のステップへの入力として使用することを計画しています。これをできるだけ簡単な方法で達成したいと考えています。あなたは何をすべきか？

**A.** BigQuery コンソールを使用してクエリを実行し、クエリ結果を新しい BigQuery テーブルに保存します。

**B.** BigQuery API を使用して BigQuery に対してクエリを実行する Python スクリプトを作成します。このスクリプトを Kubeflow パイプラインの最初のステップとして実行します。

**C.** Kubeflow Pipelines ドメイン固有言語を使用して、Python BigQuery クライアント ライブラリを使用してクエリを実行するカスタム コンポーネントを作成します。

**D.** GitHub で Kubeflow Pipelines リポジトリを見つけます。BigQuery クエリ コンポーネントを見つけて、そのコンポーネントの URL をコピーし、それを使用してコンポーネントをパイプラインに読み込みます。コンポーネントを使用して BigQuery に対してクエリを実行する

**Answer:** ([解答を表示する](#))

<https://linuxtut.com/en/f4771efee37658c083cc/>

<https://github.com/kubeflow/pipelines/blob/master/components/gcp/bigquery/query/sample.ipynb>  
; <https://v0-5.kubeflow.org/docs/pipelines/reusable-components/>

#### 最新問題: 43

あなたはストリーミング音楽サービスを運営する組織で働いています。ユーザーの最近の視聴履歴に基づいて「次の曲」の推奨を提供するカスタム実稼働モデルがあります。モデルは Vertex AI エンドポイントにデプロイされます。最近、新しいデータを使用して同じモデルを再トレーニング

グしました。モデルはオフラインで肯定的なテスト結果を受け取りました。ここで、複雑さを最小限に抑えながら、実稼働環境で新しいモデルをテストしたいと考えています。あなたは何をするべきか？

- A. 新しいモデルの新しい Vertex AI エンドポイントを作成し、新しいモデルをその新しいエンドポイントにデプロイする 運用トラフィックの 5% を新しいエンドポイントにランダムに送信するサービスを構築する リスニング時間などのエンドユーザー メトリクスを監視するエンドユーザーのメトリクスは時間の経過とともにモデル間で改善され、新しいエンドポイントに送信される運用トラフィックの割合が徐々に増加します。
- B. 既存の Vertex AI エンドポイントのモデル監視ジョブを構成します。予測のずれを検出するように監視ジョブを構成し、アラートのしきい値を設定します。エンドポイントのモデルを以前のモデルから新しいモデルに更新します。予測のずれのアラートを受信した場合は、以前のモデルに戻します。
- C. 新しいモデルを既存の Vertex AI エンドポイントにデプロイする トラフィック分割を使用して運用トラフィックの 5% を新しいモデルに送信する リスニング時間などのエンドユーザー メトリクスを監視する 時間の経過とともにモデル間でエンドユーザー メトリクスが改善するかどうか、新しいモデルに送信される本番トラフィックの割合を徐々に増やします。
- D. BigQuery で受信予測リクエストをキャプチャする Vertex AI Experiments で実験を作成する キャプチャしたデータを使用して両方のモデルのバッチ予測を実行する ユーザーが選択した曲を使用してモデルのパフォーマンスを並べて比較する 新しいモデルのパフォーマンス指標が優れている場合以前のモデルよりも新しいモデルを実稼働環境にデプロイします。

**Answer: C ([メッセージを残す](#))**

#### 最新問題: 44

あなたはチームと共同でモデルのプロトタイプを作成しています。チームのメンバー用に Vertex AI Workbench 環境を作成し、プロジェクト内の他の従業員へのアクセスを制限する必要があります。あなたは何をするべきか？

- A.
  1. 新しいサービス アカウントを作成し、それに Notebook Viewer ロールを付与します。
  2. サービス アカウントの各チーム メンバーにサービス アカウント ユーザーの役割を付与します。
  3. 各チーム メンバーに Vertex AI ユーザー ロールを付与します。
  4. 新しいサービス アカウントを使用する Vertex AI Workbench ユーザー管理のノートブック インスタンスをプロビジョニングします。
- B.
  1. Vertex AI User ロールをデフォルトの Compute Engine サービス アカウントに付与します。
  2. デフォルトの Compute Engine サービス アカウントの各チーム メンバーにサービス アカウント ユーザーの役割を付与します。
  3. デフォルトの Compute Engine サービス アカウントを使用する Vertex AI Workbench ユーザー管理のノートブック インスタンスをプロビジョニングします。
- C.
  1. 新しいサービス アカウントを作成し、Vertex AI User ロールを付与します。
  2. サービス アカウントの各チーム メンバーにサービス アカウント ユーザーの役割を付与します。

3. 各チームメンバーにノートブック閲覧者の役割を付与します。
4. 新しいサービス アカウントを使用する Vertex AI Workbench ユーザー管理のノートブック インスタンスをプロビジョニングします。

D. 1 Vertex AI User ロールをプライマリ チーム メンバーに付与します。

2. 他のチームメンバーにノートブック閲覧者の役割を付与します。
3. プライマリ ユーザーのアカウントを使用する Vertex AI Workbench ユーザー管理のノートブック インスタンスをプロビジョニングします。

**Answer: C** ([メッセージを残す](#))

チーム用の Vertex AI Workbench 環境を作成し、プロジェクト内の他の従業員へのアクセスを制限するには、次の手順に従う必要があります。

\* 新しいサービス アカウントを作成し、Vertex AI ユーザー ロールを付与します。このロールは、ノートブック インスタンスの作成と管理を含む、Vertex AI のすべてのリソースへの完全なアクセス権を付与します<sup>1</sup>。

\* サービス アカウントの各チーム メンバーにサービス アカウント ユーザーの役割を付与します。この役割により、チーム メンバーはサービス アカウントになりすましてそのアクセス許可を使用できるようになります<sup>2</sup>。

\* 各チーム メンバーにノートブック閲覧者の役割を付与します。この役割により、チーム メンバーはノートブック インスタンスを表示して接続することができますが、変更または削除することはできません<sup>3</sup>。

\* 新しいサービス アカウントを使用する Vertex AI Workbench ユーザー管理のノートブック インスタンスをプロビジョニングします。

このようにすると、ノートブック インスタンスはサービス アカウントとして実行され、サービス アカウント ユーザー ロールとノートブック閲覧者のロールを持つチーム メンバーのみがアクセスできるようになります。

参考文献:

\*1: IAM による Vertex AI アクセス制御 | グーグルクラウド

\* 2: サービス アカウントについて | クラウド IAM ドキュメント

\* 3: Vertex AI Workbench インスタンスへのアクセスを管理 | グーグルクラウド

\* [4]: Vertex AI Workbench インスタンスを作成および管理する | グーグルクラウド

### 最新問題: 45

あなたの組織のコールセンターから、各通話における顧客の感情を分析するモデルを開発するよう依頼されました。コールセンターには毎日 100 万件を超える電話がかかり、データは Cloud Storage に保存されます。収集されたデータは通話が発信された地域から出てはならず、個人を特定できる情報 (PII) を保存または分析することはできません。データサイエンスチームは、SQL ANSI-2011 準拠のインターフェイスを必要とする視覚化とアクセス用のサードパーティ ツールを備えています。データ処理および分析用のコンポーネントを選択する必要があります。データパイプラインはどのように設計すべきでしょうか？

A. 1 = データフロー、2 = BigQuery

B. 1 = Pub/Sub、2 = データストア

C. 1 = データフロー、2 = Cloud SQL

D. 1 = Cloud Function、2 = Cloud SQL

**Answer: A (メッセージを残す)**

データ パイプラインは、通常は分析、変換、または保存を目的として、1 つ以上のソースから 1 つ以上の宛先にデータを移動する一連のステップまたはプロセスです。データ パイプラインは、データ ソース、データ処理ツール、データ ストレージ システム、データ分析ツールなどのさまざまなコンポーネントを使用して設計できます<sup>1</sup>。各通話における顧客の感情を分析するためのデータ パイプラインを設計するには、次の要件と制約を考慮する必要があります。

\* コールセンターには毎日 100 万件以上の電話があり、データはクラウド ストレージに保存されます。これは、データが大きく、構造化されておらず、分散しているため、音声、テキスト、画像などのさまざまな種類のデータ形式を処理できる、スケーラブルで効率的なデータ処理ツールが必要であることを意味します。

\* 収集されたデータは通話が発信された地域から出てはならず、個人を特定できる情報 (PII) を保存または分析することはできません。これは、データが機密であり、データの影響を受けることを意味します。

\* プライバシーとコンプライアンスの規制に準拠しており、データ暗号化、アクセス制御、地域ポリシーを強制できる安全で信頼性の高いデータ ストレージ システムが必要です。

\* データ サイエンس チームは、SQL ANSI-2011 準拠のインターフェイスを必要とする視覚化とアクセス用のサードパーティ ツールを備えています。これは、データ分析ツールが外部にあり、データ パイプラインから独立しており、SQL クエリと操作をサポートできる互換性のある標準のデータ インターフェイスが必要であることを意味します。

データ処理および分析用のコンポーネントを選択するための最良のオプションの 1 つは、データ処理には Dataflow を使用し、分析には BigQuery を使用することです。Dataflow は、バッチまたはストリーム処理、抽出、変換、読み込み (ETL)、またはデータ統合などのデータ処理のための Apache Beam パイプラインを実行するためのフルマネージド サービスです。BigQuery は、サーバーレスでスケーラブルでコスト効率の高いデータ ウェアハウスであり、大規模なデータに対して高速で複雑なクエリを実行できます<sup>23</sup>。Dataflow と BigQuery を使用すると、このユースケースでは次のような利点があります。

\* Dataflow は、Cloud Storage からの大規模な非構造化データを並列分散方式で処理し、音声からテキストへの変換、センチメント スコアの抽出、PII の匿名化などのさまざまな変換を適用できます。Dataflow はバッチ処理とストリーム処理の両方を処理することもできるため、通話データのリアルタイムまたはほぼリアルタイムの分析が可能になります。

\* BigQuery は、Dataflow からの処理されたデータを安全かつ信頼性の高い方法で保存および分析し、データ暗号化、アクセス制御、地域ポリシーを適用できます。BigQuery は SQL ANSI-2011 準拠のインターフェイスもサポートしているため、データ サイエンс チームは可視化とアクセスにサードパーティ ツールを使用できます。BigQuery は、AI Platform、データスタジオ、Looker などのさまざまな Google Cloud サービスやツールと統合することもできます。

\* Dataflow と BigQuery はどちらも Google Cloud エコシステムの一部であ

り、CSV、JSON、Avro、Parquet などのさまざまなデータ形式をサポートしているため、シームレ

スに連携できません。Dataflow と BigQuery は、スケーラビリティ、パフォーマンス、費用対効果などの Google Cloud インフラストラクチャの利点も活用できます。

他のオプションは適切ではないか、実現可能ではありません。データ処理に Pub/Sub を使用し、分析に Datastore を使用することは理想的ではありません。Pub/Sub は主にデータ処理ではなくイベント駆動型の非同期メッセージング用に設計されており、Datastore は主に低レイテンシーと高スループットの Key-Value 用に設計されているためです。分析ではなく運用です。

データ処理に Cloud Function を使用し、分析に Cloud SQL を使用することは最適ではありません。Cloud Function にはメモリ、CPU、実行時間に制限があり、複雑なデータ処理はサポートされていません。また、Cloud SQL はスケールできない可能性があるリレーショナル データベース サービスであるためです。大規模なデータに適しています。データ処理に Cloud Composer を使用し、分析に Cloud SQL を使用することは関係ありません。Cloud Composer は主に、データ処理ではなく、複数のシステムにわたる複雑なワークフローをオーケストレーションするために設計されており、Cloud SQL はリレーショナル データベース サービスであり、大規模にはうまく拡張できない可能性があります。データ。

参考資料: 1: データ パイプライン 2: データフローの概要 3: BigQuery の概要 : [Dataflow ドキュメント]:

[BigQuery ドキュメント]

#### 最新問題: 46

You work for a company that provides an anti-spam service that flags and hides spam posts on social media platforms. Your company currently uses a list of 200,000 keywords to identify suspected spam posts. If a post contains more than a few of these keywords, the post is identified as spam. You want to start using machine learning to flag spam posts for human review. What is the main advantage of implementing machine learning for this business case?

- A. Posts can be compared to the keyword list much more quickly.
- B. New problematic phrases can be identified in spam posts.
- C. A much longer keyword list can be used to flag spam posts.
- D. Spam posts can be flagged using far fewer keywords.

**Answer: B** ([メッセージを残す](#))

The main advantage of implementing machine learning for this business case is that new problematic phrases can be identified in spam posts. This is because machine learning can learn from the data and the feedback, and adapt to the changing patterns and trends of spam posts. Machine learning can also capture the semantic and contextual meaning of the posts, and not just rely on the presence or absence of keywords. By using machine learning, you can improve the accuracy and coverage of your anti-spam service, and detect new and emerging types of spam posts that may not be captured by the keyword list.

The other options are not advantages of implementing machine learning for this business case for the following reasons:

\* A. Posts can be compared to the keyword list much more quickly is not an advantage, as it does not improve the quality or effectiveness of the anti-spam service. It only improves the efficiency of

the service, which is not the primary objective. Moreover, machine learning may not necessarily be faster than the keyword list, depending on the complexity and size of the model and the data.

\* C. A much longer keyword list can be used to flag spam posts is not an advantage, as it does not address the limitations or challenges of the keyword list approach. It only increases the size and complexity of the keyword list, which can make it harder to maintain and update. Moreover, a longer keyword list may not improve the accuracy or coverage of the anti-spam service, as it may introduce more false positives or false negatives, or miss new and emerging types of spam posts.

\* D. Spam posts can be flagged using far fewer keywords is not an advantage, as it does not reflect the capabilities or benefits of machine learning. It only reduces the size and complexity of the keyword list, which can make it easier to maintain and update. However, using fewer keywords may not improve the accuracy or coverage of the anti-spam service, as it may lose some information or meaning of the posts, or miss some types of spam posts.

References:

- \* Professional ML Engineer Exam Guide
- \* Preparing for Google Cloud Certification: Machine Learning Engineer Professional Certificate
- \* Google Cloud launches machine learning engineer certification
- \* Machine Learning for Spam Detection
- \* Spam Detection Using Machine Learning

有効な **Professional-Machine-Learning-Engineer** 問題集は GoShiken.com が提供された合格しやすい Professional-Machine-Learning-Engineer 試験問題集！ GoShiken.com が最新の **Professional-Machine-Learning-Engineer** 試験問題集を提供しています。GoShiken.com Professional-Machine-Learning-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Professional-Machine-Learning-Engineer 問題集をゲットする人はこちら：  
<https://www.goshiken.com/Google/Professional-Machine-Learning-Engineer-mondaishu.html>  
(29030%OFF問題集溶と正解付きで 30%w 特別割引コード: **Freepdfdumps**)

最新問題: 47

あなたはホテルで働いており、紙ベースの顧客フィードバック フォームからスキャンした顧客のコメントを含むデータセットを PDF ファイルとして保存しています。どのフォームも同じレイアウトです。各フォームに対する顧客のコメントから全体的な満足度スコアを迅速に予測する必要があります。このタスクをどのように達成する必要がありますか?」

**A.** Vision API を使用して、各 PDF ファイルのテキストを解析します。Natural Language API の分析センチメント機能を使用して、全体的な満足度スコアを推測します。

**B.** Vision API を使用して、各 PDF ファイルのテキストを解析します。Natural Language API の analyzeEntitysentiment 機能を使用して、全体的な満足度スコアを推測します。

C. Document AI カスタム エクストラクターをアップトレーニングして、各 PDF ファイルのコメント セクションのテキストを解析します。Natural Language API のセンチメント分析機能を使用して、全体的な満足度スコアを推測します。

D. Document AI カスタム エクストラクターをアップトレーニングして、各 PDF ファイルのコメント セクション内のテキストを解析します。Natural Language API のanalyzeEntitySentiment 機能を使用して、全体的な満足度スコアを推測します。

**Answer: C (メッセージを残す)**

公式試験ガイド1によると、試験で評価されるスキルの1つは、「Google Cloud テクノロジーを使用してビジネス課題を解決するための ML モデルを設計、構築、本番化する」ことです。Document AI2 は、ドキュメントから非構造化データを取得して構造化データに変換し、理解、分析、利用を容易にするドキュメント理解プラットフォームです。Document AI Workbench3 を使用すると、ドキュメントの特定のセクションのテキストを解析するカスタム抽出ツールを作成できます。Natural Language API4 は、感情分析、エンティティ分析、その他のテキスト注釈などの自然言語理解テクノロジーを提供するサービスです。AnalyzeSentiment 機能5 は、指定されたテキストを検査し、テキスト内で一般的な感情的な意見を特定し、特に書き手の態度が肯定的、否定的、または中立であるかを判断します。したがって、オプション C は、各フォームの顧客のコメントから全体的な満足度スコアを予測するタスクを達成するための最良の方法です。他のオプションは、このシナリオには関係がなく、最適でもありません。

参考文献:

- \* Professional ML Engineer 試験ガイド
- \* ドキュメントAI
- \* ドキュメント AI ワークベンチ
- \* 自然言語 API
- \* 感情分析
- \* Google プロフェッショナル機械学習認定試験 2023
- \* 最新の Google Professional Machine Learning Engineer の実際の無料試験問題

**最新問題: 48**

You are training an LSTM-based model on AI Platform to summarize text using the following job submission script:

You want to ensure that training time is minimized without significantly compromising the accuracy of your model. What should you do?

- A. Modify the 'epochs' parameter
- B. Modify the 'scale-tier' parameter
- C. Modify the 'batch size' parameter
- D. Modify the 'learning rate' parameter

**Answer: B (メッセージを残す)**

機械学習モデルのトレーニング時間は、モデルの複雑さ、データのサイズ、ハードウェア リソース、ハイパーパラメーターなどのいくつかの要因によって異なります。モデルの精度を大幅に損

なうことなくトレーニング時間を最小限に抑えるには、これらの要素を可能な限り最適化する必要があります。

トレーニング時間に大きな影響を与える可能性がある要素の 1 つは、AI Platform でのトレーニングジョブに使用するマシンのタイプと数を指定するスケール層パラメーターです。スケール層パラメーターには、BASIC、STANDARD\_1、PREMIUM\_1、BASIC\_GPU などの事前定義値のいずれか、またはマシンタイプ、ワーカーの数、パラメーターサーバーの数を構成できるカスタム値を指定できます<sup>1</sup>。AI Platform で LSTM ベースのモデルをトレーニングする場合は、スケール層パラメーターを変更して、より多くの CPU、GPU、TPU など、より多くの計算リソースを提供する上位層またはカスタム構成を使用する必要があります。これにより、モデルトレーニングの並列性とスループットが向上し、トレーニング時間を短縮できます。ただし、より高い階層またはカスタム構成ではより高い料金が発生する可能性があるため、トレーニング時間とコストのトレードオフも考慮する必要があります<sup>2</sup>。他のオプションはそれほど効果的ではないか、モデルの精度に悪影響を与える可能性があります。モデルがデータセット全体を参照する回数を指定するエポックパラメーターを変更すると、トレーニング時間が短縮される可能性があります。モデルの収束とパフォーマンスにも影響します。バッチごとのサンプル数を指定するバッチサイズパラメーターを変更すると、モデルの安定性と汎化能力、メモリ使用量と勾配更新頻度に影響を与える可能性があります。勾配降下法最適化のステップサイズを指定する学習率パラメーターを変更すると、モデルの収束とパフォーマンスに影響を与えるだけでなく、オーバーシュートや極小値で行き詰まるリスクにも影響する可能性があります<sup>3</sup> 参考資料: 1: 事前定義されたマシンタイプの使用 2: 分散トレーニング3: ハイパーパラメータ調整の概要

#### 最新問題: 49

You are an ML engineer at a regulated insurance company. You are asked to develop an insurance approval model that accepts or rejects insurance applications from potential customers. What factors should you consider before building the model?

- A. Redaction, reproducibility, and explainability
- B. Traceability, reproducibility, and explainability
- C. Federated learning, reproducibility, and explainability
- D. Differential privacy federated learning, and explainability

**Answer: B (メッセージを残す)**

Before building an insurance approval model, an ML engineer should consider the factors of traceability, reproducibility, and explainability, as these are important aspects of responsible AI and fairness in a regulated domain. Traceability is the ability to track the provenance and lineage of the data, models, and decisions throughout the ML lifecycle. It helps to ensure the quality, reliability, and accountability of the ML system, and to comply with the regulatory and ethical standards. Reproducibility is the ability to recreate the same results and outcomes using the same data, models, and parameters. It helps to verify the validity, consistency, and robustness of the ML system, and to debug and improve the performance. Explainability is the ability to understand and interpret the logic, behavior, and outcomes of the ML system. It helps to increase the transparency, trust, and confidence of the ML system, and to identify and mitigate any

potential biases, errors, or risks. The other options are not as relevant or comprehensive as this option. Redaction is the process of removing sensitive or confidential information from the data or documents, but it is not a factor that the ML engineer should consider before building the model, as it is more related to the data preparation and protection.

Federated learning is a technique that allows training ML models on decentralized data without transferring the data to a central server, but it is not a factor that the ML engineer should consider before building the model, as it is more related to the model architecture and privacy preservation. Differential privacy is a method that adds noise to the data or the model outputs to protect the individual privacy of the data subjects, but it is not a factor that the ML engineer should consider before building the model, as it is more related to the model evaluation and deployment.

References:

- \* Responsible AI documentation
- \* Traceability documentation
- \* Reproducibility documentation
- \* Explainability documentation

#### 最新問題: 50

BigQuery に保存されたデータを使用して ML モデルをトレーニングしています。このデータには、個人を特定できる情報 (PII) とみなされるいくつかの値が含まれています。モデルをトレーニングする前に、データセットの感度を下げる必要があります。すべての列はモデルにとって重要です。どのように進めるべきでしょうか？

- A. Dataflow を使用して、機密データを含む列を BigQuery から取り込み、各機密列の値をランダム化します。
- B. Cloud Data Loss Prevention (DLP) API を使用して機密データをスキャンし、Dataflow と DLP API を使用して形式保持暗号化で機密値を暗号化します。
- C. Cloud Data Loss Prevention (DLP) API を使用して機密データをスキャンし、Dataflow を使用して暗号化アルゴリズム AES-256 とソルトを使用してすべての機密データを置き換えます。
- D. トレーニングの前に、BigQuery を使用して機密データを含まない列のみを選択します。権限のない個人が機密値にアクセスできないように、承認されたデータのビューを作成します。

**Answer: D (メッセージを残す)**

このアプローチでは、モデルをトレーニングする前に個人を特定できる情報 (PII) を削除することで、データセットの機密性を下げながら、データの重要な列を維持できます。データの承認されたビューを作成することで、権限のない個人が機密の値にアクセスできないようにすることができます。

#### 最新問題: 51

あなたはオンライン小売業者で働いています。あなたの会社にはライフサイクルの短い製品が数千あります。あなたの会社では、5年間分の売上データが BigQuery に保存されています。あなたは、各製品の月次売上予測を行うモデルを構築するように依頼されました。最小限の労力で迅速に実装できるソリューションを使用したいと考えています。あなたは何をするべきか？

- A. Vertex AI Training で Prophet を使用してカスタム モデルを構築します。
- B. Vertex AI Forecast を使用して、NN ベースのモデルを構築します。
- C. BigQuery ML を使用して統計 AR1MA\_PLUS モデルを構築します。
- D. Vertex AI Training で TensorFlow を使用してカスタム モデルを構築します。

**Answer: C** ([メッセージを残す](#))

Web 検索結果によると、BigQuery ML1 は SQL クエリを使用して BigQuery で機械学習モデルを作成および実行できるサービスです。BigQuery ML は、線形回帰、ロジスティック回帰、K 平均法クラスタリング、行列分解、ディープ ニューラル ネットワーク、時系列予測など、さまざまなタイプのモデルをサポートしています1。ARIMA\_PLUS2 は、BigQuery ML に組み込まれている時系列予測用の統計モデルです。ARIMA\_PLUS は、AutoRegressive Integrated Moving Average with exogenous regressors の略です。ARIMA\_PLUS は、ターゲット変数とその過去の値の間の関係、およびターゲット変数に影響を与える可能性のあるその他の外部要因をモデル化します。ARIMA\_PLUS は、複数の時系列、季節性、休日、欠損値を処理できます2。したがって、オプション C は、特定のユースケースに対して最小限の労力で迅速に実装できるソリューションを使用する最良の方法です。これにより、データを移動したりカスタムを記述したりせずに、SQL クエリを使用して BigQuery で予測モデルを構築して実行できます。コード。他のオプションは、このシナリオには関係がなく、最適でもありません。参考文献:

- \* BigQuery ML
- \* ARIMA\_PLUS
- \* Google プロフェッショナル機械学習認定試験 2023
- \* 最新の Google Professional Machine Learning Engineer の実際の無料試験問題

#### 最新問題: 52

あなたは、複数の地域に店舗を持つ大手食料品小売店の ML エンジニアです。在庫予測モデルを作成するように依頼されました。モデルの特徴には、地域、場所、過去の需要、季節的な人気が含まれます。アルゴリズムが毎日新しい在庫データから学習できるようにしたいと考えています。モデルの構築にはどのアルゴリズムを使用する必要がありますか？

- A. 分類
- B. 強化学習
- C. リカレント ニューラル ネットワーク (RNN)
- D. 畳み込みニューラル ネットワーク (CNN)

**Answer: C** ([メッセージを残す](#))

毎日新しい在庫データから学習するアルゴリズム」= 時系列モデル、時系列を扱うための最良のオプションは forsure RNN

<https://builtin.com/data-science/recurrent-neural-networks-and- lstm>

#### 最新問題: 53

あなたは、さまざまなユースケースに対応する最先端のアルゴリズムを開発している大手医療会社で働いています。カスタムラベルが付いた非構造化テキストデータを持っています。これらのラベルを使用してさまざまな医療用語を抽出して分類する必要があります。何をすべきですか？

- A. Healthcare Natural Language API を使用して医療エンティティを抽出します。
- B. BERT ベースのモデルを使用して、医療エンティティ抽出モデルを微調整します。
- C. AutoML Entity Extraction を使用して、医療エンティティ抽出モデルをトレーニングします。
- D. TensorFlow を使用して、カスタム医療エンティティ抽出モデルを構築します。

**Answer: B (メッセージを残す)**

医療エンティティの抽出は、電子医療記録、臨床ノート、研究論文などの非構造化テキスト データから医療用語や概念を特定して分類するタスクです。医療エンティティの抽出は、情報検索、知識発見、意思決定支援、データ分析などのさまざまなユースケースに役立ちます<sup>1</sup>。

医療エンティティ抽出を実行するために考えられるアプローチの 1 つは、BERT ベースのモデルを使用して医療エンティティ抽出モデルを微調整することです。BERT (Bidirectional Encoder Representations from Transformers) は、特定のトークンの左側と右側の両方からコンテキスト情報をキャプチャできる、事前トレーニングされた言語モデルです<sup>2</sup>。BERT は、事前トレーニングされたモデルの上にタスク固有のレイヤーを追加し、少量のラベル付きデータでモデル パラメータを更新することで、医療エンティティの抽出などの特定のタスクで微調整できます<sup>3</sup>。BERT ベースのモデルは、一般的なドメインのコーパスでの大規模な事前トレーニングとドメイン固有のデータでの微調整を活用することで、医療エンティティの抽出で高いパフォーマンスを達成できます。たとえば、Nesterov と umerenkov<sup>4</sup> は、大規模な EHR データセットで事前トレーニングされたトランスフォーマー モデルを微調整することにより、シングルステップのマルチラベル分類タスクとして電子医療記録から医療エンティティを抽出する新しい方法を提案しました。彼らは、そのモデルが最も頻繁に使用されるエンティティに対して人間レベルの品質を達成できることを示しました。

参考文献:

- \* 1: 事前トレーニング済みの言語モデルとドメイン辞書に基づく、ラベルのない医療記録からの医療固有実体認識 | データインテリジェンス | MIT プレス
- ※ 2: BERT : 言語理解のためのディープ双方向トランスフォーマーの事前学習
- \* 3: 医療エンティティ抽出のための BERT の微調整
- ※ 4 : 人間レベルの品質で電子医療記録から遠隔監視によるエンドツーエンドの医療エンティティを抽出

**最新問題: 54**

2 つのステップを含む Vertex AI パイプラインが作成されました。最初のステップでは、10 TB のデータの前処理が約 1 時間で完了し、結果を Cloud Storage バケットに保存します。2 番目のステップでは、処理されたデータを使用してモデルをトレーニングします。さまざまなアルゴリズムをテストできるように、モデルのコードを更新する必要があります。パイプラインの変更を最小限に抑えながら、パイプラインの実行時間とコストを削減します。

- A. データ前処理ステップ用に、コンピューティング最適化マシン ファミリからより多くの CPU と RAM を備えたマシンを構成します。
- B. パイプライン ジョブのキャッシュを有効にします。モデルトレーニングステップのキャッシュを無効にします。

C. パイプラインパラメーターと追加のパイプラインステップを追加します。パラメーター一値に応じて、パイプラインステップはデータの前処理を実行またはスキップし、モデルのトレーニングを開始します。

D. 前処理ステップなしで別のパイプラインを作成し、モデルトレーニング用に前処理された Cloud Storage ファイルの場所をハードコーディングします。

**Answer: B (メッセージを残す)**

#### 最新問題: 55

あなたは、最近 ML とデータのワークロードを Google Cloud に移行することを決定した大規模な組織で働いています。

データエンジニアリングチームは、構造化データを Avro 形式で Cloud Storage バケットにエクスポートしました。分析を実行し、機能を作成し、ML モデルがオンライン予測に使用する機能をホストするワークフローを提案する必要があります。パイプラインはどのように構成すればよいでしょうか？

A. Avro ファイルを Cloud Spanner に取り込んで分析を実行します。Dataflow パイプラインを使用して特徴を作成し、オンライン予測のために BigQuery に保存します。

B. Avro ファイルを BigQuery に取り込んで分析を実行します。Dataflow パイプラインを使用して特徴を作成し、オンライン予測のために Vertex AI Feature Store に保存します。

C. Avro ファイルを BigQuery に取り込んで分析を実行します。BigQuery SQL を使用して特徴を作成し、オンライン予測用に別の BigQuery テーブルに保存します。

D. Avro ファイルを Cloud Spanner に取り込んで分析を実行します。Dataflow パイプラインを使用して機能を作成します。オンライン予測のために Vertex AI Feature Store に保存します。

**Answer: B (メッセージを残す)**

BigQuery は、スケーラブルでコスト効率の高い方法で大量のデータの保存とクエリを可能にするサービスです。BigQuery を使用して、Cloud Storage バケットから Avro ファイルを取り込み、構造化データの分析を実行できます。Avro は、複雑なデータ型とスキーマを保存できるバイナリファイル形式です。bqload コマンドまたは BigQuery API を使用して、Avro ファイルを BigQuery テーブルにロードできます。その後、SQL クエリを使用してデータを分析し、洞察を生成できます。Dataflow は、Google Cloud 上でスケーラブルでポータブルなデータ処理パイプラインを作成して実行できるサービスです。Dataflow を使用すると、データの変換、集計、エンコードなど、ML モデルの機能を作成できます。Apache Beam SDK を使用して、Python または Java で Dataflow パイプラインコードを作成できます。組み込み変換またはカスタム変換を使用して、特徴エンジニアリングロジックをデータに適用することもできます。Vertex AI Feature Store は、ML 機能を Google Cloud に保存して管理できるサービスです。Vertex AI Feature Store を使用して、ML モデルがオンライン予測に使用する機能をホストできます。オンライン予測は、入力データの個別または小さなバッチに対して低遅延の応答を提供する予測の一種です。Vertex AI Feature Store API を使用して、Dataflow パイプラインからフィーチャストアエンティティタイプにフィーチャを書き込むことができます。その後、Vertex AI Feature Store オンラインサービング API を使用して、特徴ストアから特徴を読み取り、オンライン予測のために ML モデルに渡すことができます。

す。BigQuery、Dataflow、Vertex AI Feature Store を使用すると、分析を実行し、機能を作成し、ML モデルがオンライン予測に使用する機能をホストするパイプラインを構成できます。参考文献:

\* BigQuery ドキュメント

\* データフローのドキュメント

\* Vertex AI フィーチャー ストアのドキュメント

\* Google Cloud 認定資格の準備: 機械学習エンジニア プロフェッショナル認定資格

#### 最新問題: 56

あなたは自動車会社の AI チームで働いており、TensorFlow と Keras を使用して視覚的欠陥検出モデルを開発しています。モデルのパフォーマンスを向上させるには、変換、トリミング、コントラスト調整などの画像拡張機能を組み込む必要があります。これらの関数を各トレーニング バッチにランダムに適用します。実行時とコンピューティング リソースの使用率を考慮してデータ処理パイプラインを最適化したいと考えています。あなたは何をするべきか？

A. Dataflow を使用して、考えられるすべての拡張を作成し、TFRecord として保存します。

B. Dataflow を使用して、トレーニングの実行ごとに拡張を動的に作成し、TFRecord としてステージングします。

C. 拡張関数を tf.Data パイプラインに動的に埋め込みます。

D. Keras ジェネレーターの一部として拡張関数を動的に埋め込みます。

**Answer: A (メッセージを残す)**

#### 最新問題: 57

100 を超える入力特徴 (すべての値が -1 から 1 まで) を含む線形モデルを構築しています。多くの特徴は有益ではないと思われます。有益な特徴を元の形式に保ちながら、非有益な特徴をモデルから削除したいと考えています。どのテクニックを使用する必要がありますか？

A. 主成分分析を使用して、情報量が最も少ない特徴を削除します。

B. L1 正則化を使用して、有益でない特徴の係数を 0 に削減します。

C. モデルを構築した後、Shapley 値を使用して、どの特徴が最も有益であるかを判断します。

D. 反復ドロップアウト手法を使用して、削除時にモデルが劣化しないフィーチャを特定します。

**Answer: B (メッセージを残す)**

<https://cloud.google.com/ai-platform/prediction/docs/ai-explanations/overview#sampled-shapley>

#### 最新問題: 58

機械学習スペシャリストは、ROC 曲線下面積 (AUC) を客観的な指標として Amazon SageMaker を使用して、ツリーベースのアンサンブルモデルのハイパーパラメータ調整ジョブを開始します。このワークフローは最終的に、毎晩ハイパーパラメータを再トレーニングおよび調整するパイプラインにデプロイされ、24 時間ごとに失われるデータのクリックスルーをモデル化します。これらのモデルのトレーニングにかかる時間を短縮し、最終的にコストを削減することを目的として、スペシャリストは入力ハイパーパラメータ範囲を再構成したいと考えています。

これを実現できるのはどのビジュアライゼーションですか？

A. 各トレーニング反復にわたる目標メトリクスのパフォーマンスを示す散布図。

- B. t 分布確率的近傍埋め込み (t-SNE) を使用して多数の入力変数を読みやすい次元で視覚化する、ターゲット変数によって色付けされた点を含む散布図。
- C. ツリーの最大深さと客観的なメトリックの間の相関関係を示す散布図。
- D. 最も重要な入力特徴がガウスであるかどうかを示すヒストグラム。

**Answer: B (メッセージを残す)**

#### 最新問題: 59

カスタム TensorFlow DNN リグレッサー モデルを使用して BigQuery テーブル内の 1 億レコードに対してバッチ予測を実行し、予測結果を BigQuery テーブルに保存する必要があります。この推論パイプラインの構築に必要な労力を最小限に抑えたいと考えています。あなたは何をするべきか？

- A. BigQuery ML を使用して TensorFlow モデルをインポートし、ml.predict 関数を実行します。
- B. TensorFlow BigQuery リーダーを使用してデータを読み込み、BigQuery API を使用して結果を BigQuery に書き込みます。
- C. BigQuery のデータを TFRecord に変換する Dataflow パイプラインを作成します。Vertex AI Prediction でバッチ推論を実行し、結果を BigQuery に書き込みます。
- D. TensorFlow SavedModel を Dataflow パイプラインにロードします。BigQuery I/O コネクタとカスタム関数を使用してパイプライン内で推論を実行し、結果を BigQuery に書き込みます。

**Answer: A (メッセージを残す)**

\* 選択肢 A は正解です。BigQuery ML で TensorFlow モデルをインポートし、ml.predict 関数を実行するのが、カスタム TensorFlow モデルを使用して大規模な BigQuery テーブルに対してバッチ予測を実行し、予測結果を別の BigQuery テーブルに保存する最も簡単な方法です。BigQuery ML を使用すると、Cloud Storage に保存されている TensorFlow モデルをインポートし、SQL クエリによる予測に使用できます<sup>1</sup>。ml.predict 関数は、予測値を含むテーブルを返します。このテーブルは、別の BigQuery テーブル 2 に保存できます。

\* オプション B は不正解です。TensorFlow BigQuery リーダーを使用してデータを読み込み、BigQuery API を使用して結果を BigQuery に書き込むには、オプション A よりも推論パイプラインの構築に多くの労力が必要です。TensorFlow BigQuery リーダーは、データを読み取る方法です。BigQuery を TensorFlow データセットに変換し、トレーニングや予測に使用できます<sup>3</sup>。ただし、このオプションでは、TensorFlow モデルをロードし、予測を実行し、BigQuery API を使用して結果を BigQuery<sup>4</sup> に書き戻すコードを記述する必要もあります。

\* オプション C は不正解です。Dataflow パイプラインを作成して BigQuery のデータを TFRecord に変換し、Vertex AI Prediction でバッチ推論を実行し、結果を BigQuery に書き込むには、オプション A よりも推論パイプラインの構築に多くの労力が必要です。Dataflow はサービスです。ETL (抽出、変換、ロード) やバッチ処理などのデータ処理パイプラインの作成と実行用<sup>5</sup>。Vertex AI Prediction は、オンラインまたはバッチ予測用の ML モデルをデプロイおよび提供するためのサービスです。ただし、このオプションでは、Dataflow パイプラインを作成し、データを TFRecord に変換し、バッチ推論を実行し、結果を BigQuery に書き込むためのコードを記述する必要もあります。

\* オプション D は不正解です。TensorFlow SavedModel を Dataflow パイプラインにロードし、カスタム関数を備えた BigQuery I/O コネクタを使用してパイプライン内で推論を実行し、結果を BigQuery に書き込むには、オプションよりも推論パイプラインの構築に多くの労力が必要です。A. BigQuery I/O コネクタは、Dataflow パイプライン内で BigQuery からデータを読み書きする方法です。ただし、このオプションでは、TensorFlow SavedModel を読み込み、推論用のカスタム関数を作成し、結果を BigQuery に書き込むコードを記述する必要もあります。

参考文献:

- \* BigQuery ML へのモデルのインポート
- \* インポートされたモデルを予測に使用する
- \* TensorFlow BigQuery リーダー
- \* BigQuery API
- \* データフローの概要
- \* [Vertex AI Prediction 概要]
- \* [Dataflowによるバッチ予測]
- \* [BigQuery I/O コネクタ]
- \* [Dataflow での TensorFlow モデルの使用]

#### 最新問題: 60

オンライン予測に使用されるカスタム TensorRow モデルを開発する必要があります。トレーニング データは BigQuery に保存されます。モデルのトレーニングと提供のために、インスタンスレベルのデータ変換をデータに適用する必要があります。モデルのトレーニングと提供中に同じ前処理ルーチンを使用したいと考えています。前処理ルーチンをどのように構成すればよいでしょうか？

- A. BigQuery からデータを読み取って変換する前処理関数を作成します。サービス提供時に前処理関数を呼び出す Vertex AI カスタム予測ルーチンを作成します。
- B. Apache Beam パイプラインを作成して、BigQuery からデータを読み取り、TensorFlow Transform と Dataflow を使用して前処理します。
- C. Vertex AI Pipelines でパイプラインを作成し、BigQuery からデータを読み取り、カスタム前処理コンポーネントを使用してデータを前処理します。
- D. BigQuery スクリプトを作成してデータを前処理し、結果を別の BigQuery テーブルに書き込みます。

**Answer: A** ([メッセージを残す](#))

#### 最新問題: 61

Your organization manages an online message board A few months ago, you discovered an increase in toxic language and bullying on the message board. You deployed an automated text classifier that flags certain comments as toxic or harmful. Now some users are reporting that benign comments referencing their religion are being misclassified as abusive Upon further inspection, you find that your classifier's false positive rate is higher for comments that reference

certain underrepresented religious groups. Your team has a limited budget and is already overextended. What should you do?

- A. Add synthetic training data where those phrases are used in non-toxic ways
- B. Remove the model and replace it with human moderation.
- C. Replace your model with a different text classifier.
- D. Raise the threshold for comments to be considered toxic or harmful

**Answer: A (メッセージを残す)**

The problem of the text classifier is that it has a high false positive rate for comments that reference certain underrepresented religious groups. This means that the classifier is not able to distinguish between toxic and non-toxic language when those groups are mentioned. One possible reason for this is that the training data does not have enough examples of non-toxic comments that reference those groups, leading to a biased model.

Therefore, a possible solution is to add synthetic training data where those phrases are used in non-toxic ways, which can help the model learn to generalize better and reduce the false positive rate. Synthetic data is artificially generated data that mimics the characteristics of real data, and can be used to augment the existing data when the real data is scarce or imbalanced.

References:

\* Preparing for Google Cloud Certification: Machine Learning Engineer, Course 5: Responsible AI, Week

3: Fairness

\* Google Cloud Professional Machine Learning Engineer Exam Guide, Section 4: Ensuring solution quality, 4.4 Evaluating fairness and bias in ML models

\* Official Google Cloud Certified Professional Machine Learning Engineer Study Guide, Chapter 9:

Responsible AI, Section 9.3: Fairness and Bias

有効な **Professional-Machine-Learning-Engineer** 問題集は GoShiken.com が提供された合格しやすい Professional-Machine-Learning-Engineer 試験問題集！ GoShiken.com が最新の **Professional-Machine-Learning-Engineer** 試験問題集を提供しています。GoShiken.com Professional-Machine-Learning-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Professional-Machine-Learning-Engineer 問題集をゲットする人はこちら:

<https://www.goshiken.com/Google/Professional-Machine-Learning-Engineer-mondaishu.html>

(**29030%OFF**問題集溶と正解付きで **30%w** 特別割引コード: **Freepdfdumps**)

最新問題: 62

あなたは、大規模なデータセットを使用して会社のカスタム言語モデルをトレーニングしています。Vertex AI で Reduction Server 戦略を使用する予定です。分散トレーニング ジョブのワーカープールを構成する必要があります。あなたは何をするべきか？

- A.** 最初の 2 つのワーカー プールのマシンが GPU を備え、トレーニング コードが実行されるコンテナ イメージを使用するように構成します。3 番目のワーカー プールが GPU を持つように構成し、reductionsserver コンテナ イメージを使用します。
- B.** 最初の 2 つのワーカー プールのマシンが GPU を備え、トレーニング コードが実行されるコンテナ イメージを使用するように構成します。アクセラレータなしで、reductionsserver コンテナ イメージを使用するように 3 番目のワーカー プールを構成し、帯域幅を優先するマシン タイプを選択します。
- C.** 最初の 2 つのプールのマシンが TPU を持つように構成します。トレーニング コードが実行されるコンテナ イメージを使用するには、TPU を持つように 3 番目のプールを構成し、reduccionserver コンテナ イメージを使用します。
- D.** 最初の 2 つのワーカー プールのマシンが TPU を持ち、トレーニング コードが実行されるコンテナ イメージを使用するように構成します。アクセラレータなしで 3 番目のワーカー プールを構成し、アクセラレータなしで reduccinserver コンテナ イメージを使用し、マシン タイプを選択します。帯域幅を優先します。

**Answer: B** ([メッセージを残す](#))

#### 最新問題: 63

あなたは、オンライン フォーラムを管理するペットフード会社で働いています。顧客は自分のペットの写真をフォーラムにアップロードして、他の人と共有します。毎日約 20 枚の写真がアップロードされます。アップロードされた各写真に動物が含まれているかどうかを、ほぼリアルタイムで自動的に検出したいと考えています。時間を優先し、アプリケーションの開発と展開のコストを最小限に抑えるにはどうすればよいでしょうか？

- A.** 動物の周囲の境界ボックスを使用して、以前に送信された画像に手動でラベルを付けます。Vertex AI を使用して AutoML オブジェクト検出モデルを構築します。モデルを Vertex AI エンドポイントにデプロイします。新しいユーザーが送信した画像をモデル エンドポイントに送信して、各写真が正しいかどうかを検出します。動物を飼っています。
- B.** TensorFlow Hub から物体検出モデルをダウンロードします。モデルを Vertex AI エンドポイントにデプロイします。ユーザーが送信した新しい画像をモデル エンドポイントに送信して、各写真に動物が写っているかどうかを分類します。
- C.** 以前に送信された画像に動物がいるかどうかを手動でラベル付けします。Vertex AI で画像データセットを作成します。Vertex AutoML を使用して分類モデルをトレーニングし、2 つのクラスを区別します。モデルを Vertex AI エンドポイントにデプロイします。新しいユーザーが送信した画像を送信します。モデルのエンドポイントに送信して、各写真に動物が写っているかどうかを分類します。
- D.** ユーザーが送信した画像を Cloud Vision API に送信します。オブジェクト ローカリゼーションを使用して画像内のすべてのオブジェクトを識別し、結果を動物のリストと比較します。

**Answer: B** ([メッセージを残す](#))

#### 最新問題: 64

あなたはソーシャルメディア会社で働いています。投稿された画像に車が含まれているかどうかを検出する必要があります。各トレーニング サンプルは、正確に 1 つのクラスのメンバーです。物体検出ニューラル ネットワークをトレーニングし、評価のためにモデル バージョンを AI Platform Prediction にデプロイしました。デプロイメントの前に、評価ジョブを作成し、それを AI Platform Prediction モデル バージョンに添付しました。精度がビジネス要件よりも低いことに気づきました。精度を高めるためにモデルの最終層のソフトマックスしきい値をどのように調整すればよいでしょうか？

- A. リコールを増やす
- B. リコールを減らします。
- C. 誤検知の数を増やす
- D. 偽陰性の数を減らします。

**Answer: B (メッセージを残す)**

適合率と再現率は、分類モデルのパフォーマンスを評価するための 2 つの一般的な指標です。適合率は正しい陽性予測の割合を測定し、再現率は正しく予測された陽性例の割合を測定します。適合率と再現率は反比例の関係にあり、一方を増加させると他方も減少し、その逆も同様です。精度と再現率のトレードオフは、分類問題の目標とコストによって異なります<sup>1</sup>。

投稿された画像に車が含まれているかどうかを検出するユースケースでは、ソーシャルメディア会社は誤検知、つまり車が含まれていると誤ってラベル付けされた画像の数を最小限に抑えたいと考えているため、再現率よりも精度が重要です。精度が高いということは、モデルが肯定的な予測に自信があり正確であることを意味しますが、再現率が低いということは、モデルがいくつかの肯定的な例や実際に車を含む画像を見逃す可能性があることを意味します。いくつかの肯定的な例を見逃すコストは、誤った肯定的な予測を行うコストよりも低くなります。後者は、ユーザーエクスペリエンスとソーシャルメディア会社の評判に影響を与える可能性があるためです。

ソフトマックス関数は、実数のベクトルを可能なクラスにわたる確率分布に変換する関数です。ソフトマックス関数は、各クラスに確率を割り当て、最も高い確率を持つクラスが予測として選択されるため、マルチクラス分類問題のニューラル ネットワークの最終層としてよく使用されます。ソフトマックス関数は次のように定義されます。

ソフトマックス  $(x_i) = \exp(x_i) / \sum_j \exp(x_j)$

ここで、 $x_i$  はクラス  $i$  の入力値、 $\text{softmax}(x_i)$  はクラス  $i$  の出力確率です。

ソフトマックスしきい値は、クラスが予測として選択される必要がある最小確率を決定するパラメーターです。たとえば、ソフトマックスしきい値が 0.5 の場合、最も高い確率を持つクラスが選択されるには少なくとも 0.5 が必要です。そうでない場合、予測は行われません。ソフトマックスしきい値は適合率と再現率のトレードオフを調整するために使用できます。しきい値を高くすると適合率が向上し、再現率が低下します。一方、しきい値を低くすると適合率が低下し、再現率が増加します<sup>2</sup>。

投稿された画像に車が含まれているかどうかを検出するユースケースの場合、モデルの最終レイヤーのソフトマックスしきい値を調整して精度を高める最良の方法は、再現率を下げることで、これは、モデルが信頼性の高い場合にのみ肯定的な予測を行い、誤検出を避けるために、ソフトマックスしきい値を増やす必要があることを意味します。ソフトマックスしきい値を増やすこ

とで、モデルの肯定的な予測がより選択的かつ正確になり、精度メトリクスが向上します。したがって、この使用例では再現率を下げるのが最良の選択肢となります。

参考文献:

\* 精度と再現率 - Wikipedia

\* ソフトマックススコアにしきい値を追加する方法 - コードログ

#### 最新問題: 65

あなたの会社は、ユーザーがビデオを視聴したりアップロードしたりできるビデオ共有 Web サイトを管理しています。ML モデルを作成して、新しくアップロードされたどのビデオが最も人気があるかを予測し、それらのビデオを会社の Web サイトで優先的に表示できるようにする必要があります。モデルが成功したかどうかを判断するには、どの結果を使用する必要がありますか？

A. このモデルでは、動画をアップロードしたユーザーの「いいね！」が 10,000 件を超えている場合、その動画は人気があると予測します。

B. このモデルは、クリック数で測定した最も人気のあるクリックベイト動画の 97.5% を予測します。

C. このモデルは、アップロードされてから 30 日以内の総再生時間で測定された最も人気のある動画の 95% を予測します。

D. 公開後 7 日後と 30 日後の対数変換されたビュー数の間のピアソン相関係数は 0 に等しくなります。

**Answer: C (メッセージを残す)**

<https://developers.google.com/machine-learning/problem-framing/framing#quantify-it>

#### 最新問題: 66

Compute Engine 上の GPU を搭載した仮想マシンを使用して、特定の画像に存在する政府 ID のタイプを予測するコンピュータービジョンモデルをトレーニングする必要があります。次のパラメータを使用します。

\* オプティマイザー: SGD

※ 画像形状 = 224 × 224

\* バッチサイズ = 64

\* エポック = 10

\* 詳細 = 2

トレーニング中に次のエラーが発生します: ResourceExhaustedError: tensor を割り当てるときにメモリ不足 (oom)。あなたは何をするべきか？

A. オプティマイザを変更します

B. バッチサイズを減らす

C. 学習率を変更します

D. 画像の形状を縮小します。

**Answer: (解答を表示する)**

参照 :

<https://stackoverflow.com/questions/59394947/how-to-fix-resourceexhaustederror-oom-when-allocating-tensor/59395251#:~:text=OOM%20stands%20for%20%22out%20of,in%20your%20Dense%20%2C%20Conv2D%20レイヤー>

#### 最新問題: 67

テキスト分類モデルをトレーニングしました。次の SignatureDef があります。

予測リクエストの正しい書き方は何ですか？

- A. `data = json.dumps({"signature_name": f,serving_default", "instances": [['a', 'b'], [c\ 'd'], [e\ T ] ]})`
- B. `data = json.dumps({"signature_name": "serving_default"! "instances": [['a', 'b', "c", 'd', 'e', 'f'] ]})`
- C. `data = json.dumps({"signature_name": "serving_default"\ "instances": [fab', 'be1, 'cd']])`
- D. `data = json.dumps({"signature_name": "serving_default, "instances": [['a', 'b\ 'c'1, [d\ 'e\ T']])`

**Answer:** ([解答を表示する](#))

#### 最新問題: 68

A Machine Learning Specialist is assigned a TensorFlow project using Amazon SageMaker for training, and needs to continue working for an extended period with no Wi-Fi access.

Which approach should the Specialist use to continue working?

- A. Install Python 3 and boto3 on their laptop and continue the code development using that environment.
- B. Download the TensorFlow Docker container used in Amazon SageMaker from GitHub to their local environment, and use the Amazon SageMaker Python SDK to test the code.
- C. Download TensorFlow from tensorflow.org to emulate the TensorFlow kernel in the SageMaker environment.
- D. Download the SageMaker notebook to their local environment, then install Jupyter Notebooks on their laptop and continue the development in a local notebook.

**Answer: A** ([メッセージを残す](#))

Explanation

#### 最新問題: 69

Vertex AI を使用してカスタム モデルを開発し、過去の取引データに基づいて自社製品の売上を予測した 近い将来の特徴分布と特徴間の相関関係の変化を予測した 大量の予測を受け取ることも期待したドリフト検出に Vertex AI Model Monitoring を使用する予定で、コストを最小限に抑えたいと考えています。あなたは何をするべきか？

- A. モニタリングに機能と機能属性を使用します。監視頻度の値をデフォルトよりも低く設定します。
- B. 監視用の機能を使用します。予測サンプリングレア値を 0 よりも 1 に近い値に設定します。
- C. 監視機能を使用する 監視頻度の値をデフォルトよりも高く設定します。
- D. 特徴量と特徴量属性を監視に使用します。予測サンプリング レートの値を 1 よりも 0 に近い値に設定します。

**Answer:** ([解答を表示する](#))

### 最新問題: 70

あなたは銀行に勤めており、不正行為検出のためのランダム フォレスト モデルを構築しています。トランザクションを含むデータセットがあり、そのうち 1% が不正であると特定されています。どのデータ変換戦略が分類器のパフォーマンスを向上させる可能性がありますか？

- A. 不正なトランザクションを 10 回オーバーサンプリングします。
- B. データを TFRecord に書き込みます。
- C. すべてのカテゴリ特徴量でワンホット エンコーディングを使用します。
- D. すべての数値特徴を Z 正規化します。

**Answer: A** ([メッセージを残す](#))

### 最新問題: 71

You have trained a DNN regressor with TensorFlow to predict housing prices using a set of predictive features. Your default precision is tf.float64, and you use a standard TensorFlow estimator; estimator = tf.estimator.DNNRegressor( feature\_columns=[YOUR\_LIST\_OF\_FEATURES], hidden\_units=[1024, 512, 256], dropout=None) Your model performs well, but Just before deploying it to production, you discover that your current serving latency is 10ms @ 90 percentile and you currently serve on CPUs. Your production requirements expect a model latency of 8ms @ 90 percentile. You are willing to accept a small decrease in performance in order to reach the latency requirement Therefore your plan is to improve latency while evaluating how much the model's prediction decreases. What should you first try to quickly lower the serving latency?

- A. Increase the dropout rate to 0.8 in \_PREDICT mode by adjusting the TensorFlow Serving parameters
- B. Increase the dropout rate to 0.8 and retrain your model.
- C. Switch from CPU to GPU serving
- D. Apply quantization to your SavedModel by reducing the floating point precision to tf.float16.

**Answer: (解答を表示する)**

\* Quantization is a technique that reduces the numerical precision of the weights and activations of a neural network, which can improve the inference speed and reduce the memory footprint of the model<sup>1</sup>.

\* Reducing the floating point precision from tf.float64 to tf.float16 can potentially halve the latency and memory usage of the model, while having minimal impact on the accuracy<sup>2</sup>.

\* どちらのモードでもドロップアウト率を 0.8 に増やしても、レイテンシには影響しませんが、モデルのパフォーマンスが大幅に低下する可能性があります。ドロップアウトは、過学習を防ぐためにトレーニング中にユニットをランダムにドロップする正則化手法であるためです。

\* CPU から GPU サービスに切り替えると、ハードウェア仕様とモデルの複雑さに応じて遅延が改善される場合と改善されない場合がありますが、導入に追加のコストと複雑さが発生する可能性もあります<sup>4</sup>

### 最新問題: 72

You are working on a Neural Network-based project. The dataset provided to you has columns with different ranges. While preparing the data for model training, you discover that gradient optimization is having difficulty moving weights to a good solution. What should you do?

- A. Use feature construction to combine the strongest features.
- B. Improve the data cleaning step by removing features with missing values.
- C. Use the representation transformation (normalization) technique.
- D. Change the partitioning step to reduce the dimension of the test set and have a larger training set.

**Answer:** ([解答を表示する](#))

### 最新問題: 73

Compute Engine 上の GPU を搭載した仮想マシンを使用して、特定の画像に存在する政府 ID のタイプを予測するコンピュータービジョンモデルをトレーニングする必要があります。次のパラメータを使用します。

\* オプティマイザー: SGD

※ 画像形状 = 224 × 224

\* バッチサイズ = 64

\* エポック = 10

\* 詳細 = 2

トレーニング中に次のエラーが発生します: ResourceExhaustedError: tensor を割り当てるときにメモリ不足 (oom)。あなたは何をするべきか？

- A. 画像の形状を縮小します。
- B. オプティマイザを変更します
- C. バッチサイズを減らす
- D. 学習率を変更する

**Answer:** B ([メッセージを残す](#))

### 最新問題: 74

あなたは最近、概念実証 (POC) 深層学習モデルを作成しました。全体的なアーキテクチャには満足していますが、いくつかのハイパーパラメータの値を決定する必要があります。Vertex AI でハイパーパラメータ調整を実行して、モデルで使用されるカテゴリ特徴量の適切な埋め込み次元と最適な学習率の両方を決定したいと考えています。次の設定を構成します。

埋め込みディメンションの場合、タイプを INTEGER に設定し、minValue は 16、maxValue は 64 に設定します。

学習率については、タイプを DOUBLE に設定し、minValue を 10e-05、maxValue を 10e-02 に設定します。

デフォルトのベイズ最適化チューニングアルゴリズムを使用しており、モデルの精度を最大化したいと考えています。トレーニング時間は問題ではありません。各ハイパーパラメータのハイパーパラメータ スケーリングと maxParallelTrials をどのように設定すればよいでしょうか？

- A. 埋め込みディメンションには UNIT\_LOG\_SCALE を、学習率には UNIT\_LINEAR\_SCALE を、少数の並列試行には使用します。
- B. 埋め込みディメンションには UNIT\_LINEAR\_SCALE、学習率には UNIT\_LOG\_SCALE、少数の並列試行には使用します。
- C. 埋め込みディメンションには UNIT\_LOG\_SCALE、学習率には UNIT\_LINEAR\_SCALE、および多数の並列試行には UNIT\_LINEAR\_SCALE を使用します。
- D. 埋め込みディメンションには UNIT\_LINEAR\_SCALE、学習率には UNIT\_LOG\_SCALE、および多数の並列試行には UNIT\_LOG\_SCALE を使用します。

**Answer: B** ([メッセージを残す](#))

#### 最新問題: 75

機械学習スペシャリストは、複数のデータサイエンティストがノートブックにアクセスし、モデルをトレーニングし、エンドポイントをデプロイできるように Amazon SageMaker を構成しています。最高の運用パフォーマンスを確保するには、スペシャリストは、科学者がモデルをデプロイする頻度、デプロイされた SageMaker エンドポイントの GPU と CPU の使用率、およびエンドポイントの呼び出し時に生成されるすべてのエラーを追跡する必要があります。

この情報を追跡するために Amazon SageMaker と統合されているのはどのサービスですか? (2つお選びください。)

- A. AWS CloudTrail
- B. AWS ヘルス
- C. AWS Trusted Advisor
- D. Amazon CloudWatch
- E. AWS 設定

**Answer: (**[解答を表示する](#)**)**

説明/参照: <https://aws.amazon.com/sagemaker/faqs/>

#### 最新問題: 76

あなたは、最近 ML とデータのワークロードを Google Cloud に移行することを決定した大規模な組織で働いています。データ エンジニアリング チームは、構造化データを Avro 形式で Cloud Storage バケットにエクスポートしました。分析を実行し、機能を作成し、ML モデルがオンライン予測に使用する機能をホストするワークフローを提案する必要があります。パイプラインはどのように構成すればよいでしょうか?

- A. Avro ファイルを BigQuery に取り込んで分析を実行します。Dataflow パイプラインを使用して特徴を作成し、オンライン予測のために Vertex AI Feature Store に保存します。
- B. Avro ファイルを Cloud Spanner に取り込んで分析を実行します。Dataflow パイプラインを使用して機能を作成します。オンライン予測のために Vertex AI Feature Store に保存します。
- C. Avro ファイルを BigQuery に取り込んで分析を実行します。BigQuery SQL を使用して特徴を作成し、オンライン予測用に別の BigQuery テーブルに保存します。
- D. Avro ファイルを Cloud Spanner に取り込んで分析を実行します。Dataflow パイプラインを使用して特徴を作成し、オンライン予測のために BigQuery に保存します。

**Answer: A** ([メッセージを残す](#))

有効な **Professional-Machine-Learning-Engineer** 問題集は GoShiken.com が提供された合格しやすい Professional-Machine-Learning-Engineer 試験問題集！ GoShiken.com が最新の **Professional-Machine-Learning-Engineer** 試験問題集を提供しています。GoShiken.com Professional-Machine-Learning-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Professional-Machine-Learning-Engineer 問題集をゲットする人はこちら：

<https://www.goshiken.com/Google/Professional-Machine-Learning-Engineer-mondaishu.html>

(**29030%OFF**問題集溶と正解付きで **30%w**特別割引コード: **Freepdfdumps**)

**最新問題: 77**

データサイエンスチームは、スケジュールされたモデルの再トレーニングをサポートするシステム、Docker コンテナ、およびオンライン予測リクエストの自動スケーリングと監視をサポートするサービスをリクエストしました。このシステムにはどのプラットフォームコンポーネントを選択する必要がありますか？

- A. Vertex AI パイプラインと AI プラットフォーム予測
- B. Cloud Composer、BigQuery ML、および AI プラットフォームの予測
- C. Cloud Composer、カスタム コンテナを使用した AI Platform Training、App Engine
- D. Vertex AI パイプラインと App Engine

**Answer: A** ([メッセージを残す](#))

**最新問題: 78**

カテゴリカル入力変数を含むデータセットを使用して ML モデルを開発しています。データの半分をトレーニングセットとテストセットにランダムに分割しました。トレーニングセット内のカテゴリ変数にワンホットエンコーディングを適用した後、テストセットから1つのカテゴリ変数が欠落していることがわかりました。あなたは何をすべきか？

- A. トレーニングセットに70%、テストセットに30%を使用して、データをランダムに再配布します。
- B. テストセットでスパース表現を使用します
- C. テストデータ内のカテゴリ変数にワンホットエンコーディングを適用します。
- D. すべてのカテゴリを表すさらに多くのデータを収集します

**Answer: C** ([メッセージを残す](#))

テストセット内の欠落しているカテゴリ変数に対処するための最良のオプションは、テストデータ内のカテゴリ変数にワンホットエンコーディングを適用することです。このオプションには次の利点があります。

\* ワンホットエンコーディングはカテゴリ変数をモデルで簡単に処理できるバイナリベクトルに変換するため、MLモデルのデータ形式の一貫性と互換性が保証されます。テストデータ内のカテ

ゴリ変数にワンホット エンコーディングを適用すると、テスト データ内の特徴の数と順序をトレーニング データと一致させ、モデル予測におけるエラーや不一致を回避できます。

\* ワンホット エンコーディングはカテゴリ変数の考えられる値ごとに個別の特徴を作成し、実際の値に対応する特徴に値 1 を割り当てるため、ML モデルのデータの情報と関連性が保持されず。変数、残りは 0。テスト データ内のカテゴリ変数にワンホット エンコーディングを適用すると、カテゴリ変数の元の意味と重要性を保持し、データの損失や歪みを回避できます。

他のオプションは、次の理由からあまり最適ではありません。

\* オプション A: データをトレーニング セットに 70%、テスト セットに 30% としてランダムに再配布すると、さらなる複雑さとリスクが生じます。このオプションでは、データを再度シャッフルして分割する必要があり、面倒で時間がかかる可能性があります。さらに、このオプションは、データ分布のランダム性に依存するため、欠落しているカテゴリ変数がテスト セットに存在することを保証しない場合があります。さらに、このオプションは、モデルが元のトレーニング セットから学習したデータの特徴やパターンを変更する可能性があるため、ML モデルの品質と有効性に影響を与える可能性があります。

\* オプション B: テスト セットでスパース表現を使用すると、追加のオーバーヘッドと非効率が生じます。

\* このオプションでは、テスト セット内のカテゴリ変数をスパース ベクトルに変換する必要があります。スパース ベクトルとは、ほとんどの値がゼロであり、ゼロ以外の要素のインデックスと値のみを格納するベクトルです。

ただし、モデルでは入力データがトレーニング データと同じ形式と次元を持つことを想定しており、ワンホット エンコーディングを使用するため、テスト セットでのスパース表現の使用は ML モデルと互換性がない可能性があります。さらに、テスト セットでスパース表現を使用すると、スパース ベクトルを保存および処理するために追加の計算とメモリが必要になるため、効率的またはスケーラブルではない可能性があります。

\* オプション D: すべてのカテゴリを表すより多くのデータを収集すると、追加のコストと遅延が発生します。このオプションでは、欠落しているカテゴリ変数を含むさらに多くのデータを取得してラベルを付ける必要があります。コストと時間がかかる可能性があります。さらに、データ ソースやビジネス上の問題によっては、欠落しているカテゴリ変数が利用できないか、テスト データに関連しない可能性があるため、このオプションは実行不可能または必要でない場合があります。

#### 最新問題: 79

あなたは小売会社に勤めています。Vertex AI には、3 つの異なる店舗からの売上データを含む管理された表形式データセットがあります。データセットには、店舗名や販売タイムスタンプなどのいくつかの特徴が含まれています。そのデータを使用して、間もなくオープンする新しい店舗の売上予測を行うモデルをトレーニングしたいと考えています。トレーニング、検証、およびテスト セットの間でデータを分割する必要があります。データを分割するにはどのようなアプローチを使用する必要がありますか？

A. Vertex AI 手動分割を使用し、ストア名機能を使用して各セットに 1 つのストアを割り当てます。

B. Vertex AI のデフォルトのデータ分割を使用します。

C. Vertex AI 時系列分割を使用し、販売タイムスタンプ機能をタイムバナブルとして指定します。

D. Vertex AI ランダム分割を使用して、行の 70% をトレーニング セットに割り当て、10% を検証 セットに割り当てます。

テストセットの 20%。

**Answer: B (メッセージを残す)**

3 つの異なる店舗からの売上データを含む Vertex AI の管理された表形式データセットを使用して、トレーニング、検証、テスト セット間でデータを分割するための最良のオプションは、Vertex AI のデフォルト データ分割を使用することです。

このオプションを使用すると、Vertex AI のパワーとシンプルさを活用して、データをパーセンテージで 3 つのセットに自動的かつランダムに分割できます。Vertex AI は、Google Cloud 上で機械学習ソリューションを構築およびデプロイするための統合プラットフォームです。Vertex AI は、線形回帰、ロジスティック回帰、K 平均法クラスタリング、行列因数分解、ディープ ニューラル ネットワークなど、さまざまなタイプのモデルをサポートできます。Vertex AI は、データ分析、モデル開発、モデル展開、モデル監視、モデル ガバナンスのためのさまざまなツールとサービスも提供できます。デフォルトのデータ分割は、Vertex AI によって提供されるデータ分割方法であり、ユーザーの入力や構成は必要ありません。デフォルトのデータ分割は、ランダム サンプリング方法を使用してデータをトレーニング、検証、テスト セットに分割し、データの固定パーセンテージを各セットに割り当てての役に立ちます。デフォルトのデータ分割はデータ分割プロセスを簡素化するのに役立ち、ほとんどの場合にうまく機能します。

トレーニング セットは、モデルをトレーニングし、モデル パラメーターを調整するために使用されるデータのサブセットです。トレーニング セットは、入力特徴とターゲット変数の間の関係を学習し、モデルのパフォーマンスを最適化するのに役立ちます。検証セットは、モデルを検証し、モデルのハイパーパラメーターを調整するために使用されるデータのサブセットです。検証セットは、目に見えないデータに対するモデルのパフォーマンスを評価し、過剰適合または過小適合を回避するのに役立ちます。テスト セットは、モデルのテストに使用されるデータのサブセットであり、最終的な評価指標を提供します。テスト セットは、新しいデータに対するモデルのパフォーマンスを評価し、モデルの汎化能力を測定するのに役立ちます。Vertex AI のデフォルト データ分割を使用すると、ランダム サンプリング方法を使用してデータをトレーニング、検証、テスト セットに分割し、次の割合のデータを各セットに割り当てることができます<sup>1</sup>。

他のオプションは、次の理由により、オプション B ほど優れていません。

\* オプション A: Vertex AI 手動分割を使用し、ストア名機能を使用して各セットに 1 つのストアを割り当てると、データを代表的なバランスのとれたセットに分割できず、エラーやパフォーマンスの低下が発生する可能性があります。手動分割は、ml\_use ラベルまたはデータ フィルター式を使用して、データをセットに分割する方法を制御できるデータ分割方法です。手動分割は、データ分割ロジックをカスタマイズし、複雑なデータ形式または非標準のデータ形式を処理するのに役立ちます。店名特徴量は、売上データを収集した店舗の名前を示す特徴量です。店舗名機能は、データのソースを特定し、店舗ごとにデータをグループ化するのに役立ちます。ただし、Vertex AI の手動分割を使用し、ストア名機能を使用して各セットに 1 つのストアを割り当てると、データ

を代表的なバランスの取れたセットに分割できず、エラーやパフォーマンスの低下が発生する可能性があります。書く必要があります

\* コードを作成し、ml\_use ラベルまたはデータ フィルター式を作成して構成し、セットごとに1つのストアを割り当てます。さらに、このオプションでは、各セットのデータがデータセット全体のデータと同じ分布と特性を持つことが保証されないため、データの一般的なパターンを学習できなくなり、モデルに偏りや分散が生じる可能性があります2。

\* オプション C: Vertex AI 時系列分割を使用し、販売タイムスタンプ機能を時間変数として指定すると、データを代表的なバランスのとれたセットに分割できず、エラーやパフォーマンスの低下が発生する可能性があります。時系列分割は、データの順序に基づいてデータをセットに分割できるデータ分割方法です。時系列に分割すると、データの時間的な依存関係と順序を保持し、データ漏洩を防ぐことができます。販売タイムスタンプ機能は、販売データが収集された日時を示す機能です。販売タイムスタンプ機能は、時間の経過に伴うデータの変化と傾向を追跡し、データの季節性と周期性を把握するのに役立ちます。ただし、Vertex AI 時系列分割を使用し、販売タイムスタンプ機能を時間変数として指定すると、データを代表的なバランスのとれたセットに分割できず、エラーやパフォーマンスの低下が発生する可能性があります。コードを記述し、時間変数を作成して構成し、時間変数の順序でデータを分割する必要があります。さらに、このオプションでは、各セットのデータがデータセット全体のデータと同じ分布と特性を持つことが保証されないため、データの一般的なパターンを学習できなくなり、モデルに偏りや分散が生じる可能性があります3。

\* Option D: Using Vertex AI random split, assigning 70% of the rows to the training set, 10% to the validation set, and 20% to the test set would not allow you to use the default data split method that is provided by Vertex AI, and could increase the complexity and cost of the data split process. A random split is a data split method that allows you to split your data into sets by using a random sampling method, and assign a custom percentage of the data to each set. A random split can help you split your data into representative and balanced sets, and avoid data leakage. However, using Vertex AI random split, assigning 70% of the rows to the training set, 10% to the validation set, and 20% to the test set would not allow you to use the default data split method that is provided by Vertex AI, and could increase the complexity and cost of the data split process. You would need to write code, create and configure the random split method, and assign the custom percentages to each set. Moreover, this option would not use the default data split method that is provided by Vertex AI, which can simplify the data split process, and works well in most cases1.

References:

\* About data splits for AutoML models | Vertex AI | Google Cloud

\* Manual split for unstructured data

\* Mathematical split

## 最新問題: 80

Google Cloud でディープ ニューラル ネットワーク モデルをトレーニングしました。モデルはトレーニング データでは損失が少ないですが、検証データではパフォーマンスが悪くなります。モ

デルには過学習に対する耐性を持たせる必要があります。モデルを再トレーニングするときどの戦略を使用する必要がありますか？

- A. AI プラットフォームでハイパーパラメータ調整ジョブを実行して、L2 正則化パラメータとドロップアウトパラメータを最適化します。
- B. AI プラットフォームでハイパーパラメータ調整ジョブを実行して学習率を最適化し、ニューロンの数を 2 倍に増やします。
- C. 12 の正則化パラメータ 0.4 を適用し、学習率を 10 分の 1 に下げます。
- D. ドロップアウトパラメータ 0.2 を適用し、学習率を 10 分の 1 に減少させます。

**Answer:** ([解答を表示する](#))

#### 最新問題: 81

モバイルネットワーク事業者は、Amazon Athena と Amazon S3 を使用して企業の業務を分析および最適化するための分析プラットフォームを構築しています。

ソースシステムは、データを .CSV 形式でリアルタイムに送信します。データエンジニアリングチームは、データを Amazon S3 に保存する前に Apache Parquet 形式に変換したいと考えています。

実装に最も労力がかからないソリューションはどれですか？

- A. Amazon EC2 インスタンスで Apache Kafka Streams を使用して .CSV データを取り込み、Kafka Connect S3 を使用してデータを Parquet としてシリアル化します。
- B. Amazon Kinesis Data Streams から .CSV データを取り込み、Amazon Glue を使用してデータを Parquet に変換します。
- C. Amazon EMR クラスターで Apache Spark Structured Streaming を使用して .CSV データを取り込み、Apache Spark を使用してデータを Parquet に変換します。
- D. Amazon Kinesis Data Streams から .CSV データを取り込み、Amazon Kinesis Data Firehose を使用してデータを Parquet に変換します。

**Answer:** B ([メッセージを残す](#))

説明/参照:

#### 最新問題: 82

セキュリティに敏感な企業の機械学習スペシャリストは、モデルトレーニング用のデータセットを準備しています。データセットは Amazon S3 に保存され、個人を特定できる情報 (PII) が含まれています。

データセット:

- \* VPC からのみアクセスできる必要があります。
  - \* 公共のインターネットを通過してはなりません。
- これらの要件はどのように満たされるのでしょうか？

- A. VPC エンドポイントを作成し、指定された VPC エンドポイントと VPC へのアクセスを制限するバケットアクセスポリシーを適用します。

- B. VPC エンドポイントを作成し、指定された VPC エンドポイントと Amazon EC2 インスタンスからのアクセスを許可するバケット アクセス ポリシーを適用します。
- C. VPC エンドポイントを作成し、ネットワーク アクセス コントロール リスト (NACL) を使用して、指定された VPC エンドポイントと Amazon EC2 インスタンスの間のトラフィックのみを許可します。
- D. VPC エンドポイントを作成し、セキュリティ グループを使用して、指定された VPC エンドポイントと Amazon EC2 インスタンスへのアクセスを制限します。

**Answer: B** ([メッセージを残す](#))

説明/参照: <https://docs.aws.amazon.com/AmazonS3/latest/dev/example-bucket-policies-vpc-endpoint.html>

#### 最新問題: 83

あなたは、TensorFlow フレームワークを使用して最先端の深層学習モデルを構築するチームで働いています。チームは毎週複数の ML 実験を実行しているため、実験の実行を追跡することが困難になっています。オーバーヘッド コードを最小限に抑えながら、Google Cloud 上で実行される ML 実験を効果的に追跡、可視化、デバッグするためのシンプルなアプローチが必要です。どのように進めるべきでしょうか？

- A. 指標ファイルを BigQuery テーブルに書き込んで保存するように Cloud Function を設定します。可視化のために TensorBoard をローカルでホストするように Google Cloud VM を構成します。
- B. Vertex AI Workbench ノートブック インスタンスをセットアップします。インスタンスを使用して指標データを Cloud Storage バケットに保存し、可視化のために TensorBoard をローカルにホストします。
- C. メトリクス ファイルを Cloud Storage バケットに書き込んで保存するための Cloud Function を設定します。可視化のために TensorBoard をローカルでホストするように Google Cloud VM を構成します。
- D. メトリクスとパラメーターを追跡するために Vertex AI Experiments をセットアップします。視覚化のために Vertex AI TensorBoard を構成します。

**Answer: D** ([メッセージを残す](#))

#### 最新問題: 84

You need to design a customized deep neural network in Keras that will predict customer purchases based on their purchase history. You want to explore model performance using multiple model architectures, store training data, and be able to compare the evaluation metrics in the same dashboard. What should you do?

- A. Create multiple models using AutoML Tables
- B. Create an experiment in Kubeflow Pipelines to organize multiple runs
- C. Automate multiple training runs using Cloud Composer
- D. Run multiple training jobs on AI Platform with similar job names

**Answer: D** ([メッセージを残す](#))

## 最新問題: 85

あなたは銀行で働いています。ローン申請の決定をサポートする信用リスク モデルを開発する必要がある TensorFlow のニューラル ネットワークを使用してモデルを実装することを決定する規制要件により、その機能に基づいてモデルの予測を説明できる必要がある モデルがデプロイされるときまた、モデルのパフォーマンスを時間の経過とともに監視したいと考えています。モデルの開発とデプロイメントの両方に Vertex AI を使用することにしました。どうすればよいですか？

- A. サンプリングされた Shapley 法で Vertex Explainable AI を使用し、Vertex AI モデル モニタリングを有効にしてフィーチャ分布のドリフトをチェックします。
- B. サンプリングされた Shapley 法で Vertex Explainable AI を使用し、Vertex AI モデル モニタリングを有効にしてフィーチャ分布のスキューをチェックします。
- C. XRAI メソッドで Vertex Explainable AI を使用し、Vertex AI モデル モニタリングを有効にしてフィーチャ分布のドリフトをチェックします。
- D. XRAI メソッドで Vertex Explainable AI を使用し、Vertex AI モデル モニタリングを有効にしてフィーチャ分布のスキューをチェックします。

**Answer:** ([解答を表示する](#))

規制要件を満たし、長期にわたって監視できる信用リスク モデルを開発するには、次の手順に従う必要があります。

\* サンプリングされた Shapley メソッドで Vertex Explainable AI を使用します。Vertex Explainable AI は、機械学習モデルの特徴属性を提供するサービスであり、各特徴が予測にどのように寄与するかを理解するのに役立ちます<sup>1</sup>。サンプリングされた Shapley 法は、すべての可能な特徴サブセットにわたる予測に対する各特徴の限界寄与に基づいて、各特徴の Shapley 値を推定する技術です<sup>2</sup>。サンプリングされた Shapley 法は、特徴の非線形効果や相互作用効果を捉えることができるため、ニューラル ネットワークやその他の複雑なモデルに適しています<sup>3</sup>。

\* Vertex AI モデル モニタリングを有効にして、機能分布のドリフトをチェックします。Vertex AI Model Monitoring は、デプロイされたモデルのパフォーマンスと品質を長期にわたって追跡および管理するのに役立つサービスです<sup>4</sup>。フィーチャ分布ドリフトは、入力フィーチャの分布がトレーニング データから大幅に変化するときに発生するデータ ドリフトの一種で、モデルの精度と信頼性に影響を与える可能性があります。特徴量分布のドリフトをチェックすることで、モデルの再トレーニングや新しいデータによる更新が必要な時期を検出できます。

参考文献:

※1 :Vertex Explainable AI入門 | 頂点 AI | グーグルクラウド

※2 :シャプリー値 - Wikipedia

※3 :Explainable AI :ディープラーニングを解釈・説明・可視化するAI

\*4: Vertex AI モデルモニタリングの概要 | 頂点 AI | グーグルクラウド

\* [5]: データ ドリフトについてモデルを監視する | 頂点 AI | グーグルクラウド

## 最新問題: 86

scikit-learn 分類モデルを実稼働環境にデプロイする必要があります。モデルはリクエストを処理できる必要があります

24 時間年中無休で、午前 8 時から午後 7 時までの間、運用アプリケーションに対して 1 秒あたり数百万件のリクエストが発生すると予想されます。導入コストを最小限に抑える必要があります。どうすればよいでしょうか？

A. オンライン Vertex AI 予測エンドポイントをデプロイします。最大レプリカ数を 1 に設定します。

B. オンライン Vertex AI 予測エンドポイントをデプロイします。最大レプリカ数を 100 に設定します。

C. レプリカごとに 1 つの GPU を備えたオンライン Vertex AI 予測エンドポイントをデプロイします。最大レプリカ数を次のように設定します。

1.

D. レプリカごとに 1 つの GPU を備えたオンライン Vertex AI 予測エンドポイントをデプロイします。最大レプリカ数を次のように設定します。

100。

**Answer: B (メッセージを残す)**

scikit-learn 分類モデルを本番環境にデプロイするための最良のオプションは、オンラインの Vertex AI 予測エンドポイントをデプロイし、最大レプリカ数を 100 に設定することです。このオプションにより、Google Cloud のパワーとスケーラビリティを活用して 24 時間年中無休でリクエストを処理できます。毎秒数百万のリクエストを処理します。Vertex AI は、Google Cloud 上で機械学習ソリューションを構築およびデプロイするための統合プラットフォームです。Vertex AI は、トレーニングされた scikit-learn モデルをオンライン予測エンドポイントにデプロイでき、個々のインスタンスに対して低レイテンシの予測を提供できます。オンライン予測エンドポイントは、仮想マシン上で実行されるモデルのコピーである 1 つ以上のレプリカで構成されます。最大レプリカ数は、エンドポイントに対して作成できるレプリカの最大数を決定するパラメータです。最大レプリカ数を 100 に設定すると、トラフィックが増加するとエンドポイントが 100 レプリカにスケールアップし、トラフィックが減少するとゼロレプリカにスケールダウンできるようになります。これにより、使用したリソースに対してのみ料金が発生するため、導入コストを最小限に抑えることができます。さらに、自動スケーリングアルゴリズム オプションを使用して、レイテンシと使用率のメトリクスに基づいてエンドポイントのスケールアップ動作を最適化できます

1。  
他のオプションは、次の理由により、オプション B ほど優れていません。

\* オプション A: オンラインの Vertex AI 予測エンドポイントをデプロイし、最大レプリカ数を 1 に設定しても、年中無休でリクエストに対応し、1 秒あたり数百万件のリクエストを処理することはできません。最大レプリカ数を 1 に設定すると、エンドポイントが 1 つのレプリカのみに制限されるため、トラフィックが増加するとパフォーマンスの問題やサービスの中断が発生する可能性があります。さらに、最大レプリカ数を 1 に設定すると、トラフィックが減少したときにエンドポイントがゼロレプリカにスケールダウンすることがなくなり、使用しないリソースに対して料金を支払うことになり、導入コストが増加する可能性があります1。

\* オプション C: レプリカごとに 1 つの GPU を備えたオンライン Vertex AI 予測エンドポイントをデプロイし、最大レプリカ数を 1 に設定すると、24 時間年中無休でリクエストを処理し、1 秒あたり数百万のリクエストを処理することができず、デプロイのコストが増加します。各レプリカに GPU を追加すると、エンドポイントの計算能力が向上しますが、GPU は CPU よりも高価であるため、導入コストも増加します。さらに、最大レプリカ数を 1 に設定すると、エンドポイントが 1 つのレプリカのみで制限されるため、トラフィックが増加するとパフォーマンスの問題やサービスの中断が発生し、トラフィックが減少するとエンドポイントがゼロ レプリカにスケールダウンできなくなる可能性があります<sup>1</sup>。さらに、scikit-learn は GPU アクセラレーション用に最適化されていないため、scikit-learn モデルは GPU の恩恵を受けません<sup>2</sup>。

\* オプション D: レプリカごとに 1 つの GPU を備えたオンライン Vertex AI 予測エンドポイントをデプロイし、最大レプリカ数を 100 に設定すると、年中無休でリクエストを処理し、レプリカごとに数百万のリクエストを処理できます。

\* 2 番目ですが、導入コストが増加します。各レプリカに GPU を追加すると、エンドポイントの計算能力が向上しますが、GPU は CPU よりも高価であるため、導入コストも増加します。最大レプリカ数を 100 に設定すると、エンドポイントはトラフィックが増加すると 100 個のレプリカにスケールアップし、トラフィックが減少するとゼロ レプリカにスケールダウンできるため、展開コストを最小限に抑えることができます。ただし、scikit-learn は GPU アクセラレーション用に最適化されていないため、scikit-learn モデルは GPU の恩恵を受けません<sup>2</sup>。したがって、scikit-learn モデルに GPU を使用することは不必要で無駄になります。

参考文献:

\* Google Cloud 認定資格の準備: 機械学習エンジニア、コース 3: 本番環境 ML システム、第 2 週: ML 予測の提供

\* Google Cloud Professional Machine Learning Engineer 試験ガイド、セクション 3: 本番環境での ML モデルのスケーリング、3.1 本番環境への ML モデルのデプロイ

\* 公式 Google Cloud Certified Professional Machine Learning Engineer 学習ガイド、第 6 章: 実稼働 ML システム、セクション 6.2: ML 予測の提供

\* オンライン予測

\* オンライン予測のスケーリング

\* scikit-learn FAQ

**最新問題: 87**

You have developed an application that uses a chain of multiple scikit-learn models to predict the optimal price for your company's products. The workflow logic is shown in the diagram Members of your team use the individual models in other solution workflows. You want to deploy this workflow while ensuring version control for each individual model and the overall workflow Your application needs to be able to scale down to zero. You want to minimize the compute resource utilization and the manual effort required to manage this solution. What should you do?

**A.** Expose each individual model as an endpoint in Vertex AI Endpoints. Create a custom container endpoint to orchestrate the workflow.

**B.** Create a custom container endpoint for the workflow that loads each model's individual files. Track the versions of each individual model in BigQuery.

**C.** Expose each individual model as an endpoint in Vertex AI Endpoints. Use Cloud Run to orchestrate the workflow.

**D.** Load each model's individual files into Cloud Run. Use Cloud Run to orchestrate the workflow. Track the versions of each individual model in BigQuery.

**Answer: C** ([メッセージを残す](#))

The option C is the most efficient and scalable solution for deploying a machine learning workflow with multiple models while ensuring version control and minimizing compute resource utilization. By exposing each model as an endpoint in Vertex AI Endpoints, it allows for easy versioning and management of individual models. Using Cloud Run to orchestrate the workflow ensures that the application can scale down to zero, thus minimizing resource utilization when not in use. Cloud Run is a service that allows you to run stateless containers on a fully managed environment or on Google Kubernetes Engine. You can use Cloud Run to invoke the endpoints of each model in the workflow and pass the data between them. You can also use Cloud Run to handle the input and output of the workflow and provide an HTTP interface for the application.

References:

\* Vertex AI Endpoints documentation

\* Cloud Run documentation

\* Preparing for Google Cloud Certification: Machine Learning Engineer Professional Certificate

### 最新問題: 88

Your data science team has requested a system that supports scheduled model retraining, Docker containers, and a service that supports autoscaling and monitoring for online prediction requests. Which platform components should you choose for this system?

**A.** Vertex AI Pipelines and App Engine

**B.** Vertex AI Pipelines, Vertex AI Prediction, and Vertex AI Model Monitoring

**C.** Cloud Composer, BigQuery ML, and Vertex AI Prediction

**D.** Cloud Composer, Vertex AI Training with custom containers, and App Engine

**Answer: B** ([メッセージを残す](#))

\* Option A is incorrect because Vertex AI Pipelines and App Engine do not meet all the requirements of the system. Vertex AI Pipelines is a service that allows you to create, run, and manage ML workflows using TensorFlow Extended (TFX) components or custom components<sup>1</sup>. App Engine is a service that allows you to build and deploy scalable web applications using standard or flexible environments<sup>2</sup>. However, App Engine does not support Docker containers in the standard environment, and does not provide a dedicated service for online prediction and monitoring of ML models<sup>3</sup>.

\* Option B is correct because Vertex AI Pipelines, Vertex AI Prediction, and Vertex AI Model Monitoring meet all the requirements of the system. Vertex AI Prediction is a service that allows you to deploy and serve ML models for online or batch prediction, with support for autoscaling

and custom containers<sup>4</sup>. Vertex AI Model Monitoring is a service that allows you to monitor the performance and fairness of your deployed models, and get alerts for any issues or anomalies<sup>5</sup>.

\* Option C is incorrect because Cloud Composer, BigQuery ML, and Vertex AI Prediction do not meet all the requirements of the system. Cloud Composer is a service that allows you to create, schedule, and

\* manage workflows using Apache Airflow. BigQuery ML is a service that allows you to create and use ML models within BigQuery using SQL queries. However, BigQuery ML does not support custom containers, and Vertex AI Prediction does not support scheduled model retraining or model monitoring.

\* Option D is incorrect because Cloud Composer, Vertex AI Training with custom containers, and App Engine do not meet all the requirements of the system. Vertex AI Training is a service that allows you to train ML models using built-in algorithms or custom containers. However, Vertex AI Training does not support online prediction or model monitoring, and App Engine does not support Docker containers in the standard environment or online prediction and monitoring of ML models<sup>3</sup>.

References:

- \* Vertex AI Pipelines overview
- \* App Engine overview
- \* Choosing an App Engine environment
- \* Vertex AI Prediction overview
- \* Vertex AI Model Monitoring overview
- \* [Cloud Composer overview]
- \* [BigQuery ML overview]
- \* [BigQuery ML limitations]
- \* [Vertex AI Training overview]

### 最新問題: 89

Keras と TensorFlow を使用して不正検出モデルを開発しています。顧客のトランザクションの記録は BigQuery の大きなテーブルに保存されます。これらのレコードをモデルのトレーニングに使用する前に、コスト効率が高く効率的な方法で前処理する必要があります。トレーニングされたモデルは、BigQuery でバッチ推論を実行するために使用されます。前処理ワークフローはどのように実装すればよいのでしょうか？

- A.** Apache Spark を使用して前処理パイプラインを実装し、Dataproc でパイプラインを実行します。前処理されたデータを CSV ファイルとして Cloud Storage バケットに保存します。
- B.** SQL を使用して BigQuery で前処理を実行します。TensorFlow の BigQueryClient を使用して、BigQuery からデータを直接読み取ります。
- C.** Apache Beam を使用して前処理パイプラインを実装し、Dataflow でパイプラインを実行します。前処理されたデータを CSV ファイルとして Cloud Storage バケットに保存します。
- D.** データを pandas DataFrame にロードします。panda の変換を使用して前処理ステップを実装します。DataFrame 上でモデルを直接トレーニングします。

**Answer: C (メッセージを残す)**

**最新問題: 90**

あなたは、オンライン フォーラムを管理するペットフード会社で働いています。顧客は自分のペットの写真をフォーラムにアップロードして、他の人と共有します。毎日約 20 枚の写真がアップロードされます。アップロードされた各写真に動物が含まれているかどうかを、ほぼリアルタイムで自動的に検出したいと考えています。時間を優先し、アプリケーションの開発と展開のコストを最小限に抑えるにはどうすればよいでしょうか？

- A.** ユーザーが送信した画像を Cloud Vision API に送信します。オブジェクト ローカリゼーションを使用して画像内のすべてのオブジェクトを識別し、結果を動物のリストと比較します。
- B.** TensorFlow Hub からオブジェクト検出モデルをダウンロードします。モデルを Vertex AI エンドポイントにデプロイします。  
ユーザーが送信した新しい画像をモデル エンドポイントに送信して、各写真に動物が写っているかどうかを分類します。
- C.** 動物の周囲の境界ボックスを使用して、以前に送信された画像に手動でラベルを付けます。Vertex AI を使用して AutoML オブジェクト検出モデルを構築します。モデルを Vertex AI エンドポイントにデプロイします。新しいユーザーが送信した画像をモデル エンドポイントに送信して、各写真が正しいかどうかを検出します。動物を飼っています。
- D.** 以前に送信された画像に動物がいるかどうかを手動でラベル付けします。Vertex AI で画像データセットを作成します。Vertex AutoML を使用して分類モデルをトレーニングし、2 つのクラスを区別します。モデルを Vertex AI エンドポイントにデプロイします。新しいユーザーが送信した画像を送信します。モデルのエンドポイントに送信して、各写真に動物が写っているかどうかを分類します。

**Answer: A (メッセージを残す)**

Cloud Vision API は、事前トレーニングされた機械学習モデルを使用して画像を分析できるサービスです<sup>1</sup>。Cloud Vision API を使用すると、顔検出、テキスト抽出、ロゴ認識、オブジェクトの位置特定などのさまざまなタスクを実行できます<sup>1</sup>。オブジェクトの位置特定は、画像内の複数のオブジェクトを検出し、それらの周囲に境界ボックスを描画できる機能です<sup>2</sup>。検出された各オブジェクトのラベルと信頼スコアを取得することもできます<sup>2</sup>。

ユーザーが送信した画像を Cloud Vision API に送信すると、オブジェクトの位置特定を使用して画像内のすべてのオブジェクトを識別し、結果を動物のリストと比較できます。AnnotateImageRequest で OBJECT\_LOCALIZATION 機能タイプを使用して、オブジェクトのローカリゼーションをリクエストできます<sup>3</sup>。

その後、AnnotateImageResponse の localizedObjectAnnotations フィールドを使用して、検出されたオブジェクト、そのラベル、および信頼スコアのリストを取得できます。ラベルを犬、猫、鳥などの動物の事前定義されたリストと比較し、画像に動物が含まれているかどうかを判断できます。

このオプションは、手動のラベル付け、モデルのトレーニング、またはモデルの展開を必要とせず、アップロードされた各写真に動物が写っているかどうかを自動的にほぼリアルタイムで検

出できるため、シナリオに最適です。また、機械学習の専門知識やインフラストラクチャを必要とせずに、すぐに使えるサービスとして Cloud Vision API を使用できるため、アプリケーションの開発とデプロイメントにかかる時間を優先し、コストを最小限に抑えることができます。

他のオプションは、手動のラベル付け、モデルのトレーニング、またはモデルのデプロイメントを必要とするため、アプリケーションの開発とデプロイメントの時間とコストが増加するか、より複雑でオブジェクト検出モデルを使用するため、シナリオには適していません。オブジェクト位置特定モデルよりも計算コストが高く、画像に動物がいるかどうかを検出するという単純なタスクには必要ありません。

参考文献:

\* クラウドビジョン API | グーグルクラウド

\* オブジェクトのローカリゼーション | クラウドビジョンAPI | グーグルクラウド

\* AnnotateImageRequest | クラウドビジョンAPI | グーグルクラウド

\* [AnnotateImageResponse | クラウドビジョンAPI | グーグルクラウド]

### 最新問題: 91

あなたはモバイル ゲーム会社の ML エンジニアです。あなたのチームのデータサイエンティストが最近 TensorFlow モデルをトレーニングしました。あなたはこのモデルをモバイル アプリケーションにデプロイする責任があります。現在のモデルの推論レイテンシが運用要件を満たしていないことがわかりました。推論時間を 50% 削減する必要があり、レイテンシー要件を達成するためにモデル精度の多少の低下は許容します。新しいモデルをトレーニングしない場合、レイテンシを短縮するためにどのモデル最適化手法を最初に試すべきでしょうか？

A. 重みの枝刈り

B. ダイナミックレンジ量子化

C. モデル蒸留

D. 次元削減

**Answer: B (メッセージを残す)**

ダイナミックレンジ量子化は、モデルの重みとアクティベーションの数値精度を低下させるレイテンシーを削減するためのモデル最適化手法です。この手法により、モデルのサイズ、メモリ使用量、推論時間を最大 4 分の 1 に削減でき、精度の損失も無視できます。ダイナミックレンジ量子化は、再トレーニングせずにトレーニング済みの TensorFlow モデルに適用でき、低遅延と低消費電力を必要とするモバイル アプリケーションに適しています。

重み枝刈り、モデルの蒸留、次元削減もレイテンシを短縮するためのモデル最適化手法ですが、ダイナミックレンジ量子化と比較していくつかの制限や欠点があります。

\* 重み枝刈りは、予測にわずかな影響しか及ぼさないモデル内のパラメーターを削除することによって機能します。プルーニングされたモデルはディスク上のサイズが同じで、実行時の待ち時間も同じですが、より効果的に圧縮できます。このため、プルーニングはモデルのダウンロードサイズを削減するための有用な手法になりますが、推論時間を短縮するための手法ではありません。

\* モデルの蒸留は、より小さく単純なモデル (生徒) をトレーニングして、より大きく複雑なモデル (教師) の動作を模倣することによって機能します。蒸留されたモデルは、元のモデルよりも遅延とメモリ使用量が少なくなりますが、再トレーニングが必要であり、教師モデルの精度が維持されない可能性があります。

\* 次元削減は、入力データまたはモデル レイヤー内のフィーチャまたは次元の数を減らすことによって機能します。次元削減によりモデルの計算効率と汎化能力が向上しますが、一部の情報が失われたり、データやモデルにノイズが発生したりする可能性もあります。次元削減には、モデル アーキテクチャの再トレーニングまたは変更も必要です。

参考文献:

\* [TensorFlow モデルの最適化]

\* [TensorFlow モデル最適化ツールキット - トレーニング後の整数量子化]

\* [レイテンシーを削減し、新しいデータに適応するためのモデル最適化方法]

有効な **Professional-Machine-Learning-Engineer** 問題集は GoShiken.com が提供された合格しやすい Professional-Machine-Learning-Engineer 試験問題集！ GoShiken.com が最新の **Professional-Machine-Learning-Engineer** 試験問題集を提供しています。GoShiken.com Professional-Machine-Learning-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Professional-Machine-Learning-Engineer 問題集をゲットする人はこちら：  
<https://www.goshiken.com/Google/Professional-Machine-Learning-Engineer-mondaishu.html>  
(**29030%OFF**問題集溶と正解付きで **30%w**特別割引コード: **Freepdfdumps**)

最新問題: 92

チームは、画像に運転免許証、パスポート、クレジットカードが含まれているかどうかを予測するモデルを構築する必要があります。データ エンジニアリング チームはすでにパイプラインを構築し、運転免許証の画像 10,000 枚、パスポートの画像 1,000 枚、クレジットカードの画像 1,000 枚で構成されるデータセットを生成しました。次に、ラベル マップ ['driverslicense', 'passport', 'credit\_card'] を使用してモデルをトレーニングする必要があります。どの損失関数を使用する必要がありますか？

- A. カテゴリカル ヒンジ
- B. バイナリクロスエントロピー
- C. カテゴリカルクロスエントロピー
- D. スパースのカテゴリカルクロスエントロピー

**Answer: C (メッセージを残す)**

カテゴリカル クロス エントロピーは、ターゲット変数が 3 つ以上の可能な値を持つ、マルチクラス分類問題に適した損失関数です。カテゴリカル クロス エントロピーは、ターゲットクラスの真の確率分布とモデルの予測確率分布の間の差異を測定します。それは次のように定義されます。

$$L = - \sum(y_i * \log(p_i))$$

ここで、 $y_i$  はクラス  $i$  の真の確率、 $p_i$  はクラス  $i$  の予測確率です。カテゴリカルクロスエントロピーは、不正確な予測を行うモデルにペナルティを与え、モデルが正しいクラスに高い確率を割り当て、不正確なクラスに低い確率を割り当てるように促します。

画像に運転免許証、パスポート、またはクレジットカードが含まれているかどうかを予測するモデルを構築するユースケースでは、カテゴリカルクロスエントロピーが使用する適切な損失関数です。これは、問題がマルチクラス分類問題であり、ターゲット変数に3つの可能な値があるためです: ['drivers\_license', 'パスポート', 'クレジットカード']。ラベルマップは、クラス名をクラスインデックスにマップするリストです。次のようになります。

「drivers\_license」はインデックス0に対応し、「passport」はインデックス1に対応し、「credit\_card」はインデックス2に対応します。モデルは各画像の3つのクラスにわたる確率分布を出力する必要があり、カテゴリカルクロスエントロピー損失関数は比較する必要があります。真のラベルを含む出力。したがって、カテゴリカルクロスエントロピーは、この使用例に最適な損失関数です。

#### 最新問題: 93

機械学習スペシャリストは、幅広い経済要因に基づいて将来の雇用率を予測するモデルを構築しています。データを調査しているときに、スペシャリストは入力特徴の大きさが大きく異なることに気づきました。スペシャリストは、より大きな大きさの変数がモデルを支配することを望んでいません。

モデルトレーニング用のデータを準備するためにスペシャリストは何をすべきでしょうか？

- A. 分位値ビンングを適用してデータをカテゴリビンにグループ化し、大きさを分布に置き換えることでデータ内の関係を維持します。
- B. デカルト積変換を適用して、大きさに依存しないフィールドの新しい組み合わせを作成します。
- C. 正規化を適用して、各フィールドの平均が0、分散が1になるようにして、有意な大きさを削除します。
- D. 直交スパスバイグラム (OSB) 変換を適用して固定サイズのスライディングウィンドウを適用し、同様の大きさの新しい特徴を生成します。

**Answer: C** ([メッセージを残す](#))

説明/参照: <https://docs.aws.amazon.com/machine-learning/latest/dg/data-transformations-reference.html>

#### 最新問題: 94

あなたは、全員が AI Platform を使用する 50 名を超えるデータサイエンティストからなる成長しつつあるチームで働いています。あなたは、ジョブ、モデル、バージョンをクリーンかつスケラブルな方法で整理するための戦略を設計しています。どの戦略を選択する必要がありますか？

- A. AI プラットフォームノートブックに制限的な IAM 権限を設定して、単一のユーザーまたはグループのみが特定のインスタンスにアクセスできるようにします。

- B. 各データサイエンティストの作業を別のプロジェクトに分離して、各データサイエンティストが作成したジョブ、モデル、バージョンにそのユーザーのみがアクセスできるようにします。
- C. ラベルを使用して、リソースを説明的なカテゴリに整理します。作成された各リソースにラベルを適用すると、ユーザーはリソースを表示または監視するときにラベルで結果をフィルタリングできます。
- D. AI Platform リソースの使用状況に関する情報を取得するために適切にフィルタリングされた Cloud Logging ログ用の BigQuery シンクを設定します。BigQuery で、ユーザーを使用しているリソースにマッピングする SQL ビューを作成します。

**Answer:** ([解答を表示する](#))

ラベルは、ジョブ、モデル、バージョン、エンドポイントなどの AI Platform リソースに添付できるキーと値のペアです<sup>1</sup>。ラベルは、リソースをプロジェクト、チーム、環境、目的などのわかりやすいカテゴリに整理するのに役立ちます。ラベルを使用すると、リソースを一覧表示または監視するときに結果をフィルタリングしたり、請求やクォータの目的でリソースをグループ化したりすることができます<sup>2</sup>。ラベルを使用すると、不必要な複雑さやオーバーヘッドを生じさせることなく AI Platform リソースを管理できるシンプルでスケラブルな方法です。したがって、ラベルを使用してリソースを整理することが、この使用例にとって最良の戦略です。

参考文献:

- \* ラベルの使用
- \* ラベルによるフィルタリングとグループ化

#### 最新問題: 95

あなたは、複数のデータサイエンスワークロードを扱うスタートアップ企業で働いています。現在、コンピューティングインフラストラクチャはオンプレミスにあります。データサイエンスワークロードは PySpark にネイティブです。チームはデータサイエンスワークロードを Google Cloud に移行することを計画しています。1つのデータサイエンスジョブを Google Cloud に移行するための概念実証を構築する必要があります。最小限のコストと必要な移行プロセスを提案したい努力。まず何をすべきでしょうか？

- A. 基本的なノードプール構成を使用して Google Kubernetes Engine クラスタを作成し、Java Scala とそれに依存する Apache Spark をインストールします。
- B. インスタンスタイプ n2-standard-4 で Vertex AI Workbench ノートブックを作成します。
- C. n2-standard-4 VM インスタンスを作成し、それに Java、Scala、および Apache Spark の依存関係をインストールします。
- D. 標準 (マスター 1 つ、ワーカー 3 つ) Dataproc クラスタを作成し、その上で Vertex AI Workbench ノートブック インスタンスを実行します。

**Answer:** D ([メッセージを残す](#))

#### 最新問題: 96

最近、新しい Google Cloud プロジェクトを作成しました。Cloud Shell から Vertex AI Pipeline ジョブを送信できることをテストした後、Vertex AI Workbench のユーザー管理のノートブックインスタンスを使用して、そのインスタンスからコードを実行したいと考えています。インスタ

ンスを作成して実行しました。コードを記述しましたが、今回は権限が不十分なエラーでジョブが失敗します。あなたは何をするべきか？

- A. 作成した Workbench インスタンスが、使用する Vertex AI Pipelines リソースの同じリージョンにあることを確認してください。
- B. Vertex AI Workbench インスタンスが、使用する Vertex AI Pipeline リソースの同じサブネットワーク上にあることを確認します。
- C. Vertex AI Workbench インスタンスに Identity and Access Management (IAM) Vertex AI User role が割り当てられていることを確認します。
- D. Vertex AI Workbench インスタンスに Identity and Access Management (IAM) Notebooks Runner ロールが割り当てられていることを確認します。

**Answer: C (メッセージを残す)**

Vertex AI Workbench は、Google Cloud 上で Jupyter ノートブックを作成して実行できる統合開発環境 (IDE) です。Vertex AI Pipelines は、Vertex AI コンポーネントを使用して機械学習ワークフローを作成および管理できるサービスです。Vertex AI Workbench インスタンスから Vertex AI Pipeline ジョブを送信するには、Vertex AI リソースにアクセスするための適切な権限が必要です。Identity and Access Management (IAM) Vertex AI ユーザー ロールは、モデル、エンドポイント、パイプラインの作成やデプロイなど、Vertex AI サービスを使用するために必要な最小限の権限を付与する事前定義されたロールです。Vertex AI ユーザー ロールを Vertex AI Workbench インスタンスに割り当てることで、インスタンスに Vertex AI Pipeline ジョブを送信するための十分な権限があることを確認できます。Cloud Console、gcloud コマンドライン ツール、または Cloud IAM API を使用して、インスタンスにロールを割り当てることができます。参考資料: 答えは、Google Cloud の公式ドキュメントと、Vertex AI Workbench、Vertex AI Pipelines、IAM に関連するリソースから確認できます。

\* Vertex AI ワークベンチ | グーグルクラウド

\* Vertex AI パイプライン | グーグルクラウド

\* Vertex AI の役割 | グーグルクラウド

\* リソースへのアクセス権の付与、変更、取り消し | グーグルクラウド

#### 最新問題: 97

データサイエンス チームは、さまざまな機能、モデル アーキテクチャ、ハイパーパラメーターを迅速に実験する必要があります。さまざまな実験の精度メトリクスを追跡し、API を使用して時間の経過とともにメトリクスをクエリする必要があります。手作業を最小限に抑えながら実験を追跡し、報告するには何を使用すればよいのでしょうか？

- A. Kubeflow Pipelines を使用して実験を実行します。メトリクス ファイルをエクスポートし、Kubeflow Pipelines API を使用して結果をクエリします。
- B. AI Platform Training を使用して実験を実行します。精度指標を BigQuery に書き込み、BigQueryAPI を使用して結果をクエリします。
- C. AI Platform Training を使用して実験を実行します。精度指標を Cloud Monitoring に書き込み、Monitoring API を使用して結果をクエリします。

D. AI Platform Notebooks を使用して実験を実行します。結果を共有 Google スプレッドシート ファイルに収集し、Google Sheets API を使用して結果をクエリします

**Answer: C** ([メッセージを残す](#))

AI Platform Training は、さまざまな機能、モデル アーキテクチャ、ハイパーパラメータを使用して、Google Cloud 上で機械学習の実験を実行できるサービスです。AI Platform Training を使用すると、実験をスケールアップし、分散トレーニングを活用し、GPU や TPU などの特殊なハードウェアにアクセスできます<sup>1</sup>。

Cloud Monitoring は、Google Cloud、AWS、その他のソースからメトリクス、ログ、トレースを収集して分析するサービスです。Cloud Monitoring を使用すると、データに基づいてダッシュボード、アラート、レポートを作成できます<sup>2</sup>。モニタリング API は、プログラムでモニタリング データにアクセスして操作できるようにするインターフェイスです<sup>3</sup>。

AI Platform Training と Cloud Monitoring を使用すると、手作業を最小限に抑えながら実験を追跡してレポートできます。AI Platform Training Python パッケージ<sup>4</sup> を使用して、実験の精度指標を Cloud Monitoring に書き込むことができます。その後、Monitoring API を使用して結果をクエリし、さまざまな実験のパフォーマンスを比較できます。Cloud Console で指標を視覚化したり、カスタム ダッシュボードやアラートを作成したりすることもできます<sup>5</sup>。したがって、AI Platform Training と Cloud Monitoring を使用することが、このユースケースには最適なオプションです。

参考文献:

- \* AI プラットフォーム トレーニングのドキュメント
- \* クラウドモニタリングのドキュメント
- \* モニタリング API の概要
- \* AI Platform トレーニングでの Cloud Monitoring の使用
- \* 評価指標の表示

**最新問題: 98**

You are developing a model to predict whether a failure will occur in a critical machine part. You have a dataset consisting of a multivariate time series and labels indicating whether the machine part failed. You recently started experimenting with a few different preprocessing and modeling approaches in a Vertex AI Workbench notebook. You want to log data and track artifacts from each run. How should you set up your experiments?

- A.
- B.
- C.
- D.

**Answer: B** ([メッセージを残す](#))

**最新問題: 99**

あなたは、需要が大幅に増加している玩具メーカーで働いています。品質管理検査官が製品の欠陥をチェックするのに費やす時間を短縮するには、ML モデルを構築する必要があります。より迅

速な欠陥検出が優先事項です。工場には信頼できる Wi-Fi がありません。あなたの会社は、新しい ML モデルをできるだけ早く実装したいと考えています。どのモデルを使用すべきですか？

- A. AutoML Vision Edge モバイル高精度-1モデル
- B. AutoML Vision モデル
- C. AutoML Vision Edge mobile-low-latency-1 モデル
- D. AutoML Vision Edge mobile-versatile-1 モデル

**Answer: B** ([メッセージを残す](#))

#### 最新問題: 100

機械学習スペシャリストは、Amazon SageMaker にカスタムアルゴリズムを導入したいと考えています。スペシャリストは、Amazon SageMaker によってサポートされる Docker コンテナにアルゴリズムを実装します。

Amazon SageMaker がトレーニングを正しく起動できるように、スペシャリストは Docker コンテナをどのようにパッケージ化する必要がありますか？

- A. コンテナ内の `bash_profile` ファイルを変更し、トレーニング プログラムを開始する `bash` コマンドを追加します。
- B. トレーニング プログラムをディレクトリ `/opt/ml/train` にコピーします。
- C. トレーニング プログラムを `ENTRYPOINT` という名前のトレインとして構成します
- D. Dockerfile の `CMD` 構成を使用して、トレーニング プログラムをイメージの `CMD` として追加します

**Answer: D** ([メッセージを残す](#))

#### 最新問題: 101

企業の機械学習スペシャリストは、TensorFlow を使用した時系列予測モデルのトレーニング速度を向上させる必要があります。現在、トレーニングはシングル GPU マシンで実装されており、完了までに約 23 時間かかります。トレーニングは毎日実行する必要があります。

モデルの精度は許容範囲内ですが、同社はトレーニング データのサイズが継続的に増加し、モデルを毎日ではなく時間ごとに更新する必要があると予想しています。同社はまた、コーディングの労力とインフラストラクチャの変更を最小限に抑えたいと考えています。

将来の需要に合わせて拡張できるようにするために、機械学習スペシャリストはトレーニング ソリューションに対して何をすべきでしょうか？

- A. 組み込みの AWS SageMaker DeepAR モデルの使用に切り替えます。ビジネス目標を達成するために必要な数のマシンに対してトレーニングを並列化します。
- B. TensorFlow コードを変更して、Amazon SageMaker でサポートされる Horovod 分散フレームワークを実装します。ビジネス目標を達成するために必要な数のマシンに対してトレーニングを並列化します。
- C. TensorFlow コードを変更しないでください。トレーニングを高速化するには、より強力な GPU を搭載したマシンに変更します。
- D. トレーニングを Amazon EMR に移動し、ビジネス目標を達成するために必要な数のマシンにワークロードを分散します。

**Answer: B** ([メッセージを残す](#))

最新問題: 102

Google Cloud 上で構造化データ用の ML パイプラインを再構築したいと考えています。あなたは PySpark を使用して大規模なデータ変換を行っていますが、パイプラインの実行に 12 時間以上かかっています。開発とパイプラインの実行時間を短縮するには、サーバーレス ツールと SQL 構文を使用する必要があります。生データはすでに Cloud Storage に移動されています。速度と処理の要件を満たしながら、Google Cloud 上にパイプラインを構築するにはどうすればよいでしょうか？

- A. PySpark を SparkSQL クエリに変換してデータを変換し、Dataproc でパイプラインを実行してデータを BigQuery に書き込みます。
- B. BigQuery Load を使用してデータを BigQuery に取り込み、PySpark コマンドを BigQuery SQL クエリに変換してデータを変換し、その変換を新しいテーブルに書き込みます。
- C. データを Cloud SQL に取り込み、PySpark コマンドを SQL クエリに変換してデータを変換し、機械学習に BigQuery のフェデレーション クエリを使用します。
- D. Data Fusion の GUI を使用して変換パイプラインを構築し、データを BigQuery に書き込みます。

**Answer: A** ([メッセージを残す](#))

最新問題: 103

あなたは、カスタマー サポートの電子メールを分類するモデルを開発しています。オンプレミスシステム上の小規模なデータセットを使用して TensorFlow Estimator でモデルを作成しましたが、高いパフォーマンスを確保するには大規模なデータセットを使用してモデルをトレーニングする必要があります。モデルを Google Cloud に移植し、オンプレミスからクラウドへの移行を容易にするために、コードのリファクタリングとインフラストラクチャのオーバーヘッドを最小限に抑えたいと考えています。あなたは何をするべきか？

- A. 分散トレーニングに Vertex AI Platform を使用する
- B. トレーニング用に Dataproc 上にクラスタを作成します
- C. 自動スケーリングを使用してマネージド インスタンス グループを作成する
- D. Kubeflow Pipelines を使用して、Google Kubernetes Engine クラスタでトレーニングします。

**Answer: A** ([メッセージを残す](#))

AI プラットフォームには kubeflow パイプラインも含まれています。それを使用するためにインフラストラクチャをセットアップする必要はありません。D の場合は、kubernetes クラスタエンジンをセットアップする必要があります。この質問は、インフラストラクチャのオーバーヘッドを最小限に抑えるよう求めています。

最新問題: 104

あなたは、オンライン マルチプレイヤー ゲームを作成するモバイル ゲームのスタートアップ企業で働いています。最近、あなたの会社では、ゲーム内で不正行為を行うプレイヤーが増加し、収

益の損失とユーザー エクスペリエンスの低下につながっていることが確認されました。ゲームセッション終了後にプレイヤーが不正行為をしたかどうかを判断するバイナリ分類モデルを構築し、不正行為を行ったプレイヤーを禁止するメッセージを他の下流システムに送信しました。モデルはテスト中に良好に動作しました。次に、モデルを次の場所にデプロイする必要があります。さらなる収益の損失を避けるために、ゲームセッションの完了後にすぐに分類を提供するソリューションを提供したいと考えています。あなたは何をするべきか？

- A. モデルを Vertex AI Model Registry にインポートします。Vertex Batch Prediction サービスを使用して、バッチ推論ジョブを実行します。
- B. モデル ファイルを Cloud Storage バケットに保存します。モデル ファイルを読み取り、Cloud Function 上でオンライン推論リクエストを行う Cloud Function を作成します。
- C. モデル ファイルを VM に保存します。予測リクエストがあるたびにモデル ファイルをロードし、VM 上で推論ジョブを実行します。
- D. モデルを Vertex AI モデル レジストリにインポートします。モデルをホストする Vertex AI エンドポイントを作成し、オンライン推論リクエストを作成します。

**Answer: D (メッセージを残す)**

オンライン推論は、単一または少数の予測リクエストをモデルに送信し、即座に応答を得るプロセスです<sup>1</sup>。オンライン推論は、オンラインゲームでの不正行為の検出など、タイムリーな予測が必要なシナリオに適しています。オンライン推論では、モデルがエンドポイントにデプロイされる必要があります。エンドポイントは、予測リクエストのサービス URL を提供するリソースです<sup>2</sup>。

Vertex AI Model Registry は、ML モデルのライフサイクルを管理できる中央リポジトリです<sup>3</sup>。カスタム モデルや AutoML モデルなど、さまざまなソースからモデルをインポートし、異なるバージョンやエイリアスに割り当てることができます<sup>3</sup>。オンライン予測用のサービス URL を提供するリソースであるエンドポイントにモデルをデプロイすることもできます<sup>2</sup>。

モデルを Vertex AI モデル レジストリにインポートすると、Vertex AI 機能を利用してモデルを監視および更新できます<sup>3</sup>。Vertex AI Experiments を使用すると、精度、精度、再現率、AUC など、さまざまなモデルバージョンのメトリクスを追跡および比較できます。Vertex AI Explainable AI を使用して、各入力フィーチャがモデルの予測にどの程度寄与したかを示すフィーチャ属性を生成することもできます。

モデルをホストする Vertex AI エンドポイントを作成すると、Vertex AI Prediction サービスを使用してオンライン推論リクエストを処理できます<sup>2</sup>。Vertex AI Prediction は、スケーラビリティ、信頼性、セキュリティ、ログなどのさまざまな利点を提供します<sup>2</sup>。Vertex AI API または Google Cloud コンソールを使用して、オンライン推論リクエストをエンドポイントに送信し、即座に分類を取得できます<sup>4</sup>。

したがって、シナリオに最適なオプションは、モデルを Vertex AI Model Registry にインポートし、モデルをホストする Vertex AI エンドポイントを作成し、オンライン推論リクエストを行うことです。

他のオプションは、バッチ予測の使用やモデル ファイルの毎回の読み込みなどの即時分類を提供しないか、より多くの開発とメンテナンスの労力が必要となる Vertex AI Prediction を使用しないため、シナリオには適していません。クラウド関数や VM の作成など。

参考文献:

- \* オンライン予測とバッチ予測 | 頂点 AI | グーグルクラウド
- \* モデルをエンドポイントにデプロイする | 頂点 AI | グーグルクラウド
- \* Vertex AI モデル レジストリの紹介 | グーグルクラウド
- \* オンライン予測を取得 | 頂点 AI | グーグルクラウド

#### 最新問題: 105

あなたの組織のコールセンターから、各通話における顧客の感情を分析するモデルを開発するよう依頼されました。コールセンターには毎日 100 万件を超える電話がかかり、データは Cloud Storage に保存されます。収集されたデータは通話が発信された地域から出てはならず、個人を特定できる情報 (PII) を保存または分析することはできません。データサイエンスチームは、SQL ANSI-2011 準拠のインターフェイスを必要とする視覚化とアクセス用のサードパーティ ツールを備えています。データ処理および分析用のコンポーネントを選択する必要があります。データパイプラインはどのように設計すべきでしょうか？

- A. 1 = データフロー、2 = Cloud SQL
- B. 1 = Pub/Sub、2 = データストア
- C. 1 = データフロー、2 = BigQuery
- D. 1 = Cloud Function、2 = Cloud SQL

**Answer:** ([解答を表示する](#))

#### 最新問題: 106

あなたは自動車会社の AI チームで働いており、TensorFlow と Keras を使用して視覚的欠陥検出モデルを開発しています。モデルのパフォーマンスを向上させるには、変換、トリミング、コントラスト調整などの画像拡張機能を組み込む必要があります。これらの関数を各トレーニング バッチにランダムに適用します。実行時とコンピューティング リソースの使用率を考慮してデータ処理パイプラインを最適化したいと考えています。あなたは何をするべきか？

- A. 拡張関数を tf.Data パイプラインに動的に埋め込みます。
- B. Keras ジェネレーターの一部として拡張関数を動的に埋め込みます。
- C. Dataflow を使用して、考えられるすべての拡張を作成し、TFRecord として保存します。
- D. Dataflow を使用して、トレーニングの実行ごとに拡張を動的に作成し、TFRecord としてステージングします。

**Answer: A** ([メッセージを残す](#))

実行時およびコンピューティング リソースの使用率に関してデータ処理パイプラインを最適化するための最良のオプションは、拡張関数を tf.Data パイプラインに動的に埋め込むことです。このオプションには次の利点があります。

\* データの追加コピーを作成したり保存したりすることなく、オンザフライでデータ拡張を実行できます。これにより、ストレージ容量が節約され、データ転送時間が短縮されます。

\* tf.Data API の並列性とパフォーマンスを活用し、複数の CPU コアまたは GPU デバイスを使用して、拡張関数を複数のデータ バッチに並列で効率的に適用できます。tf.Data API は、データ処理速度を向上させ、待ち時間を短縮するために、キャッシュ、プリフェッチ、自動チューニングなどのさまざまな最適化手法もサポートしています。

\* TensorFlow および Keras モデルとシームレスに統合され、tf.Data データセットをトレーニングと評価の入力として使用できます。tf.Data API は、画像、テキスト、オーディオ、ビデオなどのさまざまなデータ形式や、ファイル、データベース、Web サービスなどのさまざまなデータ ソースもサポートします。

他のオプションは、次の理由からあまり最適ではありません。

\* オプション B: Keras ジェネレーターの一部として拡張関数を動的に埋め込むと、いくつかの制限とオーバーヘッドが発生します。Keras ジェネレーターは、トレーニングまたは評価用のデータのバッチを生成する Python ジェネレーターです。ただし、Keras ジェネレーターは、トレーニングを複数のデバイスまたはマシンに分散するために使用される tf.distribute API と互換性がありません。さらに、Keras ジェネレーターは単一の Python スレッドで実行され、並列処理や最適化手法をサポートしていないため、tf.Data API ほど効率的またはスケラビリティがありません。

\* オプション C: Dataflow を使用して可能なすべての拡張を作成し、それらを TFRecord として保存すると、追加の複雑さとコストが発生します。Dataflow は、データの処理と変換のために Apache Beam パイプラインを実行するフルマネージド サービスです。ただし、Dataflow を使用して可能なすべての拡張を作成するには、大量の拡張画像を生成して保存する必要があり、大量のストレージスペースを消費し、ストレージとネットワークのコストが発生する可能性があります。さらに、Dataflow を使用して拡張機能を作成するには、別の Dataflow パイプラインを作成してデプロイする必要があり、面倒で時間がかかる可能性があります。

\* オプション D: Dataflow を使用してトレーニングの実行ごとに拡張を動的に作成し、それらを TFRecord としてステージングすると、追加の複雑さと遅延が生じます。Dataflow は、データの処理と変換のために Apache Beam パイプラインを実行するフルマネージド サービスです。ただし、Dataflow を使用してトレーニングの実行ごとに拡張を動的に作成するには、モデルがトレーニングされるたびに Dataflow パイプラインを実行する必要があるため、レイテンシーが発生し、トレーニング プロセスが遅れる可能性があります。さらに、Dataflow を使用して拡張機能を作成するには、別の Dataflow パイプラインを作成してデプロイする必要があり、面倒で時間がかかる可能性があります。

参考文献:

\* [tf.data: TensorFlow 入力パイプラインの構築]

\* [画像拡張 | TensorFlow コア]

\* [データフローのドキュメント]

有効な **Professional-Machine-Learning-Engineer** 問題集は GoShiken.com が提供された合格しやすい Professional-Machine-Learning-Engineer 試験問題集！ GoShiken.com が最新の **Professional-Machine-Learning-Engineer** 試験問題集を提供しています。GoShiken.com Professional-Machine-Learning-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Professional-Machine-Learning-Engineer 問題集をゲットする人はこちら：  
<https://www.goshiken.com/Google/Professional-Machine-Learning-Engineer-mondaishu.html>  
(29030%OFF問題集溶と正解付きで 30%w特別割引コード: **Freepdfdumps**)

#### 最新問題: 107

あなたの組織はオンライン掲示板を管理しています。数か月前、掲示板上で有害な言葉やいじめが増加していることに気づきました。特定のコメントに有害または有害であるとフラグを付ける自動テキスト分類器をデプロイしました。現在、一部のユーザーから、自分の宗教に言及した良性的コメントが誤って虐待として分類されていると報告されています。さらに詳しく調べてみると、過小評価されている特定の宗教団体に言及したコメントの分類器の誤検知率が高いことがわかりました。チームの予算は限られており、すでに予算が過剰になっています。あなたは何をするべきか？

- A. これらのフレーズが有害でない方法で使用される合成トレーニング データを追加します。
- B. モデルを削除し、人間によるモデレーションに置き換えます。
- C. モデルを別のテキスト分類子に置き換えます。
- D. コメントが有毒または有害であると見なされるしきい値を上げます。

**Answer: A (メッセージを残す)**

このアプローチは、毒性のない方法で使用されている宗教的なフレーズの例をさらに提供することで、分類子のパフォーマンスを向上させるのに役立ちます。これにより、分類子は、これらの宗教団体に言及する有害なコメントと有害でないコメントをより適切に区別できるようになります。さらに、合成データは、人的リソースに多大な投資を必要とせずに、既存のモデルのパフォーマンスを向上させるための費用対効果の高い方法です。

#### 最新問題: 108

機械学習スペシャリストは、企業の Web サイトに製品の推奨を提供するモデルをデプロイしました。当初、このモデルのパフォーマンスは非常に良好で、顧客は平均してより多くの製品を購入するようになりました。しかし、過去数か月以内に、スペシャリストは、製品推奨の効果が薄れ、顧客が支出を減らすという元の習慣に戻り始めていることに気づきました。

このモデルは1年以上前の最初の導入から変わっていないため、スペシャリストは何が起こったのかわかりません。

スペシャリストはモデルのパフォーマンスを向上させるためにどの方法を試すべきですか？

- A. モデルは、元のトレーニング データと製品在庫の変化に応じた新しいデータを使用して定期的に再トレーニングする必要があります。
- B. 製品在庫の変更に対応できないため、モデルを完全に再設計する必要があります。

C. 製品在庫の変更を処理するために正規化項を追加しながら、元のデータを使用してモデルを最初から定期的に再トレーニングする必要があります。

D. ドリフトを防ぐために、モデルのハイパーパラメータを定期的に更新する必要があります。

**Answer: A (メッセージを残す)**

#### 最新問題: 109

あなたは世界的な自動車メーカーの ML エンジニアです。世界中のさまざまな都市での自動車販売を予測するには、ML モデルを構築する必要があります。車のタイプと販売台数の間の都市固有の関係をトレーニングするには、どの特徴または特徴クロスを使用する必要がありますか？

A. ビン化された緯度、ビン化された経度、およびワンホット エンコードされた車種の 3 つの個別の機能

B. 緯度、経度、車種の要素ごとの積として取得される 1 つの特徴

C. ビン化された緯度、ビン化された経度、およびワンホット エンコードされた車種の間要素ごとの積として取得される 1 つの特徴

D. 2 つの特徴が要素ごとの積として交差します。1 つ目はビン化された緯度とワンホット エンコードされた車種の間であり、2 つ目はビン化された経度とワンホット エンコードされた車種の間です。

**Answer: C (メッセージを残す)**

特徴クロスは、2 つ以上の既存の特徴を結合することによって、通常はそれらの積または連結によって取得される合成特徴です。特徴クロスは、元の特徴間の非線形効果と相互作用効果を捕捉し、モデルの予測パフォーマンスを向上させるのに役立ちます。特徴クロスは、数値特徴、カテゴリ特徴、地理空間特徴など、さまざまなタイプの特徴に適用できます<sup>1</sup>。

世界中のさまざまな都市での自動車販売を予測する ML モデルを構築するユースケースの場合、最適なオプションは、ビン化された緯度、ビン化された経度、およびワンホット エンコードされた車種の間要素ごとの積として取得された 1 つの特徴を使用することです。このオプションには、ビン化された緯度、ビン化された経度、ワンホット エンコードされた車種という 3 つの個別のフィーチャを組み合わせたフィーチャ クロスの作成が含まれます。ビン化は、範囲を等間隔、つまりビンに分割することによって、連続的な数値特徴を離散的なカテゴリ特徴に変換する手法です。ワンホット エンコーディングは、カテゴリ特徴量をバイナリ ベクトルに変換する手法です。各要素は考えられるカテゴリに対応し、特徴量がそのカテゴリに属する場合は値 1、それ以外の場合は 0 になります。緯度、経度、および車種の特徴にビン化とワンホット エンコーディングを適用することで、ビンと車種の各組み合わせが異なる都市を表すことができるため、特徴クロスは車種と販売台数の間の都市固有の関係を捉えることができます。そして特定の車種に対する好み。

たとえば、特徴クロスは、緯度ビンが [40, 50]、経度ビンが [-80, -70]、車種が SUV の都市が、緯度ビンは [-10, 0]、経度ビンは

【0, 20】、セダンの車種。したがって、ビン化された緯度、ビン化された経度、およびワンホット エンコードされた車種の間要素ごとの積として取得された 1 つの特徴を使用することが、このユースケースに最適なオプションです。

## 参考文献:

\* フィーチャークロス | 機械学習短期集中コース

### 最新問題: 110

あなたは、データセットをクリーンアップして Cloud Storage バケットに保存するパイプラインを開発したデータ エンジニアリング チームと協力します。ML モデルを作成し、新しいデータが利用可能になったらすぐにそのデータを使用してモデルを更新したいと考えています。CI/CD ワークフローの一部として、Google Kubernetes Engine (GKE) で Kubeflow Pipelines トレーニング ジョブを自動的に実行したいと考えています。このワークフローをどのように設計すればよいでしょうか？

- A. App Engine を使用して、Cloud Storage に新しいファイルを継続的にポーリングする軽量の Python クライアントを作成します。ファイルが到着したらすぐにトレーニング ジョブを開始します。
- B. Cloud Scheduler を使用して、定期的な間隔でジョブをスケジュールします。仕事の第一歩として、Cloud Storage バケット内のオブジェクトのタイムスタンプを確認します。前回の実行以降に新しいファイルがない場合は、ジョブを中止します。
- C. ストレージ バケットで新しいファイルが利用可能になったときに Pub/Sub トピックにメッセージを送信するように Cloud Storage トリガーを構成します。Pub/Sub トリガーの Cloud Function を使用して、GKE クラスタでトレーニング ジョブを開始する
- D. Dataflow を使用してパイプラインを構成します。これにより、ファイルが Cloud Storage に保存されます。ファイルが保存されたら、GKE クラスタでトレーニング ジョブを開始します。

**Answer: C** ([メッセージを残す](#))

### 最新問題: 111

あなたは、プロトタイプのコードを実稼働環境にデプロイするという任務を与えられました。特徴エンジニアリング コードは PySpark にあり、Dataproc サーバーレスで実行されます。モデルのトレーニングは、Vertex AI カスタム トレーニング ジョブを使用して実行されます。2つのステップは接続されていないため、現時点では、特徴エンジニアリング ステップの終了後にモデル トレーニングを手動で実行する必要があります。エンドツーエンドで実行され、ステップ間の接続を追跡する、スケーラブルで保守可能な生産プロセスを作成する必要があります。あなたは何をすべきか？

- A. Vertex AI Workbench ノートブックを作成する ノートブックを使用して Dataproc サーバーレス機能エンジニアリング ジョブを送信する 同じノートブックを使用してカスタム モデル トレーニング ジョブを送信する ノートブックのセルを順番に実行して、ステップをエンドツーエンドで結び付ける
- B. Vertex AI Workbench ノートブックを作成します。ノートブックで Apache Spark コンテキストを開始し、PySpark 機能エンジニアリング コードを実行します。同じノートブックを使用して、TensorFlow でカスタム モデル トレーニング ジョブを実行します。ノートブックのセルを順番に実行して、ステップを結び付けます。エンドツーエンドで一緒に

C. Kubeflow パイプライン SDK を使用して、2 つのコンポーネントを指定するコードを作成します

- 1 つ目は、機能エンジニアリング ジョブを起動する Dataproc サーバーレス コンポーネントです。

- 2 番目は、

create\_cusrora\_raining\_job\_from\_ccraponent Utility that launches the custom model training job.

D. Vertex AI Pipelines ジョブを作成して両方のコンポーネントをリンクして実行します

Kubeflow パイプライン SDK を使用して 2 つのコンポーネントを指定するコードを作成します

- 最初のコンポーネントは、PySpark 機能エンジニアリング コードを実行する Apache Spark コンテキストを開始します。

- 2 番目のコンポーネントは、TensorFlow カスタム モデル トレーニング コードを実行します。

Vertex AI Pipelines ジョブを作成して、両方のコンポーネントをリンクして実行します。

**Answer: C** ([メッセージを残す](#))

#### 最新問題: 112

あなたは製造会社のデータサイエンス チームで働いています。あなたは、数億件のレコードが含まれる会社の過去の売上データを確認しています。探索的なデータ分析では、平均、中央値、最頻値などの記述統計を計算する必要があります。仮説検証のための複雑な統計的テストを実施する。計算リソースを最小限に抑えながら、分析にできるだけ多くの販売データを使用したいと考えています。あなたは何をするべきか？

A. Vertex AI Workbench ユーザー管理ノートブック インスタンスをスピンアップし、データセットをインポートします。このデータを使用して統計分析と視覚分析を作成します。

B. Google データポータルで時間プロットを視覚化します。データセットを Vertex AI Workbench ユーザー管理ノートブックにインポートします。このデータを使用して記述統計を計算し、統計分析を実行します。

C. BigQuery を使用して記述統計を計算します。Vertex AI Workbench のユーザー管理ノートブックを使用して、時間プロットを視覚化し、統計分析を実行します。

D BigQuery を使用して記述統計を計算し、Google データスタジオを使用して時間プロットを視覚化します。Vertex AI Workbench のユーザー管理ノートブックを使用して、統計分析を実行します。

**Answer: C** ([メッセージを残す](#))

計算リソースを最小限に抑えながら大規模で複雑なデータセットを分析するための最良のオプションは、BigQuery と Vertex AI Workbench を組み合わせて使用することです。BigQuery は、ペタバイト規模のデータに対して高速かつインタラクティブなクエリを実行できる、サーバーレスでスケーラブルでコスト効率の高いデータ ウェアハウスです。BigQuery は、AVG、PERCENTILE\_CONT、MODE などの SQL 関数を使用して、平均、中央値、最頻値などの記述統計を計算できます。。Vertex AI Workbench は、データサイエンスと機械学習のための統合開発環境を提供するマネージド サービスです。Vertex AI Workbench を使用すると、ユーザーは Google Cloud 上で Jupyter ノートブックを作成して実行し、データの視覚化や統計分析のため

のさまざまなツールやライブラリにアクセスできます。Vertex AI Workbench は BigQuery に接続し、クエリの結果を使用して時間プロットを作成し、仮説検証のための統計テストを実行できます。BigQuery と Vertex AI Workbench を使用することで、ユーザーは Google Cloud のパワーと柔軟性を活用して、大規模で複雑なデータセットに対して探索的なデータ分析を実行できます。

参考文献:

- \* Google Cloud 認定資格の準備: 機械学習エンジニア、コース 2: Google Cloud での ML のためのデータ エンジニアリング、第 1 週: ML のためのデータ エンジニアリングの概要
- \* Google Cloud Professional Machine Learning Engineer 試験ガイド、セクション 1: ローコード ML ソリューションの設計、1.1 BigQuery ML を使用した ML モデルの開発
- \* 公式 Google Cloud Certified Professional Machine Learning Engineer Study Guide、第 3 章: ML のためのデータ エンジニアリング、セクション 3.2: ML のための BigQuery

**Valid Professional-Machine-Learning-Engineer Dumps** shared by GoShiken.com for Helping Passing Professional-Machine-Learning-Engineer Exam! GoShiken.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the GoShiken.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** GoShiken.com Professional-Machine-Learning-Engineer dumps with Test Engine here: <https://www.goshiken.com/Google/Professional-Machine-Learning-Engineer-mondaishu.html> (290 Q&As Dumps, **30%OFF Special Discount: Freepdfdumps**)