

Oracle Machine Learning AutoML Autonomous Database でのユーザー・インタフェース

データ・サイエンティスト向けの機械学習の自動化、
シチズン・データ・サイエンティスト向けの機械学習の簡略化、
アプリケーション開発者向けの機械学習モデルの迅速な
デプロイメント

2021 年 3 月、バージョン 1.0

Copyright © 2021, Oracle and/or its affiliates

Public

目的

このドキュメントでは、Oracle Machine Learning AutoML ユーザー・インタフェースに含まれる機能の概要を説明します。このドキュメントは、OML AutoML ユーザー・インタフェースがビジネスに及ぼすメリットの評価および IT プロジェクトのプランニングを支援することのみを目的としています。

免責事項

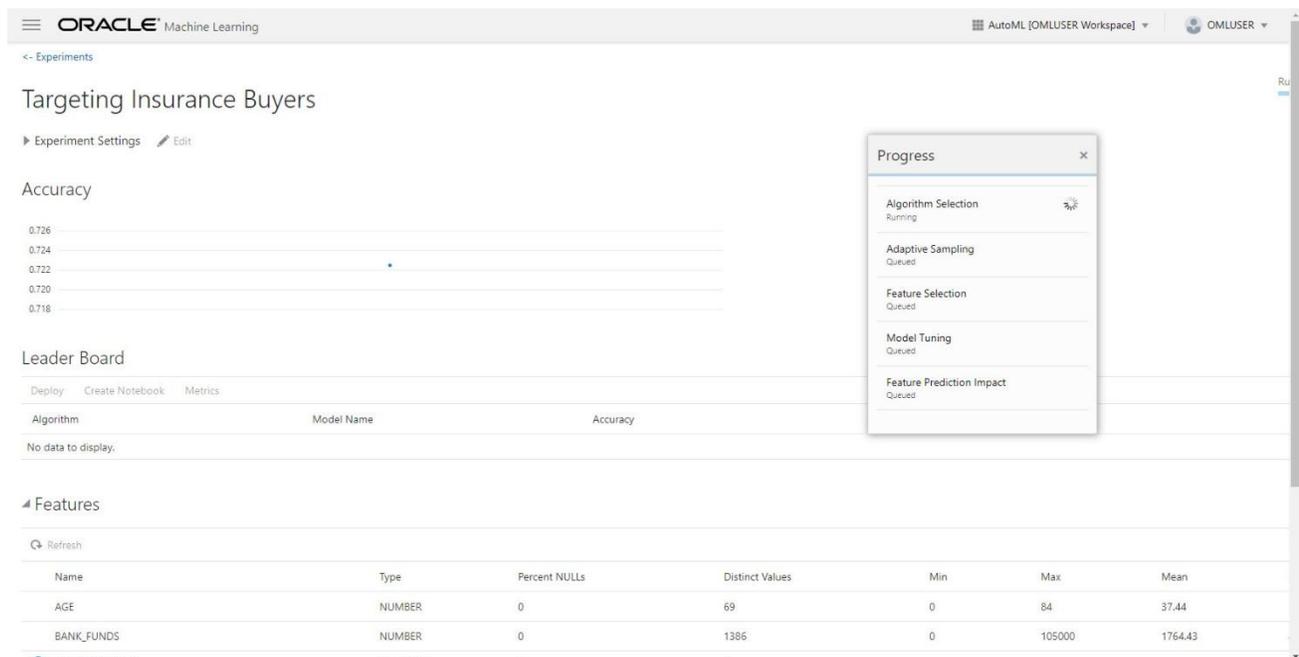
このドキュメントには、ソフトウェアまたは印刷物などの形式を問わず、オラクルが独占的な権利を有する財産的情報が含まれています。この機密資料へのアクセスと使用は、オラクルとの間で締結され遵守に同意したオラクル・ソフトウェア・ライセンスおよびサービス契約の条件に従うものとします。このドキュメントとその内容の開示、コピー、複製および配布には、オラクルによる事前の書面による承諾を必要とします。このドキュメントはライセンス契約の一部となるものではなく、オラクルおよびその子会社や関連会社との契約を構成するものではありません。

このドキュメントは参照専用であり、製品機能の実装とアップグレードのプランニングを支援することのみを目的としています。マテリアル、コード、または機能の提供をお約束するものではなく、購入決定を行う際の判断材料とすべきではありません。このドキュメントに記載されている特徴または機能の開発、リリースおよび時期については、オラクルの裁量により決定されます。製品アーキテクチャの性質上、コードが大幅に不安定化するリスクなしに、本書に記載されているすべての機能を安全に含めることができない場合があります。

概要

Oracle Autonomous Database の Oracle Machine Learning (OML)では、データベース内機械学習(ML)のアルゴリズムと機能にアクセスできます。OML AutoML ユーザー・インターフェイス(UI)には、通常は専門のデータ・サイエンティストが長い時間を要していた反復タスクを自動化する使いやすいインターフェイスが提供されているため、専門家以外のユーザーも機械学習を簡単に利用できます。OML AutoML UI により、モデル作成からモデル・デプロイメントまでの機械学習プロセスをスピードアップできます。

数回のクリック操作でデータ表およびターゲット属性を指定すると、いくつかのモデルが OML AutoML UI によって作成され、検討を開始できます。OML AutoML UI により、データの前処理、データベース内アルゴリズムを実験するための最適候補の選択、モデル品質を高めるための適切な入力データ・サンプルと特徴の選択、およびモデル・チューニングのスピードアップが自動的に実行されます。OML AutoML UI では、選択したアルゴリズムを使用して OML のモデルが作成され、ハイパーパラメータがチューニングされて精度メトリックが表示されるため、ユーザーのニーズに最適なモデルを選択できます。



Oracle Machine Learning AutoML の使いやすいユーザー・インターフェイスでは、アルゴリズムと特徴の選択、モデルのチューニングとデプロイメントなど、機械学習の通常のステップがデータベース内の OML アルゴリズムを使用して自動的に実行されます。OML AutoML UI では、シチズン・データ・サイエンティストが機械学習モデルを作成および選択できるため、データ・サイエンティストの生産性が向上し、モデルのデプロイメントがスピードアップします。

OML AutoML ユーザー・インターフェイスの特徴

Oracle Machine Learning AutoML ユーザー・インターフェイスを使用すると、シチズン・データ・サイエンティストと専門のデータ・サイエンティストの両方が機械学習モデルを簡単に作成およびデプロイできます。Oracle Autonomous Database の Oracle Machine Learning の新しいコンポーネントである OML AutoML UI では、機械学習のモデリング・プロセスを自動化する、コード不要でブラウザベースのインターフェイスが提供されており、数回のクリック操作で簡単にデプロイできます。

データ・サイエンティストは OML4Py AutoML 機能を使用して自分でコードを作成できますが、OML AutoML UI を利用すると、ML を仕事で使えることはわかっているものの、使用に耐えるモデルに仕上げるほど特定のアルゴリズムの詳細やモデリング・プロセスに詳しくないユーザーも機械学習を利用できるようになります。

同時に、データ・サイエンティストが OML AutoML UI をモデリング・アクセラレータとして利用すると、初期モデルの作成を自動化して特定のハイパーパラメータを継続的にチューニングしたり、OML AutoML UI で生成したノートブックのモデルを直接拡張して特定のモデルの OML4Py コードを再生成できるため、生産性が大幅に向上します。

OML AutoML UI を使用すると、機械学習が簡略化され、時間がかかる定型ステップが自動化されてデータ・サイエンティストの生産性が向上し、機械学習モデルの提供に必要な計算時間を削減できます。OML AutoML UI では、次のことを行えます。

- シチズン・データ・サイエンティストは、数回のクリック操作で機械学習モデルをすばやく作成してノートブックを生成し、データ・サイエンス・チームと協力して、ノートブックを拡張し実行をスケジュールできます。
- データ・サイエンティストは、OML のデータベース内アルゴリズムおよび機械学習機能を利用しながら、機械学習のベストプラクティスが自動適用された複数のモデルを簡単に作成、調査および評価して生産性を向上できます。
- アプリケーション開発者は、データベース内スコアリングを利用して機械学習モデルを迅速にデプロイし、生成された OML ノートブックを変更して、OML サービスの REST エンドポイントを使用してアプリケーションを統合できます。

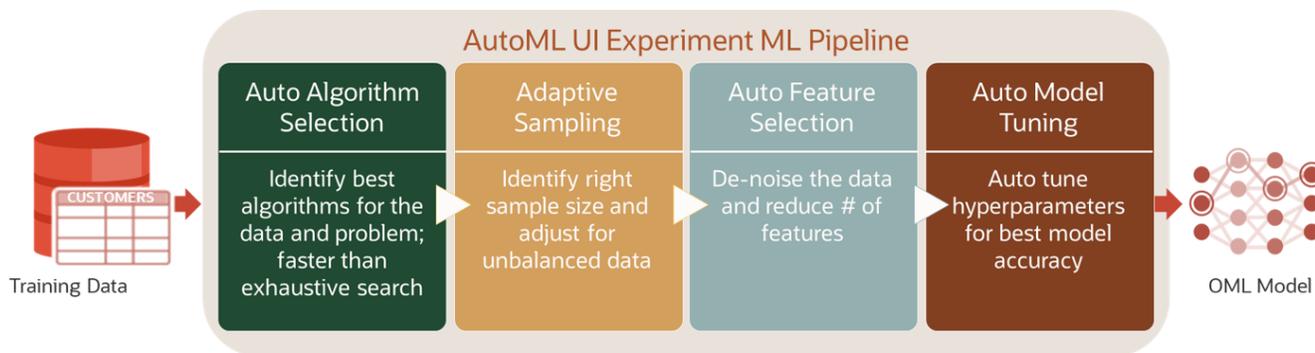
人工知能、機械学習、自律システムは、以前のものとは根本的に異なるため、コンピュータ・テクノロジーの新しい世代の始まりを感じます。

ラリー・エリソン

オラクル社会長兼最高技術責任者

機械学習パイプラインの自動化

OML AutoML UI では、**実験**という概念を使用して、アルゴリズムと特徴の選択、データ・サンプリング、モデルの作成と評価、ハイパーパラメータのチューニングなど、データ・サイエンティストが長い時間を要していた反復タスクを自動化する機械学習パイプラインが作成されます。



OML AutoML ユーザー・インタフェースでは、アルゴリズムの選択、データ・サンプリング、特徴の選択、モデルのチューニングという、時間と根気が必要な 4 つの主要ステップが自動化され、データ・サイエンティストとシチズン・データ・サイエンティストの生産性が大幅に向上しました。

アルゴリズムの選択

AutoML では、指定されたデータおよびターゲットのモデルの作成に対して、最も有望なアルゴリズムが実験ごとに特定されます。広範なデータセットから作成されたモデルを使用したメタ学習により、アルゴリズムが自動的に選択されます。ここでは、値の分布つまりデータのメタ特徴に基づいて、どのアルゴリズムが最善の結果になりそうかが事前作成済のモデルにより予測されます。その後、最高スコアのアルゴリズムがモデルのチューニングに使用されます。これにより、データ・サイエンティストおよび専門家以外のユーザーが、しらみつぶし探索をするよりも早く最善のアルゴリズムを見つけることができます。これにより、計算コストも削減できます。

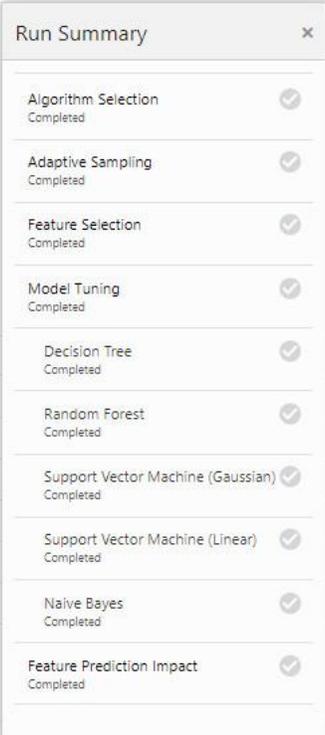
アダプティブ・サンプリング

OML AutoML UI では、アダプティブ・サンプリングを使用して、サンプル・サイズを大きくしてもモデル品質がそれ以上向上しなくなるようなサンプリングの適正割合が見極められます。これにより、モデルの作成もスピードアップします。さらに、アダプティブ・サンプリングでは、モデルの作成に悪影響を及ぼす不均衡なデータセットが検出され、必要に応じて層別サンプリングが採用されるため、よりよいモデルの作成用に均衡の取れたデータセットが作成されます。

特徴の選択

ターゲット属性と相関がない属性、定数や欠落値が多すぎる属性、カーディナリティが高すぎる属性はモデル品質を悪くする可能性があり、モデルの作成時間およびデータ・スコアリング時間も長くなります。OML AutoML UI では、入力データが前処理され、ほとんど情報が含まれていない属性やノイズのような属性が自動的に削除されます。OML AutoML UI では、次のいくつかの方法で、まず特徴がランク付けされ、ランキングに基づいてサブセットが評価されます。

- ピアソンの相関係数(SQL CORR 関数の関数)によって計算される相関スコア
- トレーニングされたランダム・フォレスト・モデルからの属性重要度
- トレーニングされた一般線形モデルからの係数
- 最小記述長モデルからの属性重要度



Run Summary	
Algorithm Selection Completed	✓
Adaptive Sampling Completed	✓
Feature Selection Completed	✓
Model Tuning Completed	✓
Decision Tree Completed	✓
Random Forest Completed	✓
Support Vector Machine (Gaussian) Completed	✓
Support Vector Machine (Linear) Completed	✓
Naive Bayes Completed	✓
Feature Prediction Impact Completed	✓

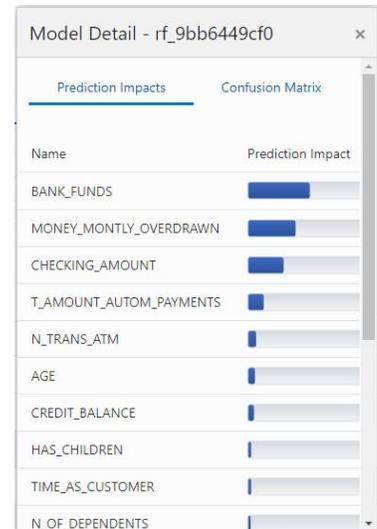
応答時間を速くするために、特徴の自動選択でメタ学習を内部的に使用して、評価対象になる可能性のあるランキングおよび特徴サブセットの数を絞り込んでいます。これにより、モデルの作成およびハイパーパラメータのチューニング時間を短縮しています。OML AutoML UI では、候補モデルの入力特徴のランキングおよび選択に対して、このような多面的なアプローチを採用しています。入力特徴は、ユーザーが手動で上書きして追加または削除できます。

モデルのチューニング

特徴を選択すると、OML AutoML UI により、OML のデータベース内処理を利用して、複数の機械学習モデルが作成およびチューニングされます。OML AutoML UI では、選択したアルゴリズムごとにモデルのチューニングが実行され、しらみつぶし探索をしなくても、モデルが大幅に向上するハイパーパラメータ設定が見つかります。たとえば、ディジション・ツリーでは、AutoML により、アルゴリズムの最大ツリー深度と最小分割割合が探索され最適化されます。ニューラル・ネットワークでは、AutoML により、一定範囲の非表示レイヤーおよびレイヤーごとのニューロンが探索されます。OML AutoML UI では、時間がかかる可能性がある「より良い精度」か、良いモデルをより速く作成しようとする「より速い結果」をユーザーが選択できます。OML ではモデルの作成時に、データベース内の自動データ準備(ADP)を利用して、欠落値、データのビン化、データの正規化、データのスケールリングが処理されます。

特徴予測の影響

OML AutoML UI では、モデルを作成するたびに、モデルに依存しないグローバルな説明方法でモデルの動作に対するインサイトが提供されます。最上位属性およびターゲット属性に対するその相対的な影響が表示されます。「特徴予測の影響」では、モデルが作成された後、各属性の現在の値がランダムに入れ替えられて様々なレコードに割り当てられ、トレーニングされた機械学習モデルの予測で入れ替える際の各特徴に及ぼす影響に基づいて、特徴の重要度が予測されてランク付けされます。



サポート対象の機械学習の機能、アルゴリズムおよびモデル・メトリック

OML AutoML UI では、次の機械学習の機能、データベース内アルゴリズムおよびスコア・メトリックがサポートされます。

マイニング機能	サンプル・ユース・ケース	アルゴリズム	モデル・メトリック
分類(バイナリおよびマルチクラス) コレクション内のアイテムをターゲット・カテゴリまたはクラスに割り当てます	ローンの申込みの与信リスクを低、中、高に特定します 可能性が高い買い手を予測します 従業員の自然減を予測します	単純ベイズ 一般化線形モデル 一般化線形モデル(リッジ回帰) Support Vector Machine(線形) Support Vector Machine(ガウス) ディビジョン・ツリー ランダム・フォレスト ニューラル・ネットワーク	精度 バランスの取れた精度 F1 適合率 再現率 ROC AUC 再現率マイクロ 再現率マクロ 再現率加重 適合率マイクロ 適合率マクロ 適合率加重 混同行列 予測の影響
回帰 連続体に沿った数値を予想します	家の価値を予測します 顧客生涯価値(LTV)を見積もります 次を見積もります	一般化線形モデル 一般化線形モデル(リッジ回帰) Support Vector Machine(線形) Support Vector Machine(ガウス) ニューラル・ネットワーク	R2 乗 負の平均二乗誤差 負の平均絶対誤差 負の中央絶対誤差 予測の影響

注意: OML AutoML UI では、非構造化テキスト・データ、トランザクション・データの集計の結果生じるネストされたデータやパーティション化されたモデルはサポートされません。

モデル開発のスピードアップ

OML AutoML UI で機械学習モデルを作成すると、データサイエンティストとアプリケーション開発者は複数の戦略を使用して OML モデルをデプロイできます。

Oracle Machine Learning Notebooks で使用するノートブックの生成

OML AutoML UI でモデルの作成が終わると、実験で作成した各モデルから OML Notebooks でノートブックを生成できます。このようなノートブックはカスタマイズ可能で、ジョブとして自動実行をスケジュールして、データ・スコアリングに使用できます。データサイエンティストとアプリケーション開発者は、OML4Py および OML4SQL のインターフェースを使用して、生成されたノートブックを変更および拡張できます。

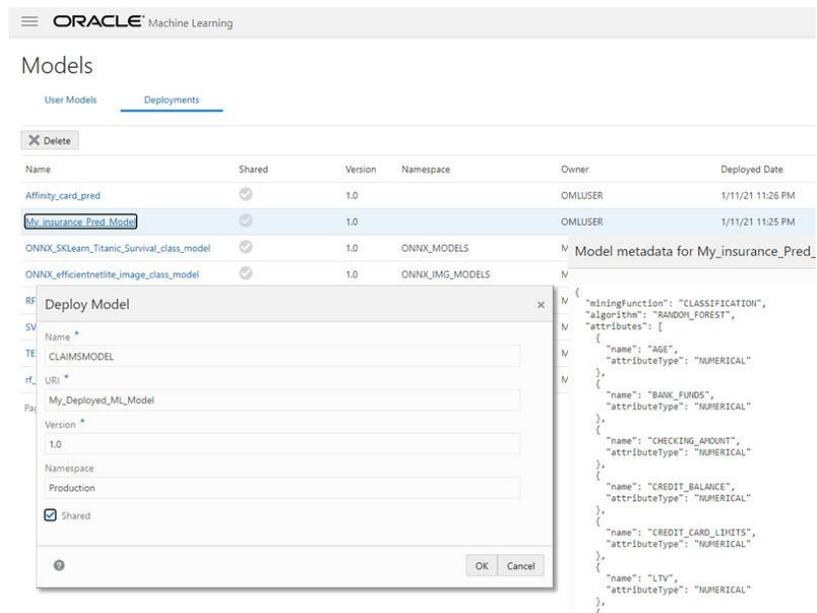
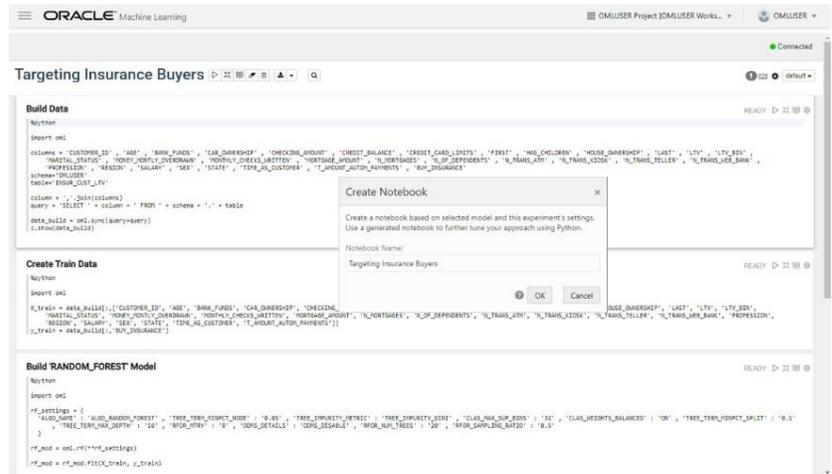
モデル・リポジトリ

Oracle Autonomous Database の OML では、モデルの管理およびデータベース内モデルのデプロイメントを可能にする「モデル」インターフェースもサポートされています。

モデルのデプロイメント

データベース内のモデルはいくつかの方法でデプロイできます。前述のとおり、OML AutoML UI のユーザーは、実験で生成されたモデルから OML ノートブックを生成できます。さらに、データベース内モデルを SQL 問合せで直接使用して、SQL Developer、SQL Developer Web、Oracle Applications Express (APEX) および Oracle Analytics Cloud などのツールからデータをスコアリングできます。オンプレミスでもクラウドでも、モデルを作成したのと同じデータベースまたは別の Oracle Database でモデルを使用し、モデルのエクスポートおよびインポート機能を使用して OML モデルを本番化します。

さらに、OML AutoML UI のユーザーは数回のクリック操作で、Oracle Machine Learning Services にモデルをデプロイできます。このサービスでは、OML モデル用の REST エンドポイントでモデルの管理とデプロイメントがサポートされています。



概要

OML AutoML UI では、通常は専門のデータサイエンティストが長い時間を要していた反復タスクを自動化する使いやすいインターフェースが提供されており、専門家以外のユーザーも機械学習を簡単に利用できます。OML AutoML UI により、すべてのユーザーのモデル作成からモデル・デプロイメントまでの機械学習プロセスをスピードアップできます。OML AutoML UI により、機械学習は大きく進歩し、よりシンプルで強力が機能豊富になり、企業が実行可能なインサイトにデータをすばやく変換できるようになります。

関連資料

[OML AutoML UI User's Guide](#)

[OML User's Guide](#)

[OML4Py User's Guide](#)

[OML4SQL User's Guide](#)

[Oracle Machine Learning](#)

オラクルの情報を発信しています

+1.800.ORACLE1までご連絡いただくか、[oracle.com](#)をご覧ください。北米以外の地域では、次のサイトで最寄りの営業所をご確認いただけます：[oracle.com/contact](#).

 [blogs.oracle.com](#)

 [facebook.com/oracle](#)

 [twitter.com/oracle](#)

Copyright © 2021, Oracle and/or its affiliates. このドキュメントは情報提供のみを目的としており、記載内容は予告なしに変更される場合があります。このドキュメントに誤りが無いことの保証はいたしかねます。また、口頭で明示されたか法律で黙示されたかを問わず、商品性もしくは特定の目的に対する適合性についての黙示的な保証を含め、いかなる保証や条件を提供するものでもありません。オラクル社はこのドキュメントに関するいかなる法的責任も明確に否定し、直接的、間接的を問わず、本ドキュメントにより、いかなる契約上の義務を構成するものではありません。このドキュメントは、オラクル社による事前の書面による承諾を得ることなく、目的の如何を問わず、電子的手段または印刷によるものも含めていかなる形式や手段によっても複製または送信することは禁じられています。

このデバイスは連邦通信委員会の規則で求められている承認を受けていません。このデバイスは、承認が取得されるまで、販売またはリリースされない場合があります。

OracleおよびJavaはオラクルおよびその関連会社の登録商標です。その他の社名、商品名等は各社の商標または登録商標である場合があります。

Intel, Intel Insideは、Intel Corporationの商標または登録商標です。すべてのSPARCの商標はライセンスのもとに使用し、SPARC International, Inc.の商標または登録商標です。AMD、Epyc、AMDロゴは、Advanced Micro Devices, Inc.の商標または登録商標です。UNIXは、The Open Groupの登録商標です。0120

免責事項：データシートに免責事項が必要かどうか分からない場合は、収益認識ポリシーをお読みください。内容および免責事項の要件について質問がある場合は、REVREC_US@oracle.comに電子メールでご連絡ください。