

Oracle Machine Learning 技術概要

データではなくアルゴリズムを移動する

2023年4月、バージョン1.0

Copyright © 2023, Oracle and/or its affiliates

公開

本書の目的

本書では、Oracle Machine Learningの機能の概要と強化された点が説明されています。本書は、Oracleのデータ管理およびデータ分析プラットフォームとOracle Machine Learningの利用によるビジネス上の利点の評価、およびデータ主導のデータ・サイエンス、機械学習、情報テクノロジーの各プロジェクトの計画立案を支援することのみを目的としています。

対象読者

この技術概要は、経営幹部、事業部門マネージャー、およびスケーラブルな機械学習機能を企業全体のデータ・サイエンティスト、データ・アナリスト、データ・エンジニア、アプリケーションおよびダッシュボード開発者に提供しようとしている担当者を対象としています。

免責事項

本文書には、ソフトウェアや印刷物など、いかなる形式のものも含め、オラクルの独占的な所有物である占有情報が含まれます。この機密文書へのアクセスと使用は、締結および遵守に同意したOracle Software License and Service Agreementの諸条件に従うものとします。本文書と本文書に含まれる情報は、オラクルの事前の書面による同意なしに、公開、複製、再作成、またはオラクルの外部に配布することはできません。本文書は、ライセンス契約の一部ではありません。また、オラクル、オラクルの子会社または関連会社との契約に組み込むことはできません。

本書は情報提供のみを目的としており、記載した製品機能の実装およびアップグレードの計画を支援することのみを意図しています。マテリアルやコード、機能の提供をコミットメント（確約）するものではなく、購買を決定する際の判断材料になさらないでください。本書に記載されている機能の開発、リリース、および時期については、弊社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。

目次

本書の目的	2
対象読者	2
免責事項	2
はじめに	4
概要	8
機械学習とは	13
データではなくアルゴリズムを移動する	13
Oracle Machine Learningとは	13
SQLとオープンソースのRおよびPythonのサポート	14
Oracle Machine Learningによるインデータベース処理	15
自動化による生産性の促進	15
オープンソースのRおよびPythonとOracle Databaseとの統合	16
Oracle Machine Learning AutoMLユーザー・インタフェース	18
Oracle Machine Learning Services	19
Oracle Data Miner—SQL Developerの拡張機能	20
データ・レイクのOracle Machine Learning	22
Oracle Machine Learning—AIアプリケーションと ダッシュボードを最短期間で提供	22
結論	24
追加情報	24

はじめに

データの移動を最小限に抑えながら、データが存在する場所でデータを大規模に分析。データサイエンスおよび機械学習ソリューションのアーキテクチャと展開を簡素化。

オラクルのコンバインド・データベースでは、統合されたセキュアな単一プラットフォームで、機械学習やその他の重要な分析テクノロジーをすぐに利用でき、同時にデータ・アクセスを拡大してチームと共有することができます。Oracle Machine Learningは、オラクルのコンバインド・データベース戦略の重要な構成要素として、データ主導のソリューションの迅速な開発とデプロイを支援します。

Oracle AIプラットフォームの戦略的な構成要素であるOracle Machine Learningを使用すれば、データサイエンティストとデータアナリストは、Oracle DatabaseとOracle Autonomous Databaseで大量のデータを直接処理することにより、知識を抽出し、新しいインサイトを発見し、



データに基づいて予測することができます。Oracle Machine Learningは、アプリケーションやダッシュボードをサポートする機械学習モデルやソリューションを容易にデプロイして使用できるようにすることで、アプリケーション開発者を支援します。この技術概要では、Oracle Machine Learningの概要と、この製品によって企業のデータサイエンスチームがデータの価値を向上させる方法について紹介します。

Oracle Machine Learning (OML) のユーザーは操作方法を柔軟に選択できます。SQL、Python、R、RESTなどの一般的な言語を使用することも、Oracle Machine Learning Notebooks、Oracle Data Minerユーザー・インタフェース、Oracle Machine Learning AutoMLユーザー・インタフェースなどを使用することも可能です。また、Oracle Machine Learning Servicesでは、RESTインタフェースからのAutonomous Databaseによるモデルの管理とデプロイ、およびデータとモデルの監視もサポートされています。



Oracle Autonomous Databaseでは、**Oracle Machine Learning Notebooks**を使用して、SQL、PL/SQL、Python、Rのスクリプトを開発および実行でき、ノートブック内のドキュメントではマークダウン言語を使用

できます。Oracle Machine Learning Notebooksは、Oracle Autonomous Databaseを機械学習およびデータサイエンス向けのプラットフォームとして拡張したもので、データサイエンティスト、データエンジニア、データアナリスト、アプリケーション開発者、ITプロフェッショナル（DBAを含む）をサポートします。Oracle Machine Learning Notebooksは、セキュリティを維持して権限をOracle Databaseのデータに継承できるようにしながら、リアルタイムの同時編集セッション、統合グラフィック、ノートブックおよびテンプレートの容易な共有、ならびに権限ベースのアクセス、バージョンング、およびジョブ・スケジューリングを通じてコラボレーションを促進します。

Oracle Autonomous Databaseで**Oracle Machine Learning**

AutoMLユーザー・インターフェースを使用すれば、最小限の入力情報、つまり予測対象のデータとターゲットを指定するだけでインデータベースMLモデルを構築でき、後のモデル生成については、機械学習メタモデルに従って



ツールによって実行されます。Oracle Machine Learning AutoML UIにより、データ・サイエンティストの生産性が向上し、さらに専門家でないユーザーも機械学習を利用しやすくなります。Oracle Machine Learning AutoML UIは、Oracle Machine Learning Servicesにインデータベース・モデルをデプロイし、OML4Py Python APIを使用して、選択したモデル用の編集可能ノートブックを生成するためにも使用できます。関連のOracle Machine Learning Models インタフェースは、モデルとデプロイメントの管理に使用できます。

{REST:API}

Oracle Machine Learning Servicesは、モデルの管理、デプロイ、監視をサポートし、システムにより提供されるインフラストラクチャや統合データベース・アーキテクチャも活用できます。

このモデル管理およびデプロイ・サービスにより、ユーザーはオンプレミスのOracle DatabaseおよびAutonomous Databaseの両方からインデータベース機械学習モデルをデプロイして、分類、回帰、クラスタリング、特徴抽出などの機械学習技術のために使用することができます。Oracle Machine Learning Servicesは、インデータベース機械学習モデルと、ONNX形式でエクスポートされたモデルの両方をサポートしています。ONNX形式のモデルは、Tensorflow、PyTorch、Scikit-learnなどのサード・パーティのパッケージからでも個別に生成可能です。また、Oracle Machine Learning Servicesは、トピックおよびキーワード抽出、センチメント分析、テキスト・サマリー、類似度を含むコグニティブ・テキスト解析をサポートしています。Oracle Machine Learning Servicesは、ストリーム・アプリケーションやリアルタイム・アプリケーション向けのスコアリングのために最適化されています。さらに、他のソリューションでは24時間365日の可用性を確保するためにVMのプロビジョニングが必要となりますが、Oracle Machine Learning ServicesはOracle Autonomous Databaseに付属しているため、実際の予測を生成する際の追加の計算リソースにだけ支払えば済みます。

Oracle Machine Learning for SQL (OML4SQL) は、PL/SQLを使用したモデル構築とSQL問合せを使用したデータ・スコアリングによる予測のための強力なインデータベース機械学習アルゴリズムを提供します。これらのスケーラブル・アルゴリズムは並列化され、メモリも最適化されます。これらの同じインデータベース・アルゴリズムを、PythonおよびRインタフェース、

SQL

Oracle Machine Learning Notebooks、Oracle Data Minerからも使用できます。これによりデータベースが全社的分析プラットフォームとして拡張され、顧客離れの予測、顧客セグメンテーション、不正および異常の検出、クロスセルとアップセルの機会の特定、マーケット・バスケット分析、

予測的メンテナンス、その他多くのデータ由来の問題を処理できるようになります。



Oracle Machine Learning for Python (OML4Py) により、データサイエンティストやPython開発者は、Oracle DatabaseおよびOracle Autonomous Databaseで管理されるデータ上でPython環境を利用できます。Pythonは、データ操作、データ探索、グラフィック、統計関数、

および機械学習アルゴリズム向けに一連のソフトウェア・パッケージを提供しています。OML4Pyは、次の4つの主要領域でPythonの機能を拡張しています。(i) オーバーロードされた馴染みあるPythonの構文と関数（透過的にSQLを生成）を使用して、データベース・データを探索し、準備する。(ii) 自然なPythonインタフェースを使用して、Pythonからインデータベース機械学習アルゴリズムにアクセスする。(iii) REST APIとSQL APIを使用した埋込みPythonの実行により、ユーザー定義のPython関数を容易にデプロイできる。(iv) 自動機械学習（AutoML）。OML4Pyは、透過的アクセスのために、Pandas DataFrameオブジェクトの関数などの厳選されたPython関数をオーバーロードして、インデータベース処理用のSQLを透過的に生成します。OML4Pyは、AutoMLを使用して自動アルゴリズム選択、特徴選択、モデル・チューニングをサポートすることで、生産性、モデルの精度およびパフォーマンスを向上させます。

Oracle Machine Learning for R (OML4R) は、Oracle Database およびOracle Autonomous Databaseとの統合を通じて、オープンソースの統計プログラミング言語/環境であるRで高いスケーラビリティを実現します。



Rは、データ操作、グラフィック、統計関数、および機械学習アルゴリズム向けの充実したソフトウェア・パッケージ・エコシステムを提供しています。OML4Rは、次の3つの主要領域でRの機能を拡張しています。(i) 使い慣れたRの構文と関数（透過的にSQLを生成）を使用して、データベース・データを探索し、準備する。(ii) 自然なRインタフェースを使用して、Rからインデータベース機械学習アルゴリズムにアクセスする。(iii) SQL APIを使用した埋込みR実行により、ユーザー定義のR関数を容易にデプロイできる。



Oracle Data MinerはOracle SQL Developerの拡張機能で、ドラッグ・アンド・ドロップ・ワークフローのユーザー・インタフェースを使用して、機械学習プロセスの多くの手順を自動化します。Oracle SQL Developerは、広く使用

されている無償の統合開発環境で、Oracle DatabaseおよびOracle Autonomous Databaseのデータベース開発と管理を簡素化します。Oracle Data Minerを使用して、分析ワークフローを共有したり、SQLスクリプトを生成してソリューションのデプロイを迅速化させたりすることができます。Oracle Data Minerには、ワークフローのスケジューリングと自動化のためのPL/SQL APIも用意されています。

まとめると、オラクルのコンバインド・データベース戦略を利用し、機械学習、インメモリ、ブロックチェーン、JSON、テキスト、グラフ、空間、パーティション化、シャーディングなどの複数の特定用途データベースの機能を統合することによって、複雑さを緩和して管理作業を減らし、エンド・ツー・エンドの最適なエクスペリエンスを実現できます。Oracle Machine Learningはこの戦略を活用し、最小限のデータ移動、テクノロジーの垣根を越えた利用、総所有コストの削減、全社的なデータを活用したソリューションの最速提供を可能にします。

概要

オラクルのミッションは、"人々が新たな方法でデータを参照し、インサイトを発見し、無限の可能性を解き放つことができるよう支援すること"です。そのため、オラクルは企業に対して、Oracle Machine Learning製品ファミリー内の広範囲にわたる機能を提供しています。それにより、データサイエンティスト、市民データサイエンティスト、データおよびビジネスアナリスト、データエンジニア、IT開発者、アプリケーション開発者、ダッシュボード開発者、経営陣は、オラクルのテクノロジーの価値を最大まで引き出して、インサイトを発見し、予測を行うことができます。

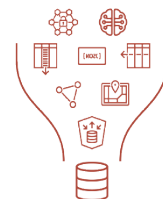
大企業から中小企業まで一様に、機械学習テクノロジーを適用しておもなビジネス上の問題を解決し、トップとは言わないまでも競争力のある地位に到達することの価値を理解しています。オラクルは、企業全体で誰もが機械学習を利用できるようにし、各ビジネスが強力な機械学習ツールをさらに活用できるようにします。

データではなくアルゴリズムを移動する

オラクルは、Oracle Machine Learningによって、オンプレミス、Oracle Cloudを問わずデータがある場所で、Oracleデータベース・インスタンス内のデータを処理する機械学習アルゴリズムを提供することで、“データ側にアルゴリズムを移動”します。この方法により、データの移動が最小化または排除され、データベースのスケーラビリティの活用、データセキュリティの保持、およびデータの取得からモデルのデプロイまでの時間の短縮が可能になります。オラクルは、機械学習アルゴリズムのスケーラブルな並列インデータベース実装を提供し、オープンソースのRおよびPython環境と統合して、エンタープライズのデータサイエンスプロジェクトで必要とされるパフォーマンス、スケーラビリティ、セキュリティ、および自動化を実現します。インデータベース・アルゴリズムでは特に、この業界を代表するOracle Databaseの並列処理、スケーラビリティ、セキュリティの機能が活用されます。

Oracle Database - コンバインド・データベース

異なるタイプのデータや分析機能を使用するために複数のデータベースやツールをプロビジョニングしたり管理したりする必要はありません。これを行うと複雑さが増し、コストがかかります。それよりも、そのような機能はすべてマルチモデルの単一のコンバインド・データベースに存在し、広範な分析機能を集約したものであるべきです。



オラクルのコンバインド・データベースは、データベース分野のあらゆるタイプのデータと処理に対応できます。リレーショナル・データ、ドキュメント・データ、空間データ、テキスト・データ、グラフ・データのすべてを同じデータベース内に効率的に格納して処理できるのです。また、OLTP、リアルタイム分析、機械学習、IoT、ブロックチェーンのワークロードのすべてを同じデータベース内で効率的に実行できます。これらのテクノロジーを利用するために大量の特定用途データベースをデプロイし、管理し、同期させる必要はもうありません。

スマートフォンのように、オラクルのコンバインド・データベースはこれらのデータタイプやワークロードをただサポートするだけでなく、特定用途データベースよりも適切にサポートします。さらに、これらすべての機能を一つのデータベース内にまとめた相乗効果が生まれることで、OLTPシステム内で直接、リアルタイムの機械学習ベースの不正検出を実行するといった新たな可能性が開かれます。また、特定用途データベースで必要とされる個別の言語ではなく宣言的なオープンSQLを使用できるため、開発者にとってこれらのテクノロジーの実装が非常に容易になっています。

—オラクル、Juan Loaiza

SQL、R、Python

調査によると、Python、R、およびSQLが2019年に機械学習プロジェクトで使用されたもっとも人気のある言語でした。異なるビジネス課題やそれらの根底をなすデータでは、異なる分析手法やアルゴリズムを正常に機能させる必要があります。同様に、異なるチームやチーム・メンバーにさまざまなスキルがあり、それらが特定の言語に集中していることもよくあります。Oracle Machine Learningでは、これら3つの人気のあるデータ・サイエンス/機械学習言語がすべてサポートされているため、チームにとって、あるいは目下の問題にとって最適な言語を活用できます。Oracle Machine Learningでは、同じプラットフォーム内でOracleデータ・プロフェッショナルからデータ・サイエンティストへと自然に進化していけます。インデータベース機械学習モデルは、権限、バックアップ、リカバリ、監査など、同じデータ管理機能の多くを備えたファーストクラスのデータベース・オブジェクトです。Oracle Machine Learningにより、企業は機械学習のモデルと結果をデプロイして、Oracle Analytics Cloud、Oracle APEXなどのツールで構築されたダッシュボードやアプリケーション、ならびにカスタムのサーバー・アプリケーションやモバイル・アプリケーションで使用でき、また、使いやすいSQLやRESTインタフェースを活用できます。

RとPythonをサポートすることにより、データ・サイエンティストや他のRおよびPythonユーザーは、高パフォーマンスな計算エンジンとしてOracle Databaseにアクセスして、透過的にSQLを生成するRおよびPythonのオーバーロード関数および構文を使用し、スケーラブルなデータ探索とデータ準備を実行できます。インデータベース・アルゴリズムには、十分に統合されたRおよびPythonインタフェース経由でアクセスできます。これは、SQL APIで提供されるものと同じアルゴリズムです。データベース・スクリプト・リポジトリにユーザー定義のRおよびPython関数をデプロイして、SQLから、そしてOracle Autonomous DatabaseではREST APIからも、それらの関数を呼び出すことができます。ユーザー定義のRおよびPython関数には、Oracle DatabaseおよびOracle Autonomous Database上のサード・パーティ・パッケージをインクルードできます。

自動化による生産性およびモデル品質の向上

機械学習プロセスの重要な部分を占めるのは、反復です。つまり、特徴の設計と選出を行い、「最適な」モデルを見つけるために複数のアルゴリズムを試行してアルゴリズムのハイパーパラメータをチューニングし、ビジネス上の特定の問題についてのモデルを比較して選択するということです。

データ・サイエンティストの生産性を上げ、専門家でないユーザーでも使用できるように、Oracle Machine Learning (OML4PyとノーコードのOracle Machine Learning AutoML UIの両方) には自動化された機械学習であるAutoMLが導入されています。

AutoMLでは、高品質のモデルを獲得するために必要な計算時間全体を短縮できます。また、AutoMLを使用すれば、市民データ・サイエンティストが強力なインデータベース・アルゴリズムにアクセスしやすくなります。アルゴリズムのハイパーパラメータの詳細やアルゴリズム固有のデータ変換について知っている必要はなく、各種スコア・メトリックに従ってモデルを比較し選択するためのコードを記述する必要もないためです。

¹ <https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html>

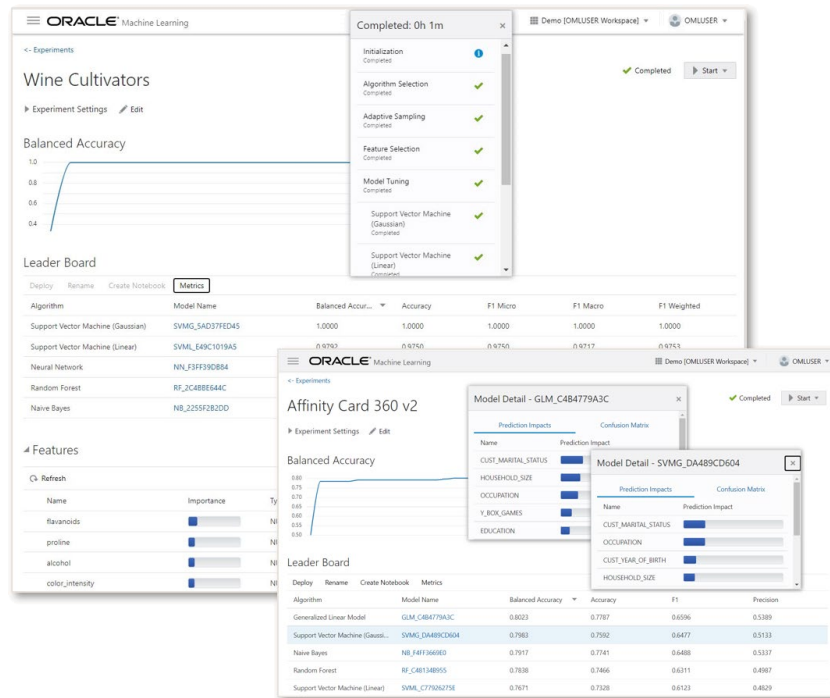


図1: Oracle Machine Learning AutoML UIは、データサイエンティストの生産性向上に役立つ一方、専門家でないユーザーに対しても、インデータベース機械学習モデルを開発してSQL問合せ経由で即座に使用するか、数回のクリックでOracle Machine Learning Servicesにデプロイできる ようにします。

OML4Py AutoMLでは、アルゴリズムの選択、特徴選択、モデルのチューニングという3つのおもな機能を使用したモデリング・パイプラインを明示的に構成できます。Oracle Machine Learning AutoML UIはOML4PyのAutoML機能をベースにして、コードゼロの使いやすいインターフェースを提供しており、このインターフェースでは予測対象のデータ表とターゲット列だけの最小限の情報指定のみが求められます。完全に組み立てられた状態のAutoML実験パイプラインによって、組込みインテリジェンスを活用し、迅速かつ簡単に適切なモデルを生成してデプロイすることが可能です。

Oracle Machine Learningのインデータベース・アルゴリズムは、次のようなアルゴリズム・レベルの自動化もサポートしています。(i) 数値データやカテゴリ・データをアルゴリズム別に変換するためのデータ準備の自動化。(ii) トークンおよびテーマ抽出のためのテキスト・マイニングの統合。(iii) 使いやすいアンサンプル・モデルを生成するパーティション化モデルの自動化。

Oracle Data Minerでは、使いやすいドラッグ・アンド・ドロップ式のグラフィカル・ワークフロー表示を使用して、簡単にデータを集約および変換し、複数のアルゴリズムを使用してモデルを構築および評価することができます。

本番デプロイメント - 重要なステップ

本番デプロイメントは、データサイエンスプロジェクトの弱点とみなされることがあります。2019年、データサイエンスプロジェクトの87%は本番運用に行き着かないことを[VentureBeat.com](https://venturebeat.com)は明らかにしました。また、[Gartner](https://gartner.com)は、「2022年までにビジネスの成果を生み出すような分析インサイトは全体の20%のみ」と指摘しました。機械学習モデルやオープンソースのRおよびPythonスクリプトを本番環境に統合しようとすると、環境や構成コンテナをプロビジョニングし、十分なバックアップ、リカバリ、セキュリティ、スケーラビリティの確保に伴う退屈でエラーの起こり

やすいタスクに対処するという現実に直面します。アルゴリズムとモデルをデータベース内に位置付けることで、これらのタスクがほぼ不要になります。Oracle Machine Learningであれば、デプロイメントはすぐに完了します。インデータベースMLモデルがデータベース内に存在し、SQL問合せから使用できるからです。ユーザー定義のRおよびPython関数をデータベースに保存して、データベース環境により起動され制御されるRおよびPythonエンジンでそれぞれ実行できます。このようにオラクルでは、“配管工事”の提供によって本番デプロイメントが簡素化されているため、エンタープライズ・チームは機械学習ソリューションに重点を置くことができます。

Oracle Machine Learning Servicesは、アプリケーション開発者に対してOracle Autonomous Database上のRESTエンドポイントをサポートすることにより、インデータベース・モデルのデプロイの代わりとなる手段を提供しています。Oracle Machine Learning Servicesでは、インデータベースのOracle Machine Learningモデルとサード・パーティの機械学習アルゴリズムによって生成されたモデルのいずれについても、REST API経由でモデルをデプロイし、モデルを管理できます。Oracle Machine Learning Servicesは、Oracle Database、Oracle Database Cloud Service、Oracle Autonomous Databaseからエクスポートされた分類、回帰、クラスタリング、特徴抽出のモデルを管理しデプロイします。このようなモデルの管理とデプロイのサービスは、Open Neural Networks Exchange (ONNX) 形式でエクスポートされたサード・パーティのモデルにも適用されます。サード・パーティのモデルは、たとえばOracle Cloud Infrastructure Data Science内で、Scikit-learn、TensorFlow、PyTorchなどのパッケージから個別に生成する場合があります。

Oracle Machine Learning Servicesはデータとモデルの監視もサポートしており、データ品質とネイティブのインデータベース・モデル品質のいずれについても、問題があった場合にアラートを受け取ることができます。データを活用する企業では、時間とともにデータに大きな変化があるかどうかや、データから構築された機械学習モデルが期待どおりに動作しているかどうかについて把握する必要があります。

Oracle Machine Learning Servicesの監視機能では、機械学習のライフサイクルとMLOpsのサポートが拡張されます。データの監視により、エンタープライズ・アプリケーションおよびダッシュボードの整合性を確保するための組織のデータ品質標準を維持できます。たとえば、アップストリーム・データの収集により、新たなセンサーがデータを異なる単位でレポートしていることや、顧客データを収集するアプリケーションにバグがあるために類似したフィールドでデータが取り違えられて誤ったレポートが生成されていることを把握できるでしょう。モデルの監視により、モデルがソリューション成功基準を満たし続けているか、あるいはモデルの再構築や再評価が必要であるかを確認できます。たとえば、ある企業が収益性を維持するには、キャンペーンに応答する可能性が高い人を予測するモデルの精度が75 %以上である必要があります。精度がこのレベルを下回ると、事業部門マネージャーの四半期目標は達成できず、検出された、あるいは早期に修正された問題がCFOのダッシュボードに表示されることになります。

また、Oracle Machine Learning Servicesはコグニティブ・テキスト解析をサポートしており、トピックおよびキーワードの抽出、テキスト・サマリーと類似度の実行については英語、スペイン語、フランス語、イタリア語で使用でき、センチメント分析について英語で使用できます。

企業の成功事例

企業では、収集するデータ量が増加し、企業データを外部データソースで補完するため、機械学習用に大規模にデータを統合および準備でき、ビジネス課題の定義からソリューションのデプロイメントまでの時間を合理化できることが、ビジネス目標を達成するために必要です。

Oracle Machine Learningのお客様は、次のような素晴らしい結果を出しています。

- 英国の公共医療提供業者のシステム、UK National Health Service (NHS) は、処方箋、薬品、医療控除、医師のリレーションシップについての何十億ものデータ・ポイントや、組織全体のコールセンター・サービスを集めて分析し、それを使用して新たな効率や効果的な治療の可能性を明らかにすることで、より適切な結果を出すことができました。NHSは、非常に迅速にROIを実現しました。最初の2年間で5億8100万英国ポンド（7億1700万米国ドル）、そして直近では5年以内で15億英国ポンド（約20億米国ドル）以上を達成し、まだ増え続けています。NHS Business Services Authorityの情報サービス担当責任者であるNina Monckton氏は次のように述べています。「オラクルを選択したのは、数十億行に及ぶ大量のデータに対処でき、データ量の増加に合わせてスケーリングできるからです。また、Oracleソリューションでは問合せ実行のためにITチームのサポートを受ける必要はなく、データ・アナリスト・チームが自己解決できるからです。オラクルのエンジニアド・システムによってデプロイが迅速化してリスクが減少しました。オラクルと一緒に仕事できたことは非常に良い経験となりました。オラクルのチームは驚くほど対応が早く、また当社が可能な限り早く運用を開始できるように多くのエキスパートが支援してくれました。1社でソリューション全体を提供していることも、当社にとって非常にやりやすい点です。ヘルプが必要なときの問合せ先がわかっているからです」
- スペインの大手金融サービス企業のCaixaBankは、銀行の支店、ATM、インターネットおよびモバイル・バンキングのデータを統合して顧客の状況を完全に把握し、パーソナライズされたバンキング・ソリューションを提供しています。CaixaBankは、より効率的に顧客にアプローチするためのメッセージングを向上させることで、顧客ロイヤルティとビジネス上の競争力を獲得し、顧客に新しい銀行サービスや製品を案内して売上を増加させています。
- Certegyは、4000以上の顧客と米国上位50社の小売業者のうち23社に対して、20億米ドル超の小切手決済サービスを提供する大手企業です。同社も、データ・サイエンティストによる統計モデリング、リスク評価、不正検出のために8億5000万件のレコードを保管するリポジトリとして、Oracle Autonomous Data Warehouseを選択しました。CertegyはOracle Analytics、Oracle Machine Learning、Oracle APEXを使用して、Microsoft PowerBIやPythonよりも優れた統合、データ品質、パフォーマンスを実現しています。Oracle Machine Learningを使用することで、Certegyはより精度の高いリアルタイムのリスク・スコアを算出しています。不正件数は推定で10%減少し、顧客サービスも向上しました。
- 世界最大のチケット市場のStubHubは、Oracle Machine Learningのインデータベース・モデルおよび統合されたR機能を使用して、リアルタイムの不正検出モデルをデータベースで実行しています。このソリューションを使用して、StubHubはオンライン不正行為を90%削減しました。これにより、コストが節約されるだけでなく、顧客体験が大幅に向上します。もちろん、詐欺師はいつシャットダウンされるかに注意を払っており、詐欺行為の方法を変えています。そのため、すばやく反応する必要があります。さもないと、不正行為が増加します。Oracle Machine Learningを使用することで、StubHubはきわめて迅速に対応できます。予測モデルの更新は、アプリケーションやデータベースの実行中に行われます。管理者が朝に問題に気付いた場合、1時間で解決でき、停止時間を発生させることなく新しいモデルをすぐにデプロイできます。

上記の事例の詳細とその他の顧客事例については、[OML Customers](#)を参照してください。

機械学習とは何ですか。

機械学習は、アルゴリズムと統計モデルを使用して、大量になりうるデータを自動的に処理して隠れていたパターンを見つけ、新しいインサイトを発見し、次のようなデータ由来の問題を予測します。

- 顧客の行動を予測し、クロスセルとアップセルの機会を識別する
- 顧客離れ、従業員や学生の在籍継続率を予測する
- 異常を検出し、税、医療、または経費の潜在的不正を検出する
- 隠れた顧客セグメントを把握し、顧客センチメントを理解する
- 機器のメンテナンスが必要となる時期を提案する
- 成果を生み出すおもな要因を識別する
- 顧客の満足度と製品品質を向上させる

機械学習は予測分析やデータ・マイニングとも呼ばれ、何十年にもわたってかなりの価値を生み出してきました。今日、機械学習ソリューションはさらに浸透して、世界中の企業において実装および導入されています。分析テクノロジーとユーザーへの普及の発展や拡大に伴い、機械学習のユースケースや"インテリジェントな"アプリケーションが毎日のように登場しています。

データではなくアルゴリズムを移動する

データは大きいですが、アルゴリズムは小さいです。したがって、データをアルゴリズムに移動するよりも、アルゴリズムをデータに移動するほうが理にかなっています。オラクルでは、Thinking Machines Corporationの機械学習テクノロジーと開発チームを獲得した1999年に、これを理にかなったものにするのに必要なアルゴリズムとビッグ・データの課題を認識しました。オラクルは、Oracle Databaseの核となる強み、特に、集計、メモリ管理、並列処理、スケーラビリティ、およびデータベース・アーキテクチャ・フレームワークを利用し、データベース・カーネル内のあらゆる計算機構を用いてモデルを構築しデータをスコアリングしました。データ、機械学習モデル、または予測やインサイトのいずれであれ、データベースを離れる必要はありません。

今日では、Oracle Machine Learningは、30を超えるインデータベース・アルゴリズム、統合テキスト・マイニング、自動データ準備、パーティション化モデル、および自動機械学習（AutoML）をはじめとする多数の革新的な機能をサポートしています。

Oracle Machine Learningとは

Oracle Machine Learningは、広範囲の強力なデータ探索とデータ準備の機能、さらにOracle DatabaseとOracle Autonomous Databaseの内部に実装される機械学習アルゴリズムを提供することによって、データ由来の問題に対応することができます。インデータベース・アルゴリズムにはSQL経由、またはPython APIとR API経由で直接アクセス可能です。

Oracle Machine Learningのアルゴリズムでは、基盤となるSQL機能を利用して、元々のスター・スキーマ表現でデータを分析できます。これには、構造化された標準の表やビュー、トランザクション・データと集約、キャラクタ・ラージ・オブジェクト（CLOB）データ型に含まれる非構造化データ（Oracle Textを使用してトークンまたはテーマを抽出）などがあります。オラクルのコンバインド・データベースは、機械学習と空間、グラフ分析が組み合わされた単一の環境にもなります。

Oracle Machine Learningのインデータベース・アルゴリズムでは、モデル構築とスコアリングの両方でデータベースの並列処理を利用し、セキュリティ・スキームと権限スキームに従い、監査追跡データベースの機能を固守します。Exadata（オンプレミス、Public Cloud、またはCloud at Customer）およびAutonomous Databaseでは、Oracle Machine Learningインデータベース・モデルを使用したスコアリングが、オラクルのスマート・スキャン・テクノロジーを使用してストレージ層で実行されます。他の機械学習プラットフォームが抱える課題の1つが、データをメモリ内に収める必要があることです。Oracle Machine Learningでは、データは必要に応じてメモリに徐々に展開されます。さらに、モデルがキャッシュされ、スコアリングでの使用時にクエリー間で共有可能です。Oracle Machine Learningはディスクを認識した構造を利用します。これは、マルチユーザー環境での効率的な割当てのために、データベース・メモリ・マネージャを活用するものです。パーティション化モデルの構築やスコアリングの際に、すべてのコンポーネント・モデルを一度にロードする必要はありません。



図2 : Oracle Machine Learningのアルゴリズムと分析機能では、重要なビジネスのユースケースを幅広くサポートしています。

Oracle Machine Learningは、データ分析と機械学習のさまざまな方法論の組み合わせをサポートしています。たとえば、トランザクション・データ、人口統計データ、顧客サービス・データ、および顧客のコメントを組み合わせ、あらゆる角度から見た顧客像を導き出すことができます。また、顧客のクラスタリングを実行してそれらの顧客を顧客セグメントに事前に割り当て、その後セグメントごとに、別々の分類モデル、回帰モデル、または異常検出モデルを構築して、精度を上げることができます。

SQLとオープンソースのRおよびPythonのサポート

SQLは、40年以上にわたってデータ管理分野の標準言語です。しかし、データ分析と機械学習の分野では、R、Python、Scala、Java、SAS、MATLABその他のさまざまな言語が競合しています。これらは長年好まれてきましたが、近年は、オープンソースのPythonやRが優勢になっており、ほとんどのデータ・サイエンス教育プログラムで指導されています。各種ブログやデータ・サイエンス・コミュニティの人気調査によると、Python、R、SQLが上位の言語となっており、Oracle Machine LearningではOML4Py、OML4R、OML4SQLとしてこれら3つの言語をサポートしています。

Oracle Machine Learningによるインデータベース処理

Oracle Machine Learningは、Oracle DatabaseとOracle Autonomous Databaseをデータ・サイエンスおよび機械学習のプラットフォームとして、強力なインデータベース分析によって拡張します。この分析では、アルゴリズムがデータベースのデータ上、インデータベースで直接実行されます。Oracle Machine Learningのインデータベース・モデルはファーストクラスのデータベース・スキーマ・オブジェクトで、PL/SQLプロシージャ、またはOML4R、OML4Pyの対応する関数を呼び出すことによって構築されます。インデータベース・モデルを使用して、OML4SQL、OML4R、またはOML4Pyの関数およびデータベース・ビュー経由で、データのスコアリングやモデル詳細（モデルにより明らかとなったパターンやインサイト）の分析を実行できます。

モデルを構築する際、Oracle Machine Learningは、標準的でスケーラブルなデータベース・テクノロジー（並列実行、ビットマップ列索引、ストレージレベルのパーティション化、クエリ最適化、集計手法など）を利用します。また、並列インフラストラクチャ、IEEE浮動小数点、Intel Math Kernel Library（MKL）、非構造化テキスト・データのサポート、ビンニング用の自動データ準備、one-hotエンコーディング、欠損値の処理など、カスタムのテクノロジーも利用します。

インデータベース・アルゴリズムのSQL関数としての有用性は、バッチ環境、オンライン・トランザクション処理（OLTP）環境を問わず、機械学習モデルのスコアリングの際によりはっきり示されます。高パフォーマンスのスコアリングが可能で、数秒で数百万レコードを処理できます。SQL条件モデルとOMLモデルは、Oracle Exadataストレージ・セルのスマート・スキャン・テクノロジーを使用して、処理のためにストレージ・レイヤーにプッシュダウンされます。どちらの場合も、条件を満たすレコードのみがディスクから引き出され、データベース内部でさらに処理されます。

たとえば、以下のような単純なSQL問合せで、解約する可能性のある米国の顧客を検索するとします。

```
SELECT cust_id
FROM customers
WHERE region='US' AND prediction_probability(churn_model,'Y' using *) > 0.8;
```

このクエリーでは、*churn_model*と呼ばれるモデルに基づく解約確率が80%を超える米国の顧客を見つけます。Oracle Machine Learningでは、オラクル専有のExadataスマート・スキャン・テクノロジーを使用して、このスコアリング計算が条件関数を介してストレージ層にプッシュされます。このスコアリングのパフォーマンスは、スマート・スキャンを使用しないインデータベース・スコアリングの2～5倍になります。

モデルの構築でもデータのスコアリングでも、インデータベース機械学習によって、データ・サイエンス・プロジェクトの期間が数日や数週から、数秒、数分、数時間の単位まで短縮されるようになります。

自動化による生産性の促進

データ・サイエンス・プロセスは、きわめて反復的になる可能性があります。CRoss Industry Standard Process for Data Mining（[CRISP-DM](#)）などの一般的な手法には、特にデータ準備、モデル構築、およびモデル評価に対して、反復的なループが組み込まれています。幸い、このプロセスの大部分は自動化が可能です。このような自動化によって、機械学習のより時間のかかる反復的な側面が処理されてデータ・サイエンティストの生産性が向上するだけでなく、専門家でないユーザーが、モデルのチューニングに関するハイパーパラメータなどのアルゴリズムの詳細を理解していなくても機械学習を利用できるようになります。それでも、自動化によって機械学習の広範なスキルの必要性がなくなるわけではありません。特に、問題定義、データ準備、ソリューション評価の分野ではそう言えます。

Oracle Machine Learningで提供される自動化には、自動データ準備（ADP）、自動テキスト処理、パーティション化モデル、自動機械学習（AutoML）による完全自動MLパイプラインなどがあります。AutoMLでは、自動アルゴリズムと特徴の選択、モデル選択、およびモデルのチューニングが提供されます。これらの機能すべてにより、機械学習を使用してビジネス上の問題からソリューションを導き出すために必要となる作業量や知識が少なく済みます。

オープンソースのRおよびPythonとOracle Databaseとの統合

Oracle Machine Learning for R（OML4R）とOracle Machine Learning for Python（OML4Py）は、Oracle DatabaseおよびOracle Autonomous Databaseとの統合を通じて、RおよびPythonの言語/環境で高いスケーラビリティを実現します。充実したエコシステムを利用できること、強力で拡張性に優れた広範なグラフィカル・パッケージ、統計パッケージ、分析パッケージが提供されていることなどが、これらの言語の特長です。ただし、RおよびPythonを使用していると、多くの場合は、メモリ制約やシングル・スレッド処理の問題、RおよびPythonのモデル・ロジックをエンタープライズのアプリケーションまたはダッシュボードにデプロイまたは変換する際の困難などに直面します。

OML4Pyは、PythonをOracle Databaseに統合し、選択したPython関数を同等のSQLおよびインデータベース・アルゴリズムへとオーバーロードします。これにより、SQLを使用せずにデータベース内のデータを操作できます。オーバーロードされた関数によって透過レイヤーが形成され、リクエストされたPython機能がインデータベース処理用の同等のSQLに透過的に変換されます。Pythonプログラマーは、透過レイヤーを通じて、データベース表およびビューを参照するPandas DataFrameプロキシ・オブジェクトを作成し、データベース内に存在するデータにアクセスして分析と操作を行うことができます。データベースでは、対応するSQLコードが自動的に最適化されて問合せの効率が向上し、列索引、表のパーティション化、およびデータベースの並列処理が活用されます。

埋込みPython実行により、システムによって起動および制御される1つまたは複数のPythonエンジンでユーザー定義Python関数を実行できます。ユーザー定義関数の開発や改良も可能で、その際にはカスタムのインストール・パッケージを使用することもできます。その後、ユーザー定義関数をデータベース・スクリプト・リポジトリにデプロイできます。ユーザー定義Python関数は、非並列、データ並列、またはタスク並列で呼び出すことが可能です。この際、PythonとSQLを使用しますが、Autonomous Databaseの場合はREST APIも使用できます。

OML4Pyは、Autonomous Database上に事前インストールされており、Oracle Machine Learning Notebooksを通じて利用できます。Oracle Autonomous DatabaseおよびOracle Databaseインスタンスへの接続をサポートするスタンドアロンクライアントを使用して、PyCharm、Jupyterインタフェースなどのサード・パーティ・ツールを使用することもできます。

The screenshot displays the Oracle Machine Learning Notebooks interface with a workflow for building and evaluating a neural network model. The workflow consists of several steps:

- Classification Modeling to Predict Target Customers using Neural Network:** This notebook introduces the goal of predicting high affinity card responders. It includes a diagram showing a neural network structure with input nodes (circles) and output nodes (triangles).
- Build a Neural Network regression model for predicting AFFINITY CARD:** This step involves creating training and test data (60/40 split) using Python code:


```
python
TRAIN, TEST = DEMO_DF.split(ratio = (0.6,0.4))
TRAIN_X = TRAIN.drop('AFFINITY_CARD')
TRAIN_Y = TRAIN['AFFINITY_CARD']
TEST_X = TEST
TEST_Y = TEST['AFFINITY_CARD']
```
- Build a Neural Network Model using default settings:** This step shows the model training process with Python code:


```
python
try:
    oml.drop(model = 'NN_CLAS_MODEL')
except:
    pass

setting = dict()
nn_mod = oml.nn('classification', **setting)
nn_mod.fit(TRAIN_X, TRAIN_Y, case_id = 'CUST_ID', model_name = 'NN_CLAS_MODEL')
```

 The model settings are displayed as follows:

setting name	setting value
ALGO_NAME	ALGO_NEURAL_NETWORK
CLAS_WEIGHTS_BALANCED	OFF
LBFIS_GRADIENT_TOLERANCE	.000000001
LBFIS_HISTORY_DEPTH	20
- Show model accuracy:** This step displays the model's accuracy score:


```
python
nn_mod.score(TEST_X, TEST_Y)
0.849014
```
- Display confusion matrix, lift chart and ROC curve:** This step provides a comprehensive evaluation of the model. It includes a confusion matrix, a lift chart, and an ROC curve.

	PREDICTED 0	PREDICTED 1
ACTUAL 0	True Negative: 1214 (88.39%)	False Positive: 142 (8.0%)
ACTUAL 1	False Negative: 126 (7.1%)	True Positive: 293 (16.51%)
Accuracy: 84.9014%	AUC: 0.8847	F1Score: 0.6862

 The evaluation metrics for the Neural Network model are:
 - Algorithm: Neural Network
 - Algorithm setting: 3-Layer - 8,8,2
 - TN: 1214
 - TP: 293
 - FP: 142
 - FN: 126
 - TPR: 0.6993
 - FPR: 0.1047
 - TNR: 0.8953
 - FNR: 0.3007
 - Precision: 0.6736
 - Accuracy: 0.849
 - NPV: 0.906
 - DetectionRate: 0.906
 - BalancedAccuracy: 0.7973
 - AUC: 0.8847
 - F1Score: 0.6862
 - MathewsCorrCoef: 0.587

図3 : Oracle Machine Learning Notebooksは、Python、R、SQL、conda、およびマークダウン・インタプリタをサポートします。このノートブックで、Python開発者はニューラル・ネットワーク・アルゴリズムを使用して分類を実行し、matplotlibライブラリを使用した視覚化機能によりモデルを評価できます。

OML4Rは、RをOracle Databaseに統合します。OML4Pyと同様に、ユーザーはR透過性レイヤーを使用して、スケーラブルなインデータベース・データ処理を実行し、インデータベース・アルゴリズムへのRインタフェースを使用してモデルを構築し、データをスコアリングすることができます。Oracle DatabaseでのOML4Rの埋込みR実行により、サード・パーティのRパッケージを使用するユーザー定義関数を作成することによって、ネイティブ機能を拡張することができます。このRパッケージは、ユーザーまたは管理者がOracle Databaseとともにデータベース・サーバー・マシンにインストールします。これらのユーザー定義のR関数の呼び出しには、R APIとSQL APIの両方を使用できます。RStudioなどの好みのR IDEを使用してOracle Databaseインスタンスに接続し、OML4Rを使用することもできます。

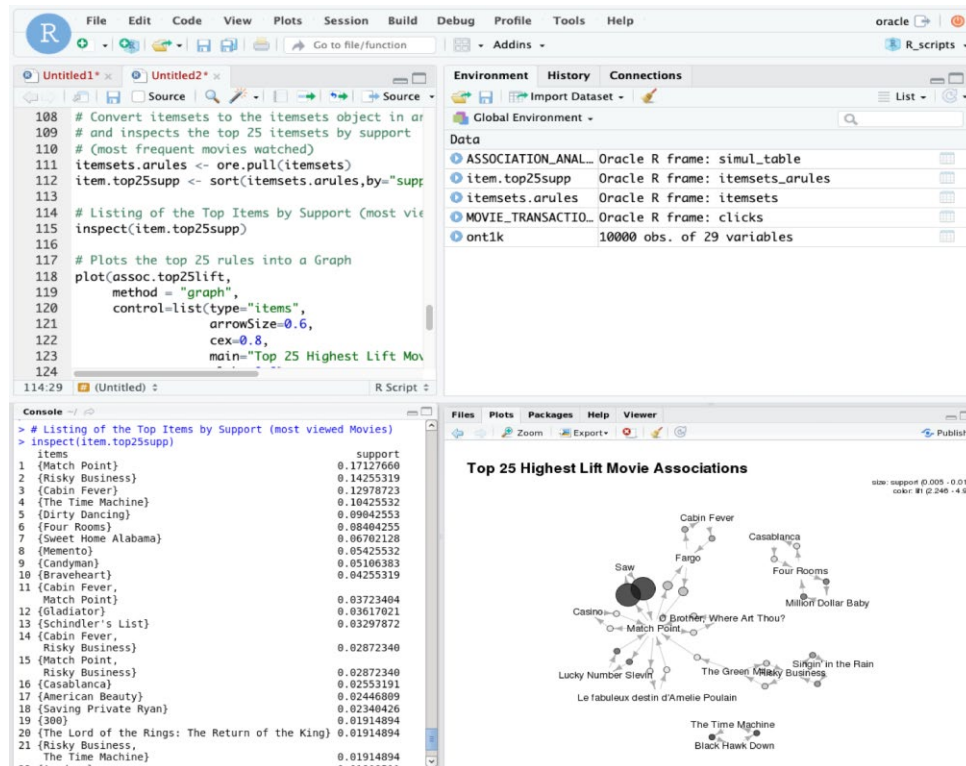


図4：Oracle Machine Learning for Rでは、インデータベース・アルゴリズム（ここではApriori相関ルール）とローカルのグラフ描画機能を使用して結果を視覚化します。これらはすべてRStudio Server内で実行されます。

Oracle Autonomous Databaseでは、Oracle Machine Learning Notebooksを通じてOML4Rにアクセスします。ユーザーがR関数を定義し、その関数にネイティブのR機能や、conda環境にインストールされているサード・パーティ・パッケージの機能をインクルードし、Embedded R実行を使用して、R、SQL、およびREST APIから関数を呼び出すことができます。

Oracle Machine Learning AutoMLユーザー・インタフェース

Oracle Machine Learning AutoMLユーザー・インタフェースは、自動機械学習をサポートするコードゼロのユーザー・インタフェースです。Oracle Machine Learning AutoML UIによって、データ・サイエンティストが生産性を高めることができ、専門家でないユーザーも分類や回帰のために強力なインデータベース・アルゴリズムにアクセスできるようになります。

Oracle Machine Learning AutoML UIは最小限の入力情報によってモデル構築を自動化します。“実験”としてデータとターゲットを指定するだけで、後の処理は自動的に実行されます。より詳細に制御したい場合は、試行するアルゴリズム・セット、ランキング対象のモデルの数などを制限するための詳細設定を使用します。

図1ですでに示したように、実験の実行によってモデルとそのパフォーマンスがランキング表示されます。ユーザーはOracle Machine Learning Servicesにデプロイするモデルを選択したり、あるいはSQLやPythonから直接モデルを使用したりできます。選択したモデルを生成するOML4Pyコードが含まれたノートブックを生成することもできます。

Oracle Machine Learning Services

Oracle Machine Learning Servicesは、モデルのデプロイ、モデル管理、データとモデルの監視ができるように、Oracle Machine Learningの機能を拡張したものです。インデータベース・モデルと、サード・パーティのアルゴリズムを使用して構築されOpen Neural Networks Exchange (ONNX) 形式でエクスポートされたモデルの両方に対応したREST APIを使用します。Oracle Machine Learning ServicesはOracle Autonomous Databaseで使用可能なRESTエンドポイントを提供します。

分類、回帰、クラスタリング、特徴抽出をサポートするインデータベース・モデルをOracle Machine Learning Servicesにデプロイできます。ONNX形式でエクスポート可能であれば、scikit-learn、TensorFlow、PyTorchなどのパッケージによるモデルを含め、分類、回帰、クラスタリングをサポートするサード・パーティのモデルについてはOracle Machine Learning Servicesにデプロイできます。

Oracle Machine Learning Servicesはデータ・ドリフトの検出によるデータ監視をサポートしています。この検出方法では、時間とともにデータの統計プロパティが大幅に変化した場合にフラグが立てられます。Oracle Machine Learningのデータ監視によって、時間の経過によるデータの変化を比較して追跡できます。これは、ビジネス・タスクが、ベースライン・データセットと呼ばれる十分に理解されたデータと、そのベースラインと比較される、定期に取得した後続のデータセットに従って進められるという考え方に基づいています。基本となるデータの特性の変化を迅速かつ確実に特定することによって、その変化が企業に悪影響を及ぼす前に、データ・スチュワードが是正措置をとることができます。

モデルの監視には、バランスの取れた精度 (*Balanced Accuracy*)、平均二乗誤差 (*MSE*) などのユーザー指定のモデル・メトリックが大幅に変化した時期の特定も含まれます。モデルの監視では、予測値の分布が初期値から逸脱し過ぎている状況も検出できます。この状況は、コンセプト・ドリフト (*Concept Drift*) とも呼ばれます。このような逸脱は、モデルの再構築が必要であることや、逸脱の根本原因調査を検討する必要があることを示している場合があります。たとえば、前年は魅力的な見込み顧客だと思われたローン申請者が、銀行の戦略や将来的なマクロ経済状況の見通しに変化したことで、魅力的とは思われなくなるようなケースがあります。同様に、製品カテゴリへの顧客の関心も時間とともに変化して、以前の期間の関心が反映されたデータに基づいて小売予測モデルが構築されていた場合には、供給の問題につながるようなケースがあります。データ監視と同様に、ユーザーはモデル精度メトリックを評価するために、新しいテスト用データセットを定期的に提供します。このメトリックが指定したしきい値を下回るとそのテスト・データにフラグが立てられます。ユーザーはそれを基に、モデルを再構築するか、もっと精度の高いメトリックを使用する別のモデルを選択するかを検討できます。異なるアルゴリズムや設定などを使用する複数のモデルの構築を含むシナリオを作成してもよいでしょう。それらのモデルはその精度メトリックに基づいて競います。このような“チャンピオン・スコア”により、特定の時点でアプリケーションに使用されるべきモデルを導き出すことができます。

さらに、Oracle Machine Learning Servicesは、トピックの探索、キーワードの抽出、要約、センチメント分析、類似度などの認知テキストの機能をサポートしています。言語として英語、スペイン語、フランス語、イタリア語がサポートされ（事前に組み込まれたWikipediaベースの明示的セマンティック分析モデルに基づく）、センチメント分析については英語がサポートされています。

また、ONNX形式のモデルのデプロイ機能を使用すれば、イメージやテンソルを使用したイメージ分類のために、サード・パーティの認知イメージ・モデルをインポートすることもできます。

Oracle Machine Learning ServicesのREST API機能一覧については、図5を参照してください。

管理	レポートリ	デプロイ	非同期ジョブ	認知テキスト
(POST) <ul style="list-style-type: none"> ADB ユーザーおよびパスワードを使用したトークン 	(POST) <ul style="list-style-type: none"> モデルの保存 モデルの名前空間の更新 	(POST) <ul style="list-style-type: none"> モデル・エンドポイントの作成 エンドポイントを使用したモデルのスコアリング 	(POST) <ul style="list-style-type: none"> ジョブの送信 ジョブの更新 ジョブ・アクションの実行 	(POST) <ul style="list-style-type: none"> 最も関連性の高いトピックの取得 最も関連性の高いキーワードの取得 サマリーの取得 センチメントの取得 セマンティック上の類似点の取得 数値的特徴
一般				
(GET) <ul style="list-style-type: none"> すべてのバージョン用のメタデータ：バージョン1メタデータ Open API 仕様 	(GET) <ul style="list-style-type: none"> モデルのリスト モデル情報 モデルのメタデータ モデルのコンテンツ 	(GET) <ul style="list-style-type: none"> エンドポイント エンドポイントの詳細 エンドポイントのOpen API 仕様 	(GET) <ul style="list-style-type: none"> ジョブのリスト ジョブの詳細 	(GET) <ul style="list-style-type: none"> エンドポイントの取得
	(DELETE) <ul style="list-style-type: none"> モデルのリスト 	(DELETE) <ul style="list-style-type: none"> エンドポイント 	(DELETE) <ul style="list-style-type: none"> ジョブの削除 	

図5 : Oracle Machine Learning Servicesで提供されるREST API機能

Oracle Data Miner—SQL Developerの拡張機能

Oracle SQL Developerデスクトップ・アプリケーションの拡張機能であるOracle Data Minerは、ドラッグ・アンド・ドロップ形式のユーザー・インタフェースで、ワークフロー・パラダイムを使用してデータベース内部のデータを直接操作するのを好むデータ・サイエンティストおよび市民データ・サイエンティストを支援します。Oracle Data Minerのワークフローは、機械学習方法を記録して文書化します。ユーザーは、ワークフローを保存して他のユーザーと共有し、機械学習方法を自動化および公開することができます。Oracle Data Minerは厳選されたインデータベース機械学習アルゴリズムをサポートしています。

すぐにデータを視覚化、探索、準備、変換し、機械学習モデルを構築および評価し、モデル詳細ビューとモデル評価ビューを使用できます。その後、Oracle Machine Learningのインデータベース・モデルを新しいデータに適用したり、SQLおよびPL/SQLスクリプトを生成して分析ワークフローをデプロイしたりします。

これらのスクリプトは、同じOracle Databaseまたは別のOracle Databaseにデプロイするために開発者に渡すことができます。アプリケーション開発者は、Oracle Data MinerのPL/SQLワークフロー-APIを使用してワークフローをプログラムによって実行でき、また、適切なSQLおよびPL/SQLスクリプトを実行するだけで、予測方法をアプリケーションに容易に組み込み、企業全体で広範に使用することができます。

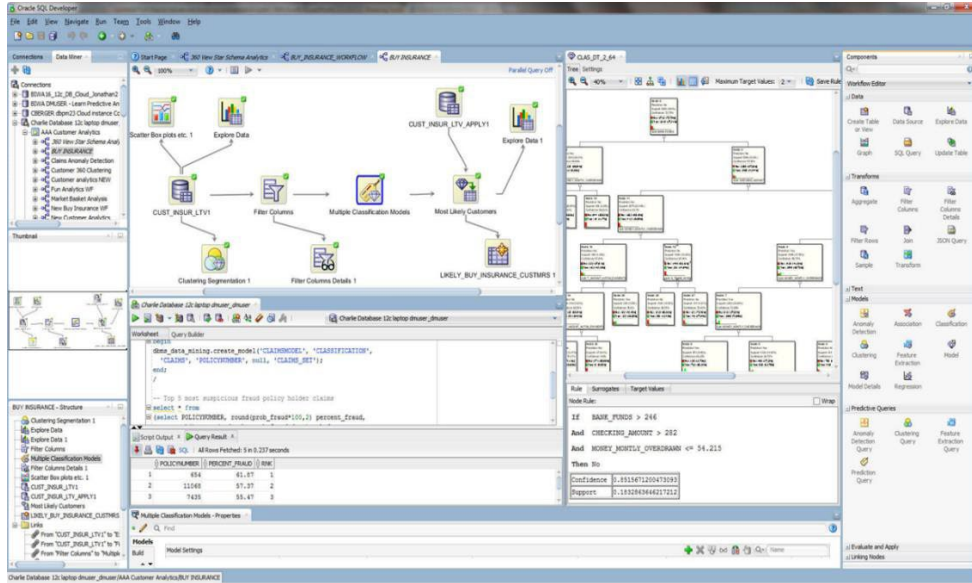


図6：SQL Developerの拡張機能であるOracle Data Minerには、ドラッグ・アンド・ドロップ式のワークフロー・ユーザー・インターフェイスが装備されており、データを探索し、予測モデルを構築/評価/適用し、分析手法をSQLスクリプトとPL/SQLスクリプトとしてデプロイできます。

データ・レイクのOracle Machine Learning

ビッグ・データは通常、分析というよりは収集が容易な構造化データおよび非構造化データを表し、収集後はデータ・リザーバ、データ・レイク、またはレイクハウスに保存されます。ただし、データ由来の問題を解決する目的において、多くの場合、おもな関心事は生データではなく、Oracle Database内でクレーションを行った他のデータと結合できる集計、パーセンテージ、および“イベント”によって提供されるデータのサマリーにのみ存在します。多くの場合、結果のデータは、機械学習を使用してビジネス上の問題を解決するために使用できます。つまり、データ・レイクを、ドメイン・エキスパートの知識を取得する新しいエンジニアド機能に“凝集”して、より適切なモデルを生成するという考えです。

Oracle Databaseの外部にデータ環境があるため、新たなデータ管理やデータ分析の課題がDBAにもたらされます。Oracle Autonomous Databaseの新しいデータ・レイク用機能はこの問題に対処するため、Oracle Object Storage内のデータをSQL処理できるように拡張されています。データ・レイクのクエリーにOracle SQLを使用できるため、Oracle Autonomous Databaseのクエリーで使用できるものであれば、任意のツールやアプリケーションでデータ・レイクのクエリーも実行できるようになります。それはつまり、Oracle Machine Learningの拡張機能を両方のソースのデータにまで適用できるようになるということです。

多くの場合、データ・レイクおよびObject Storageに保存されているデータは大量で、スパース表現（トランザクション形式など）になっています。データの大半がセンサー、IoT、“ツイート”などの大量ソースから取得される可能性があることを考えると、データ・アナリストは、以前よりも容易に、未知の価値を持つかもしれないデータを含む新たなビッグ・データ・ソースを利用し、それを要約して、データベースまたはデータウェアハウス内に管理された既知の価値を持つデータと組み合わせられるようになります。集計、最大値、最小値、しきい値、平均、スライディングSQLウィンドウの平均などのさまざまなレベルでこのデータを集計できます。

この場合、Oracle Machine Learningでは、データ・レイクで直接削減することによって“ビッグ・データ”をフィルタし、他のデータベース・データと結合してから、インデータベースの機械学習を実行できます。

Oracle Machine Learning—AIアプリケーションとダッシュボードを最短期間で提供

データ分析を簡素化するオラクルの戦略により、機械学習によって強化されたインテリジェントなアプリケーションの開発、改良、およびデプロイが容易になっています。データ、ユーザー・アクセス、セキュリティと暗号化、スケーラビリティ、アプリケーション開発環境、および機械学習が、データベース・エコシステム内に揃っています。Oracle Machine Learningでは、Oracle DatabaseへのSQL、REST、またはその他のAPIを使用して、予測やリアルタイムの実用的なインサイトを、エンタープライズのアプリケーション、ダッシュボード、またはツールに容易に追加できます。

オラクルでは、長年にわたりインテリジェントなアプリケーションの開発を継続的に行っています。オラクルでは、オンプレミスとクラウドで使用する次のような次世代の予測アプリケーションを提供します。

- Oracle Human Capital Managementワークフォース予測
- Oracle Content and Experience
- Oracle Configure, Price, Quote
- Oracle Depot Repair
- Oracle Spend Classification
- Oracle Retail Science Platform Cloud Service
- Oracle Retail XBRi Loss Prevention
- Oracle Retail Customer Segmentation
- Oracle Adaptive Intelligence Foundation for Anti-Money Laundering

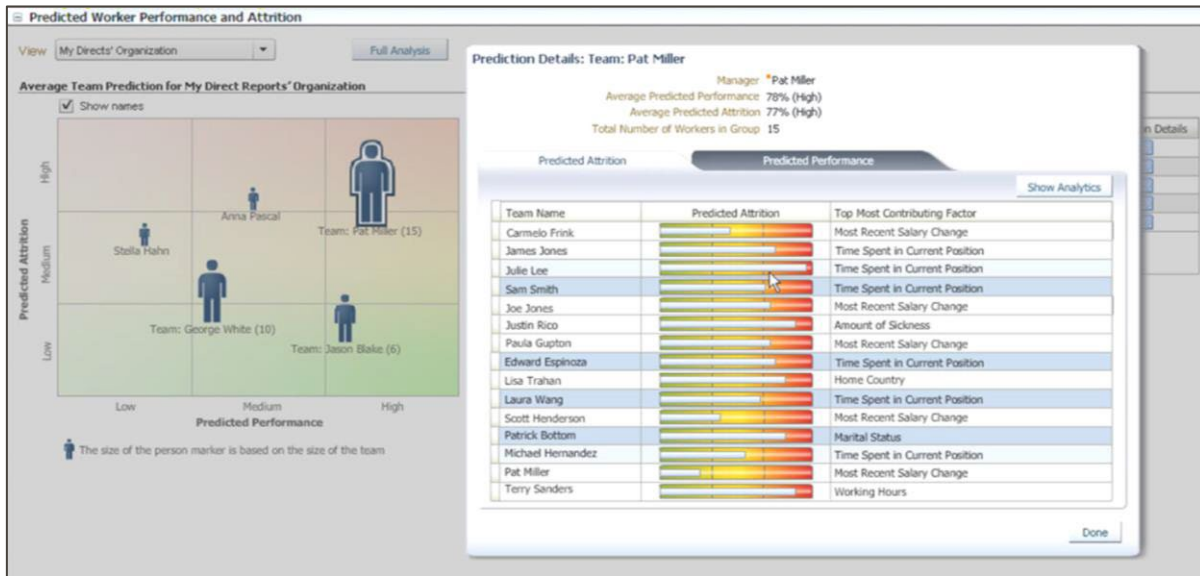


図7： Oracle HCMワークフォース予測アプリケーションでは、事前に組み込まれたOracle Machine Learningにより、従業員の減少と従業員の実績の予測分析、および“What-if”分析を実行できます。

Oracle Analytics CloudおよびOracle Analytics Serverとの統合によって、インデータベース機械学習モデルをシームレスに操作して、データの分析時やダッシュボードの生成時に予測やインサイトを追加できます。

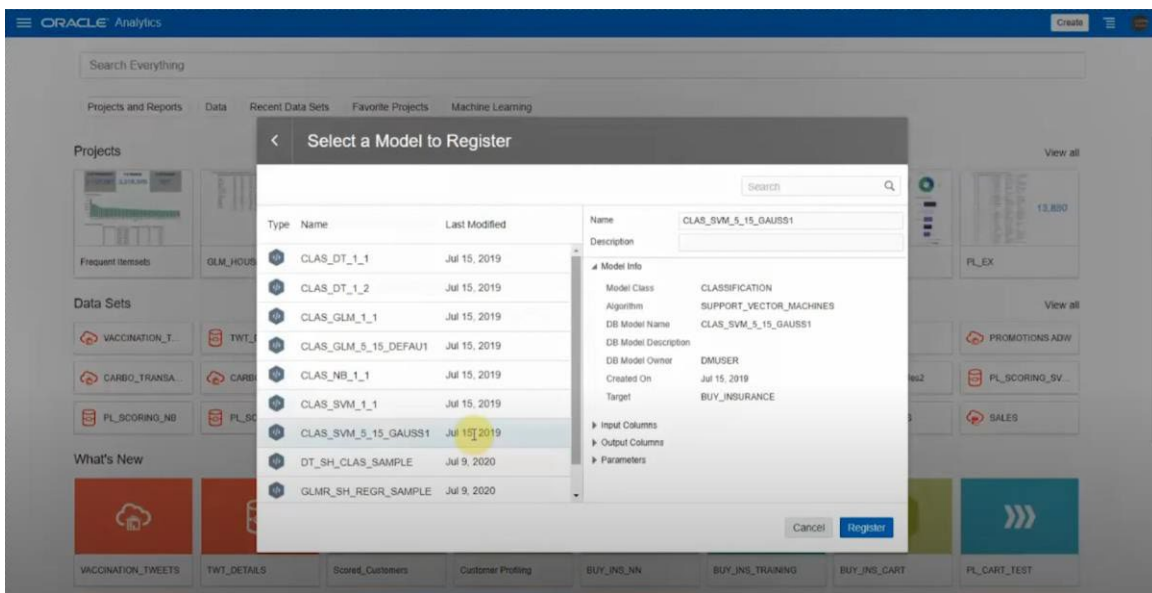


図8： Oracle Analytics Cloudを使用して、分析やダッシュボードで使用するインデータベース・モデルを選択します。

結論

Oracle Machine Learningには、オラクルのコンバインド・データベース戦略の一環として、Oracle Databaseおよびビッグ・データ環境の内部で並列処理されるスケーラブルな機械学習アルゴリズムが用意されています。Oracle Machine Learningは、分類、回帰、クラスタリング、相関ルール、特徴抽出、時系列、ランキング、行評価、異常検出に対応した機械学習アルゴリズムを提供しています。機械学習アルゴリズムはさまざまなインターフェースから使用でき、それには、Oracle Machine Learning Notebooks、Oracle Machine Learning AutoML UI、一般的な言語のAPI（SQL、R、Python）、Oracle Machine Learning Services、Oracle Data Minerが含まれます。

Oracle Machine Learningのネイティブのアルゴリズムは、Oracle Databaseのスケーラビリティ、並列処理、セキュリティ、構造化および非構造化データ処理の各機能の利点を十分に生かすことができます。これによりオラクル製品は、オンプレミスとOracle Cloudのどちらにおいても、機械学習ソリューションを開発し、それらのソリューションをエンタープライズ・アプリケーションに組み込むための強力なプラットフォームとなっています。

オラクルの数十年に及ぶ最先端のデータ管理の経験が、「データではなくアルゴリズムを移動する」という戦略と組み合わせられています。データ管理と機械学習を、集約されたOracle Databaseに統合することにより、総所有コストの削減、データ移動の解消、ソリューションの迅速デプロイというメリットが得られます。

追加情報

[Webサイト](#)

[機械学習ブログ](#)

[ドキュメント](#)

[Office Hoursセッション](#)

[Oracle Machine Learning LiveLabs](#)

Connect with us

+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、[oracle.com/contact](#)で最寄りの営業所をご確認いただけます。

 [blogs.oracle.com](#)

 [facebook.com/oracle](#)

 [twitter.com/oracle](#)

Copyright © 2023, Oracle and/or its affiliates. All rights reserved. 本文書は情報提供のみを目的として提供されており、ここに記載されている内容は予告なく変更されることがあります。本文書は、その内容に誤りがないことを保証するものではなく、また、口頭による明示的保証や法律による黙示的保証を含め、商品性ないし特定目的適合性に関する黙示的保証および条件などのいかなる保証および条件も提供するものではありません。オラクルは本文書に関するいかなる法的責任も明確に否認し、本文書によって直接的または間接的に確立される契約義務はないものとします。本文書はオラクルの書面による許可を前もって得ることなく、いかなる目的のためにも、電子または印刷を含むいかなる形式や手段によっても再作成または送信することはできません。

本デバイスは、連邦通信委員会のルールに基づいた認可を未取得です。認可を受けるまでは、このデバイスの販売またはリースを提案することも、このデバイスを販売またはリースすることはありません。

OracleおよびAvalはOracleおよびその子会社、関連会社の登録商標です。その他の名称はそれぞれの会社の商標です。

IntelおよびIntel XeonはIntel Corporationの商標または登録商標です。すべてのSPARC商標はライセンスに基づいて使用されるSPARC International, Inc.の商標または登録商標です。AMD、Opteron、AMDロゴおよびAMD Opteronロゴは、Advanced Micro Devices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：データシートにこの免責事項の記載が必要かどうか分からない場合は、収益認識方針を参照してください。本書の内容と免責事項の要件についてさらに質問がある場合は、[REVREC_US@oracle.com](#)宛てに電子メールでご連絡ください。