

Oracle Machine Learningの 概要

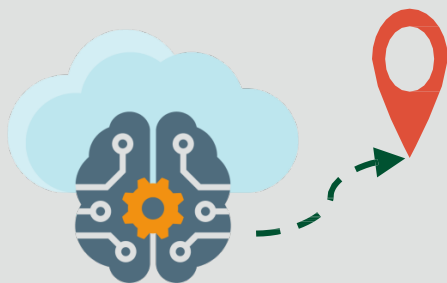
Mark Hornick, Senior Director
Oracle Machine Learning Product Management
mark.hornick@oracle.com
@MarkHornick



免責条項

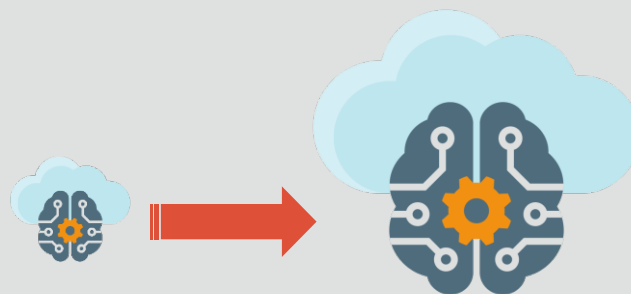
下記事項は、弊社の一般的な製品の方向性に関する概要を説明するものです。また、情報提供を唯一の目的とするものであり、いかなる契約にも組み込むことはできません。マテリアルやコード、機能の提供をコミットメント（確約）するものではなく、購買を決定する際の判断材料になさらないでください。オラクルの製品に関して記載されている機能の開発、リリース、時期、および価格は変更になる場合があります。その決定は弊社独自の裁量で行われます。

Oracle Machine Learningの主な特徴



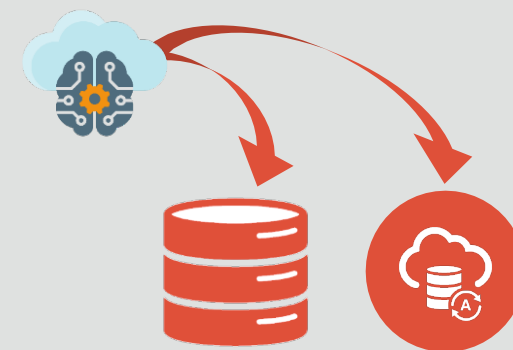
自動化

少ない労力で質の高い
結果をより早く取得 –
エキスパートでない
ユーザーでも可能



スケーラブル

並列分散アルゴリズムを使用して
ビッグ・データ規模のデータを処理
– データの移動は不要

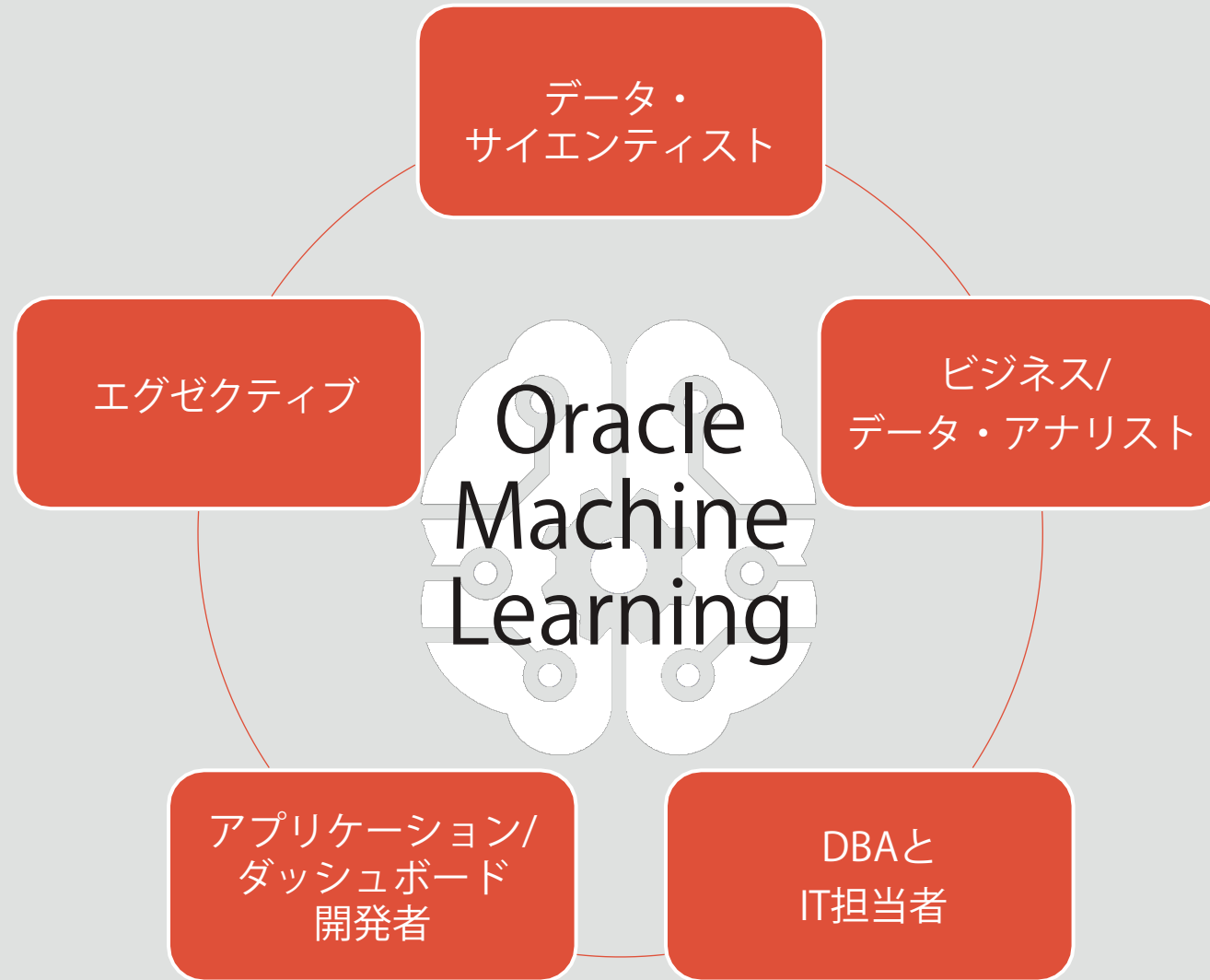


すぐに実稼働可能

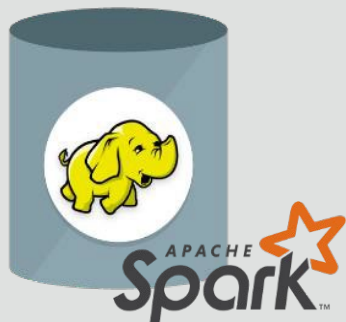
統合されたMLプラット
フォームを使用して、データ・
サイエンス・ソリューションを
より早くデプロイ、更新

生産性を向上させ、企業目標を達成し、より多くのイノベーションを実現

エンタープライズ・ユーザーを強力にサポート



ORACLE®
APPLICATIONS



Oracle Machine Learning

OML4SQL
SQL API

OMLノートブック
Autonomous Database上で
Apache Zeppelinを使用

OML4R
R API

Oracle Data Miner
Oracle SQL Developer拡張機能

OML4Py*
Python API

OML4Spark
Big Dataに対するR API

OML AutoML UI*
Autonomous Database上
でのコードフリーのAutoML
インターフェース

OML Services*
モデルのデプロイと管理、
コグニティブ・テキスト

ORACLE®
APPLICATIONS



* 近日追加予定



Oracle Machine Learningのアルゴリズムと分析

分類

ナイーブ・ベイズ
ロジスティック回帰 (GLM)
ディジション・ツリー・ランダム・
フォレストニューラル・ネット
ワーク
サポート・ベクター・マシン (SVM)
明示的セマンティック分析XGBoost*

異常検出

One-Class SVM
MSET-SPRT*

クラスタリング

階層型k-means 階層型O-Cluster
期待値最大化 (EM)

時系列

予測 - 指数平滑法
一般的なモデルを含む
例：トレンド、季節性、不規則な
時系列を扱うHolt-Winters

回帰

線形モデル
一般化線形モデル (GLM)
サポート・ベクター・マシン
(SVM) ステップワイズ線形回帰
ニューラル・ネットワーク
LASSO
XGBoost*

属性の重要度

最小記述長主成分分析 (PCA) 非管理型
ペアワイズKLダイバージェンス行および
AIのCUR分解

相関ルール

アプリオリ/マーケット・バスケット

予測問合せ

予測、クラスタ、検出、特徴

SQL分析

SQLウィンドウ
SQLパターン SQL
集計

特徴抽出

主成分分析 (PCA) Non-negative
Matrix Factorization
特異値分解 (SVD) 明示的
セマンティック分析 (ESA)

行の重要度

CUR分解

ランキング

XGBoost*

テキスト・マイニングのサポート

アルゴリズムでテキスト列をサポート
トークナイゼーションとテーマ抽出明
示的セマンティック分析 (ESA)

統計関数

最小、最大、中央値、標準偏差、t検定、
F検定、ピアソンのカイ二乗検定、
分散分析、その他

RおよびPythonのパッケージ

組込み実行によるサード・パーティの
RおよびPythonのパッケージ
Spark MLlibアルゴリズム統合





Oracle Machine Learning ノートブック

データ・サイエンス・プラットフォームとしての
Autonomous Database 共同作業に適したUI

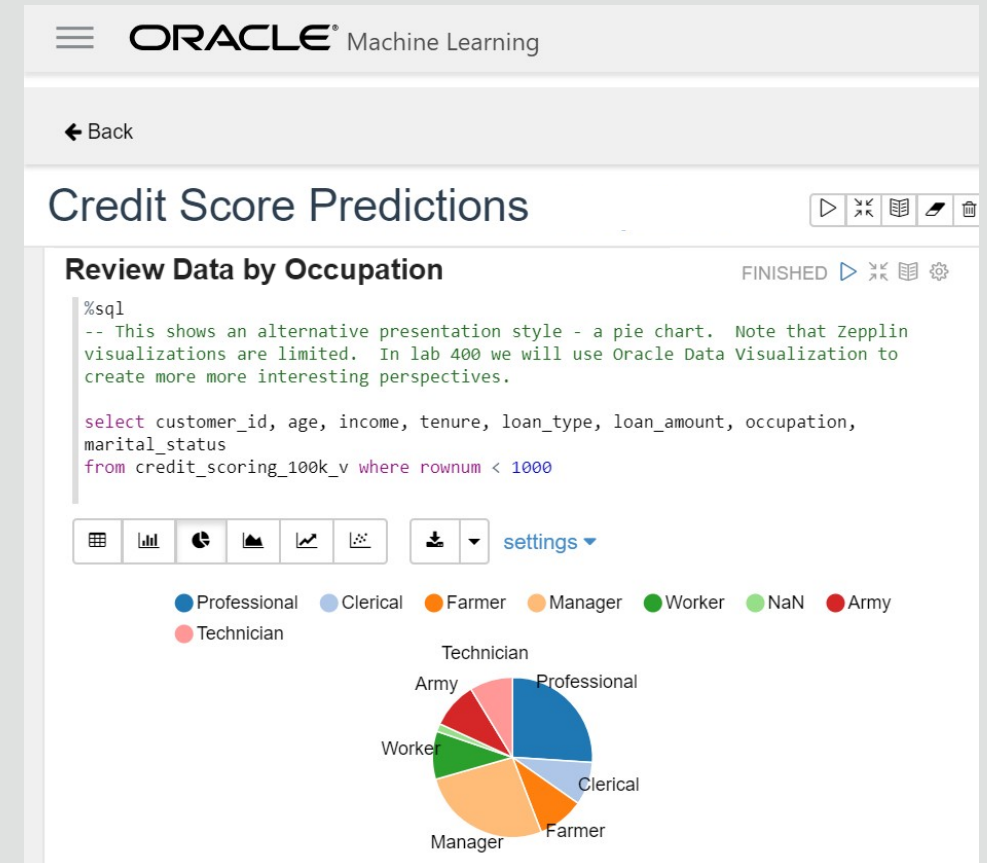
Apache Zeppelinが基盤

SQLとPython*を使用するデータ・サイエンティスト、
データ・アナリスト、

アプリケーション開発者、DBAをサポート
ノートブックとテンプレートの容易な共有
アクセス権、バージョニング、実行スケジュー
リング

Autonomous Databaseに搭載

自動のプロビジョニング、管理、バックアップ
インデータベース・アルゴリズムおよび分析関
数調査およびモデルの準備、構築、評価、デー
タのスコアリング、ソリューションのデプロイ



Oracle Machine Learning for SQL

Oracle DatabaseおよびOracle Autonomous Database内の
MLにすぐにアクセスできるようSQLユーザーをサポート

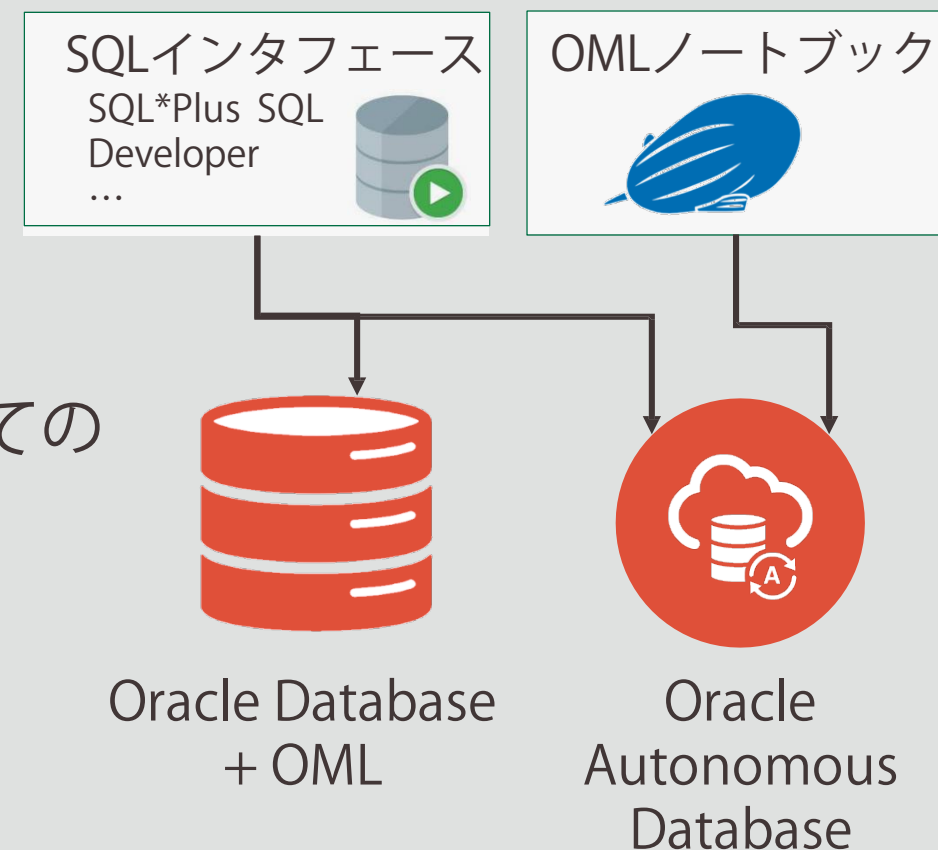
インデータベースの並列分散アルゴリズム

- MLエンジンを分離するためのデータ抽出が不要
- 高速でスケーラブル
- バッチおよびリアルタイムのスコアリング
- 分かりやすい予測の詳細

もっとも重要なデータベース・オブジェクトとしての MLモデル

- アクセス権を使用したアクセス制御
- ユーザー・アクションの監査
- データベース間のモデルのエクスポート/インポート

Oracleスタック全体でのMLの活用



Oracle Data Minerのユーザー・インタフェース

分析ワークフローの作成 – データ・サイエンティスト向けの生産性ツール - 専門家ではなくてもデータ・サイエンティストに



オンプレミスおよびDBCSの
Oracle Database用の
SQL Developer拡張機能

標準的なデータ・サイエンスの
手順を自動化

使いやすいドラッグアンド
ドロップ・インタフェース

分析ワークフローを迅速に
定義、共有

幅広いアルゴリズムとデータ
変換処理

即時デプロイメント用の
SQLコードの生成

The screenshot shows the Oracle SQL Developer interface with the Oracle Data Miner extension. The main workspace displays a workflow diagram with nodes like 'Scatter Box plots etc. 1', 'Explore Data', 'CUST_INSUR_LTV1', 'Filter Columns', 'Multiple Classification Models', 'Most Likely Customers', and 'Explore Data 1'. The 'Query Builder' window shows a SQL query for 'Top 5 most suspicious fraud policy holder claims'. The 'Script Output' window displays the results of the query. The 'Components' panel on the right lists various data mining tasks and models.

Query Builder:

```
begin
dbms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION',
'CLAIMS', 'POLICYNUMBER', null, 'CLAIMS_SET');
end;
```

Script Output:

POLICYNUMBER	PERCENT_FRAUD	RANK
654	61.87	1
11068	57.37	2
7435	55.47	3

Multiple Classification Models - Properties:

Model	Support	Confidence	Accuracy
Model 1	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 2	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 3	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 4	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 5	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 6	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 7	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 8	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 9	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 10	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 11	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 12	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 13	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 14	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 15	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 16	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 17	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 18	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 19	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 20	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 21	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 22	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 23	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 24	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 25	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 26	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 27	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 28	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 29	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 30	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 31	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 32	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 33	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 34	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 35	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 36	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 37	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 38	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 39	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 40	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 41	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 42	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 43	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 44	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 45	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 46	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 47	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 48	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 49	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%
Model 50	Support: 100 (100%)	Confidence: 80%	Accuracy: 80%

Rule Surrogates Target Values:

Rule	Surrogates	Target Values
Node Rule:		
If	BANK_FUNDS > 246	
And	CHECKING_AMOUNT > 282	
And	MONEY_MONTHLY_OVERDRAWN <= 54.215	
Then No		
Confidence	0.8515671200473093	
Support	0.1832863646217212	

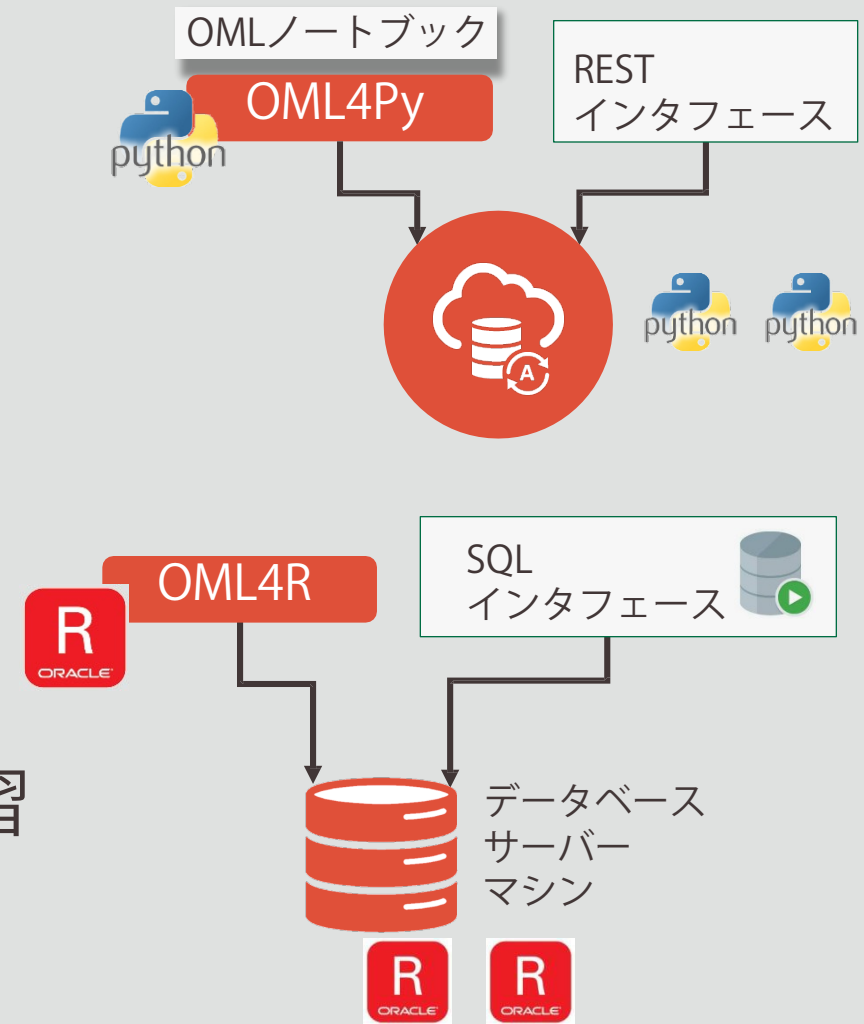
Oracle Machine Learning for R、Python

Oracle Database – R (Pythonは近日追加予定)
Oracle Autonomous Database – Python*
(Rは近日追加予定)

HPC環境としてのOracle Database並列で分散
されたインデータベースの機械学習
アルゴリズム

Oracle Database内でスクリプトと
オブジェクトを管理結果をSQLまたは
REST経由でアプリケーションおよび
ダッシュボードに統合OML4Pyの自動機械学習

オープン・ソース環境において
データ・サイエンティストをサポート



AutoML – OML4Pyの新機能

データ・サイエンティストの生産性向上 – 全体的なコンピュータ時間を削減



アルゴリズムの自動選択

- もっとも高い品質のモデルとなるインデータベース・アルゴリズムを特定
- しらみつぶし探索よりも高速に実行できる最適なアルゴリズムを検索

特徴の自動選択

- もっとも良い予測を特定して特徴数を削減
- パフォーマンスと精度の改善

モデルの自動チューニング

- アルゴリズムのハイパーパラメータを自動でチューニング
- モデル精度が大幅に向上
- 手動探索またはしらみつぶし探索手法を回避

エキスパートでないユーザーでも機械学習の活用が可能に

Oracle Machine Learning for Spark

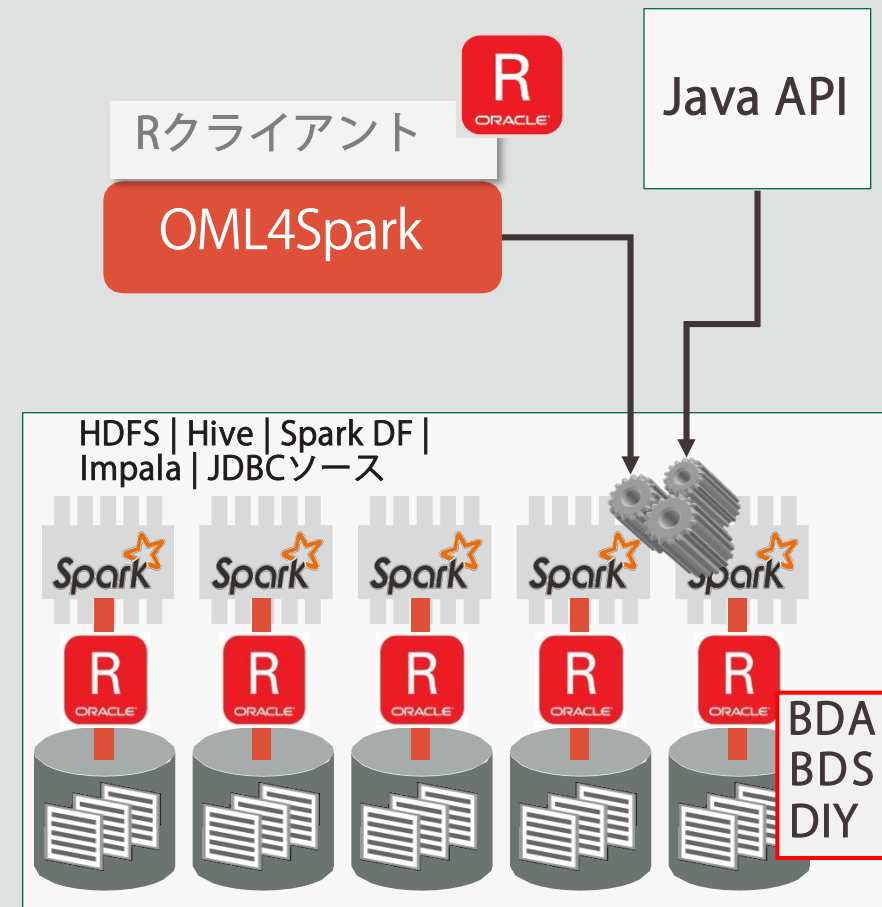


Oracle Big Data ConnectorsへのR言語APIコンポーネント

強力なデータ準備と機械学習のためにSpark 2環境を活用

幅広いデータ・レイク・ソースにわたってデータを利用Hadoopクラスタを十分に活用してスケーラビリティとパフォーマンスを達成

ネイティブ実装およびSpark MLlib実装による並列分散MLアルゴリズム



OMLを統合するOracleアプリケーション

HCM Cloud

ワークフォース予測

CRM Sales Cloud 販売予測

Retail GBU 顧客インサイト
顧客のセグメント化

Adaptive Intelligent Applications
for Manufacturing

Configure, Price, Quote Cloud

Content and Experience
非構造化データの分析

Integration Cloud
デジタル・プロセスの自動化

ORACLE
HUMAN CAPITAL
MANAGEMENT
CLOUD

ORACLE
SALES CLOUD

ORACLE
RETAIL CLOUD



ORACLE
CONFIGURE, PRICE,
AND QUOTE
CLOUD

ORACLE
CLOUD PLATFORM
CONTENT AND
EXPERIENCE



業界別データ・モデル

通信、SNA、公共事業、航空、小売りなど



EBS Spend Classification

費用を論理的なカテゴリに整理

EBS Depot Repair

製品の修理、再利用、リサイクル
の速度、コスト、品質を最適化

Identity Management

適応型のアクセス管理

FSGBU

分析アプリケーション・
インフラストラクチャ

ORACLE
E-BUSINESS SUITE

ORACLE
IDENTITY MANAGEMENT

ORACLE
FINANCIAL SERVICES

Oracleが機械学習に向いている理由

Oracleスタックとエンタープライズ全体にわたってMLを統合

MLがデータ・サイエンティスト、データ・アナリスト、開発者、DBA/IT部門の力になる

コストのかかるデータ移動と待機時間を排除

高速かつスケーラブルなデータの探索、データの準備、MLアルゴリズム

30を超えるインデータベース・アルゴリズムのサポート：回帰、分類、時系列、クラスタ化、特徴抽出、異常検出などおもなMLプロセス手順の自動化

RおよびPythonとの統合によりデータ・サイエンティストを

サポートMLモデルやR/Pythonスクリプトの容易なデプロイ

既存のバックアップ、リカバリ、セキュリティのメカニズムとプロトコルを活用

企業データの大半がある場所 – アルゴリズムをデータの側に配置

Oracle DatabaseおよびOracle Autonomous Database

今後の予定

ロードマップ：PythonによるAutonomous Databaseの拡張

データ・サイエンス・プラットフォームとしての
Autonomous Database OML

ノートブックでPythonを新たにサポート

SQL、PL/SQL、Markdownに追加

Pythonのスケラブルな実行（OML4Py）

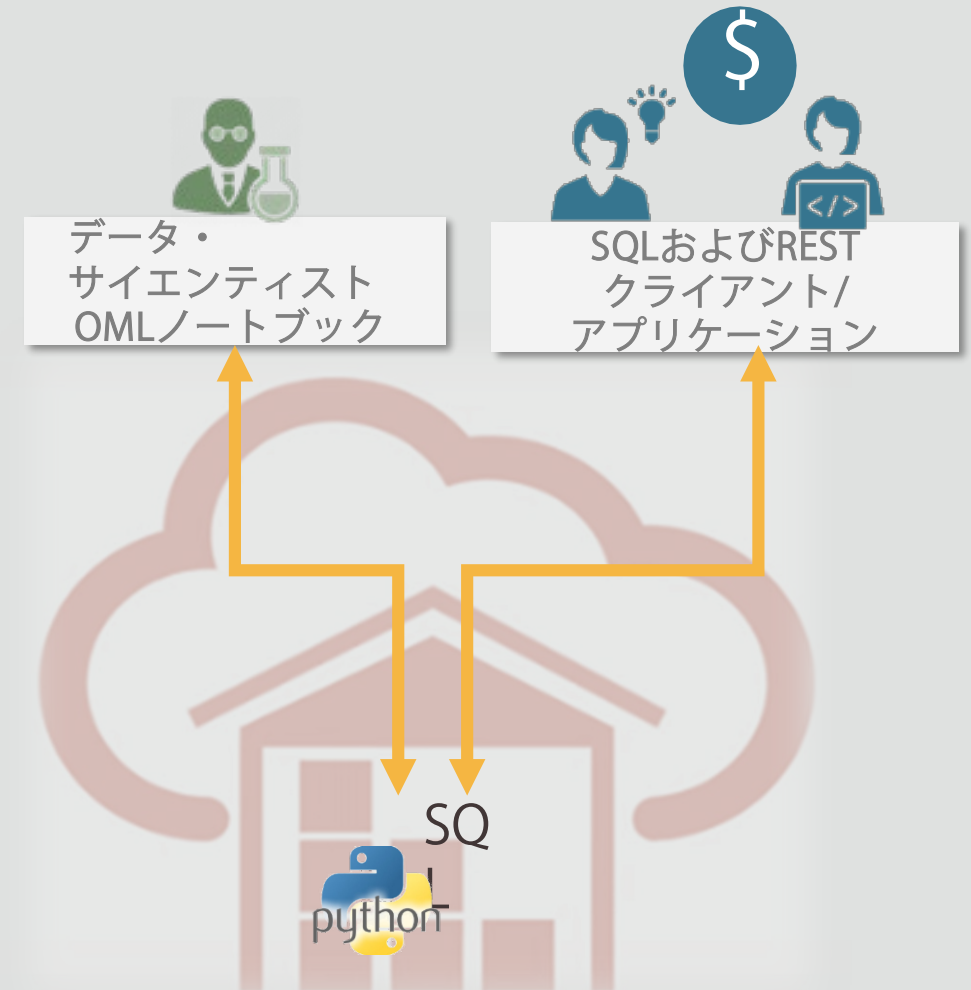
透過レイヤー対応のデータベース機能インデータベースの
機械学習アルゴリズム

自動機械学習（AutoML）

アルゴリズムと特徴の選択モデルの
チューニング

インデータベースで管理されるPython
スクリプト

OMLノートブックおよびREST APIから起動
Webアプリケーションに容易にデプロイ



ロードマップ：OML Services

OML専用のREST API - データベース外でのモデルの作成とデプロイ

モデルの管理とデプロイのサービス

OMLモデルの作成とデプロイ

OML形式とONNX形式のモデル

Tensorflow、PyTorch、MXNet、scikitlearnなどを対象とした

ONNXのインポートMLモデルの保存、バージョンング、比較

OML4Py REST APIとの認証の共有コグニティブ・

テキスト・サービス

トピックやキーワードの抽出

センチメント分析

テキストのサマリーと類似度

モデル管理

GET /models
GET /{model name}
GET /{model name}/{version}
POST /{model name}
POST /{model name}/{version}
DELETE /{model name}/{version}

モデル・デプロイ

GET /models
GET /{uri}
GET /{uri}/api
POST /{uri}
POST /{uri}/score
DELETE /{uri}

コグニティブ・ テキスト

POST /topics
POST /keywords
POST /sentiment
POST /summary
POST /similarity

ロードマップ：OML AutoML UI

エンド・ツー・エンドの自動MLをサポートする“コードフリー”のAutoML
ベースのユーザー・インターフェース

多機能で使いやすいUI

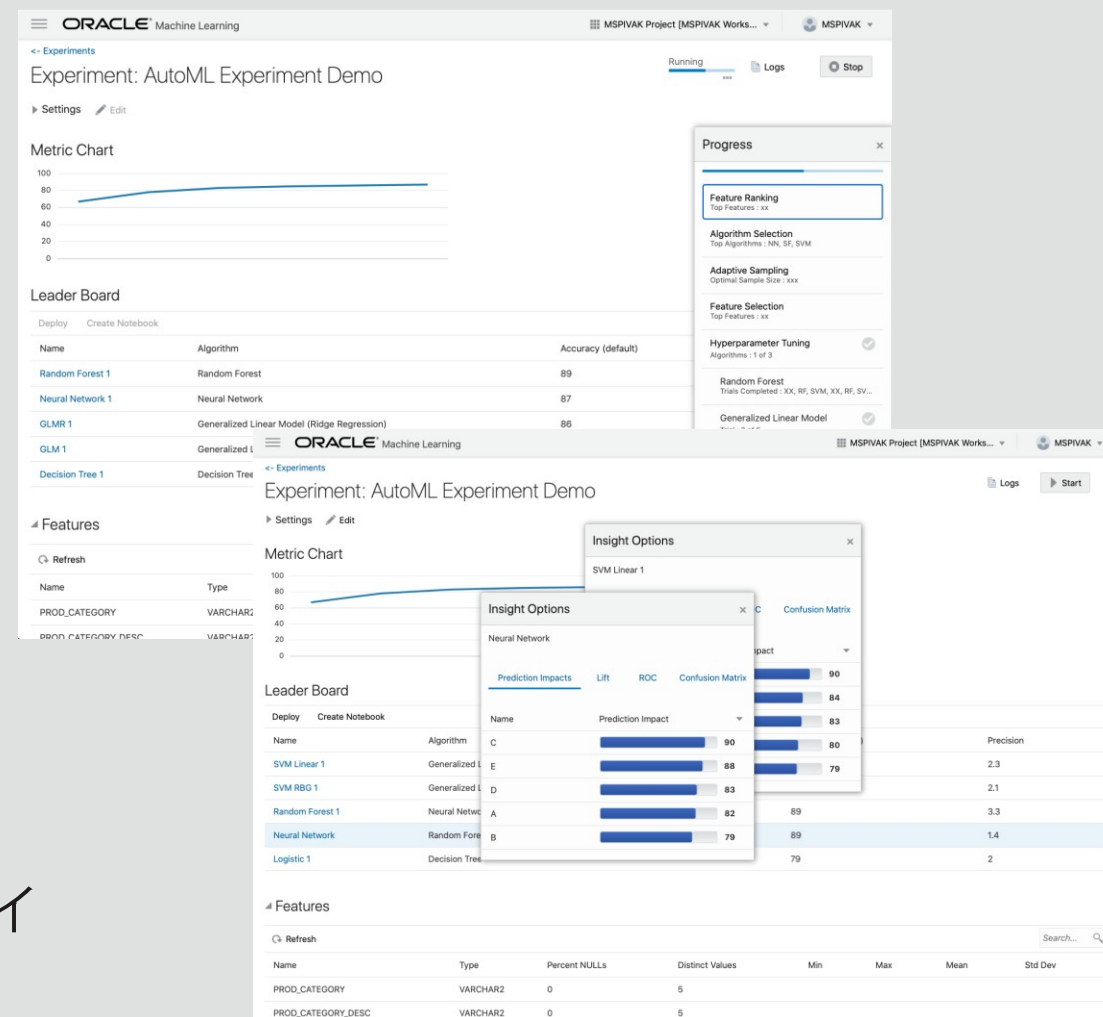
エキスパートでないユーザーもMLを
使用可能に

モデルの作成とデプロイを自動化
データ・サイエンティストの生産
性の向上モデル管理のサポート

機能

ユーザー入力は最小限：
データ、ターゲット モデルの
ランク表示

RESTエンドポイント経由のモデルのデプロイ



追加情報

oracle.com/machine-learning

Database / Technical Details /
Machine Learning



Oracle Machine Learning

The Oracle Machine Learning product family enables scalable data science projects. Data scientists, analysts, developers, and IT can achieve data science project goals faster while taking full advantage of the Oracle platform.

Oracle Machine Learning consists of complementary components supporting scalable machine learning algorithms for in-database and big data environments, notebook technology, SQL and R APIs, and Hadoop/Spark environments.

ありがとうございました

Oracle Machine Learning Product Management