

Oracle Machine Learning

Introduction

Move the Algorithms – Not the Data



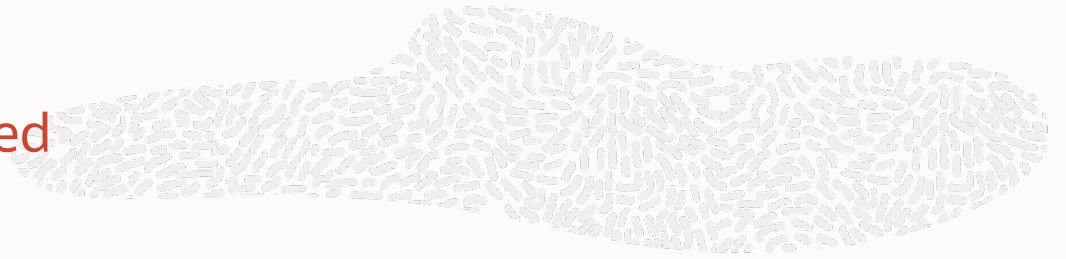
Renée Wikestad
Cloud Specialist Engineer,
Oracle Norge

April 2021



Building on a converged database

In cloud and on premises – integrated, not fragmented



Supports any Data

Relational, JSON, graph, spatial, text, OLAP, XML, multimedia

Supports any Workload

Transactions, analytics, ML, IoT, streaming, blockchain

Most Productive for Developers and Analysts

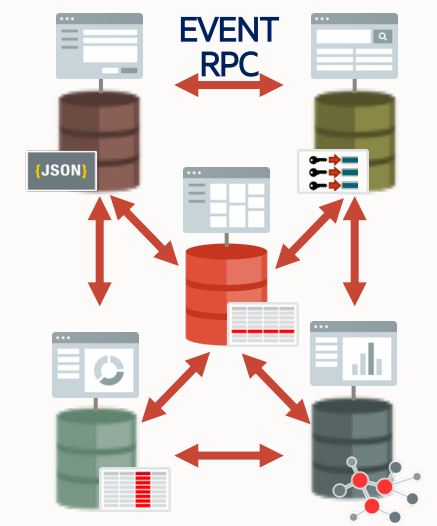
Integrated microservices, events, REST, SaaS, ML, CI/CD, Low-Code

Contrasting Architectures

Oracle – **Simpler, integrated**. Dramatically simpler for developers to invoke declarative SQL, Python and R to run machine learning, spatial, graph, JSON, unstructured, IoT, blockchain, etc. in one “**Converged**” Database



Others – **Fragmented**. Multiple fragmented databases, platforms, **tools**, and distributed execution and **data movement** across them



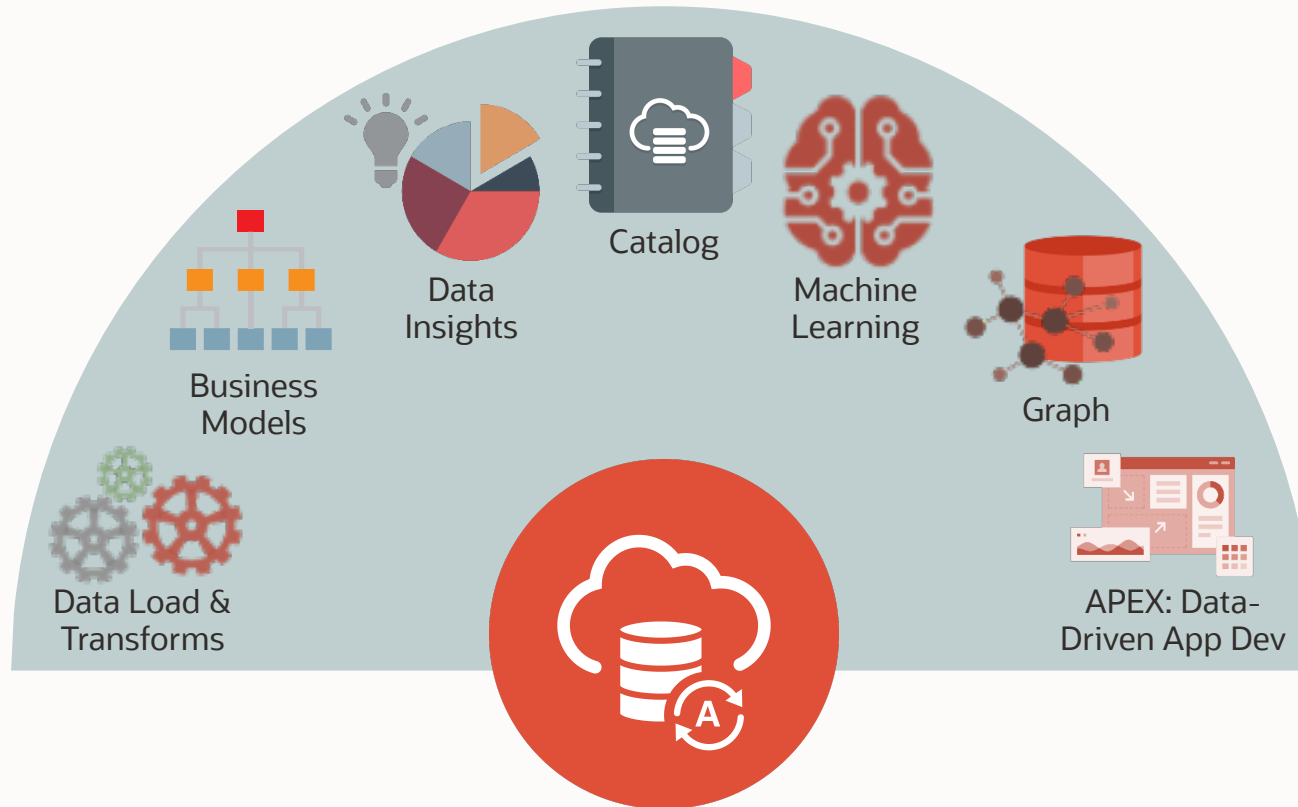
Oracle Autonomous Database is an ecosystem

Autonomous Database **with built-in tool suite**



Data Sources

Applications	
Databases	
Files	
Streaming Services	
Data Lake	



Visualization & Development

Application Express	
Oracle ML Notebook	
Oracle Analytics Cloud	
IDEs	
Spreadsheets	
3rd Party BI Tools	

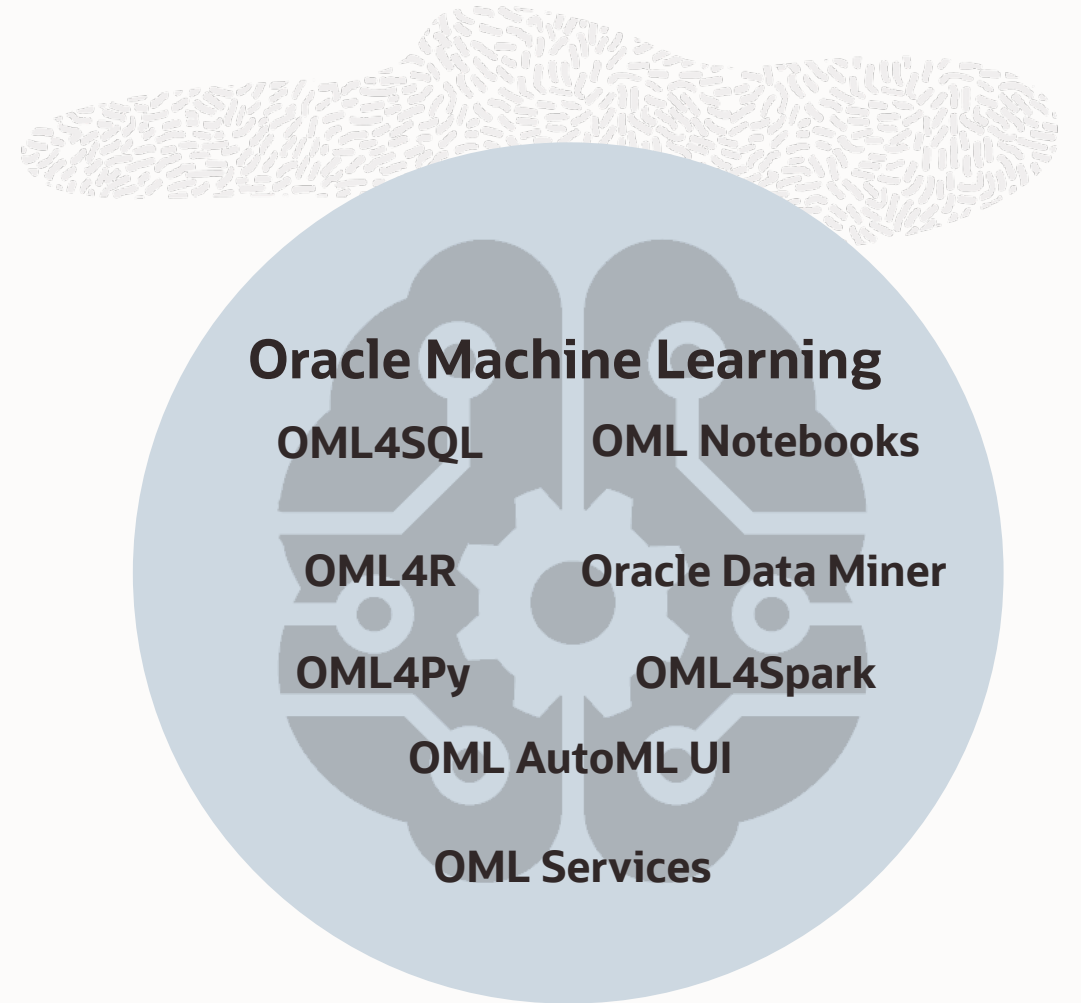


Oracle Machine Learning

Move the algorithms – not the data

Oracle Machine Learning extends Oracle Database and enables users to augment applications and dashboards with machine learning-based intelligence

OML delivers 30+ powerful in-database machine learning algorithms with automated functionality via SQL, R and Python APIs



Oracle Machine Learning database features included with Oracle Database and Oracle Autonomous Database

Oracle Machine Learning Algorithms and Analytics

CLASSIFICATION

- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine (SVM)
- Explicit Semantic Analysis
- ***XGBoost****

ANOMALY DETECTION

- One-Class SVM
- ***MSET-SPRT****

CLUSTERING

- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

TIME SERIES

- Forecasting - Exponential Smoothing
- Includes popular models
e.g. Holt-Winters with trends,
seasonality, irregularity, missing data

REGRESSION

- Linear Model
- Generalized Linear Model (GLM)
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- ***XGBoost****

ATTRIBUTE IMPORTANCE

- Minimum Description Length
- Principal Component Analysis (PCA)
- Unsupervised Pairwise KL Divergence
- CUR decomposition for row & AI

ASSOCIATION RULES

- A priori/ market basket

SQL ANALYTICS

- SQL Windows
- SQL Patterns
- SQL Aggregates

FEATURE EXTRACTION

- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

ROW IMPORTANCE

- CUR Decomposition

RANKING

- ***XGBoost****

TEXT MINING SUPPORT

- Algorithms support text columns
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA)

STATISTICAL FUNCTIONS

- min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

R AND PYTHON PACKAGES

- Third-party R and Python Packages through Embedded Execution

Oracle Machine Learning for SQL

OML4SQL



What it is

- SQL and PL/SQL interface for in-database ML
- Build, evaluate, deploy models and score data using over 30 in-database parallelized algorithms
- Automatic data preparation that is algorithm-specific
- Score data using prediction operators directly in SQL queries
- Partition models, text mining, and prediction details
- Supports numeric, categorical, text, and nested columns

Benefits

- Scalable in-database ML algorithms
- Faster time-to-market through immediate solution deployment
- Score data on a scalable platform
- In-database models exist as first-class database objects for ease of backup, recovery, and security
- Supports R and Python interfaces

Intuitive SQL API—OML4SQL

OML to Predict Customer Behavior



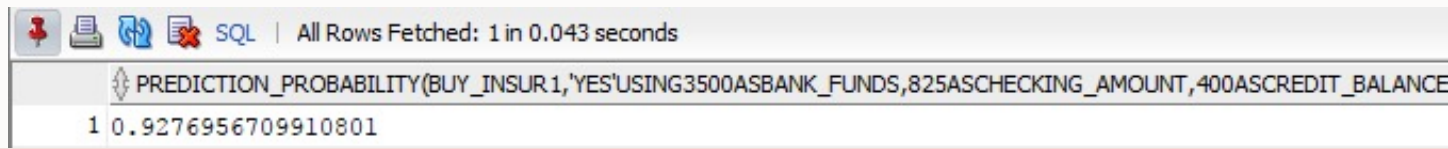
-- Build a machine learning model to determine which customers are likely buy Travel Insurance

```
DECLARE
  v_setlst DBMS_DATA_MINING.SETTING_LIST;
BEGIN
  v_setlst('ALGO_NAME') := 'ALGO_SUPPORT_VECTOR_MACHINES';
  V_setlst('PREP_AUTO') := 'ON';
  DBMS_DATA_MINING.CREATE_MODEL2 (
    MODEL_NAME           => 'BUY_TRVL_INSUR',
    MINING_FUNCTION      => 'CLASSIFICATION',
    DATA_QUERY          => 'select * from CUSTOMERS',
    SET_LIST             => v_setlst,
    CASE_ID_COLUMN_NAME => 'CUST_ID',
    TARGET_COLUMN_NAME  => 'BUY_TRAVEL_INSURANCE');
END;
```



-- Apply a machine learning model to predict which customers are likely to buy

```
SELECT prediction_probability(BUY_TRVL_INSUR, 'Yes'
  USING 3500 as bank_funds, 37 as age, 'Married' as marital_status, 2 as num_previous_cruises)
FROM dual;
```



Oracle Autonomous Database and OML Notebooks

Notebook with model created in SQL



ORACLE Machine Learning

FINISHED

Using Support Vector Machine Regression to Predict Volume of Bikes

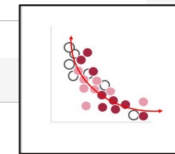
This notebook shows how to predict numerical values using Support Vector Machine (SVM) Regression. Given demographic information when renting bikes to predict the volume of bikes rented out in column Volume.

– Build the SVM regression model

```
DBMS_DATA_MINING.CREATE_MODEL(  
MODEL_NAME           => 'BIKES_REGR_MOD',  
MINING_FUNCTION      => DBMS_DATA_MINING.REGRESSION,  
DATA_TABLE_NAME     => 'BIKES_TRAIN',  
CASE_ID_COLUMN_NAME => 'DAY_ID',  
TARGET_COLUMN_NAME  => 'VOLUME',  
SETTINGS_TABLE_NAME => 'BIKES_REGR_SETTINGS',  
XFORM_LIST           => V_XLST);
```

```
%sql  
select * from BIKES_REGR_SETTINGS;
```

SETTING_NAME	SETTING_VALUE
ALGO_NAME	ALGO_SUPPORT_VECTOR_MACHINES
SVMS_KERNEL_FUNCTION	SVMS_GAUSSIAN
PREP_AUTO	ON



Updated March 2020 by Renée Wikestad.

Apply the model to a singel record

```
%sql  
select round(prediction(BIKES_REGR_MOD using 'Winter' AS season, 'Light Rain' as weathersit,0.45 as temp,0.2 as windspeed),0) VolumPrediction from dual  
  
--temp:  $X_{nom} = \frac{X - X_{min}}{X_{max} - X_{min}}$  0.45 = (1.5+30)/(40+30)
```

VOLUMPREDICTION

2620



Oracle Machine Learning for R

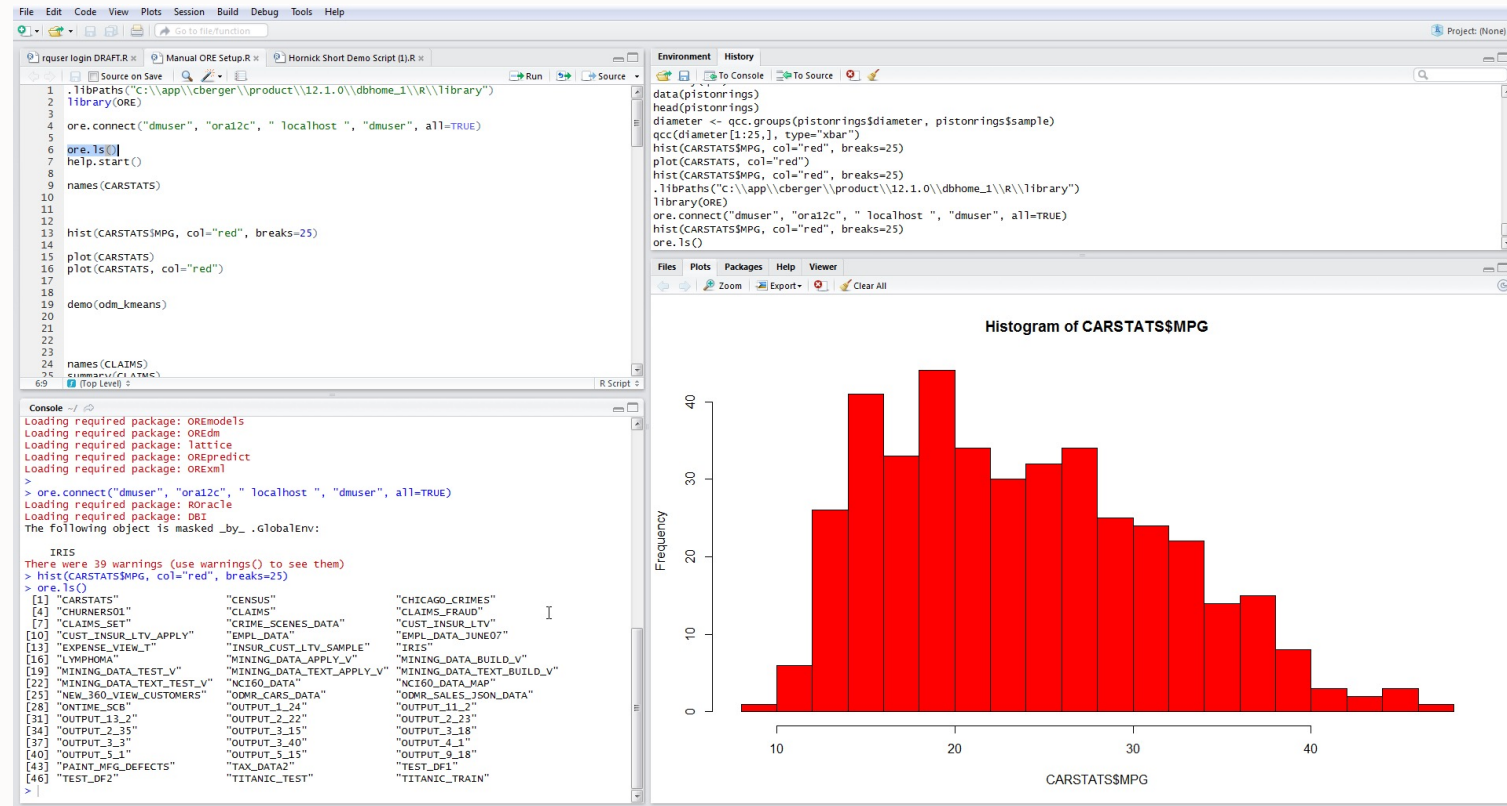
OML4R

R Language API to OML Algorithms and Integration with R



Key features

- Use Oracle Database as HPC environment via R→SQL transparency layer
- Access in-database parallelized machine learning algorithms
- Manage scripts and objects in Oracle Database
- Integrate results into applications and dashboards via SQL
- Use third-party packages with Embedded R Execution



Oracle Machine Learning for Python

OML4Py

Supported in Oracle Autonomous Database with OML Notebooks



Use Oracle Database as HPC environment

- Explore, transform, and analyze data faster and at scale

Use in-database parallelized and distributed ML algorithms

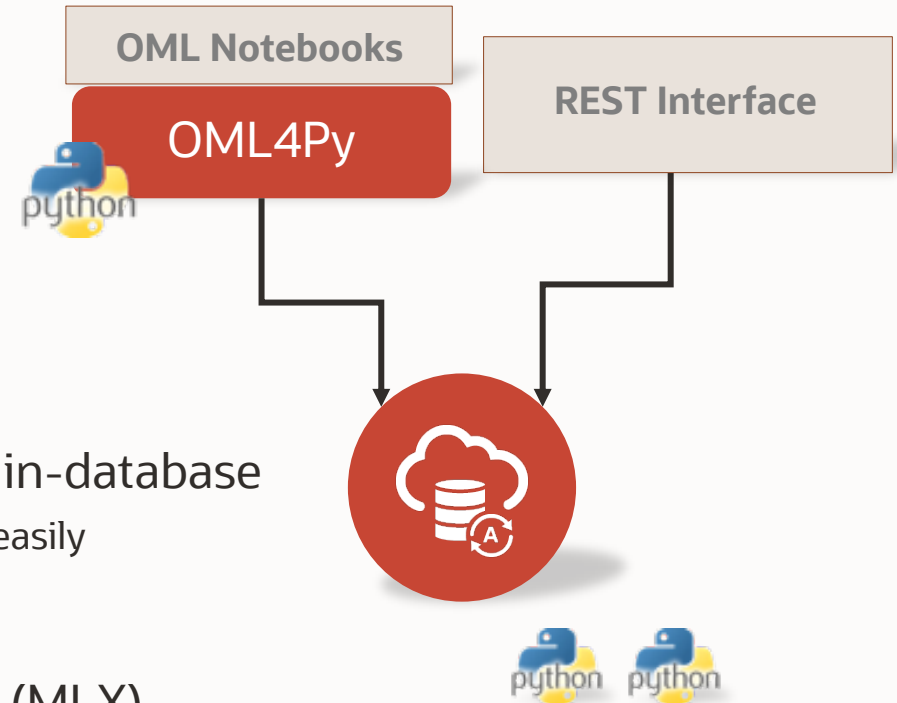
- Build more models on more data, and score large volume data – faster
- Use in-database algorithms from OML4SQL via natural Python API
- Increased productivity from automatic data preparation, partitioned models, and integrated text mining capabilities

Run user-defined Python functions and manage Python objects in-database

- Collaborate: hand-off data science products from data scientist to developers easily
- Return structured and image results in Python and REST API

Automated machine learning (AutoML) and model explainability (MLX)

- Algorithm selection, feature selection, hyperparameter tuning, model selection
- Model-agnostic identification of important features that impact model predictions



Oracle Machine Learning for Python Notebooks UI

Autonomous Database as a Data Science Platform



Collaborative UI

- Based on Apache Zeppelin
- Supports data scientists, data analysts, application developers, DBAs with **SQL** and **Python**
- Easy notebook sharing
- Permissions, versioning, and scheduling of notebooks

Included with Autonomous Database

- Automatically provisioned and managed
- In-database algorithms and analytics functions
- Explore and prepare, build and evaluate models, score data, deploy solutions

The screenshot displays the Oracle Machine Learning for Python Notebooks UI. The interface includes a header with the Oracle logo and 'Machine Learning' text, a user profile 'OMLUSER', and a 'Connected' status indicator. The main content area shows three notebooks:

- Overloaded data visualization functions:** This notebook explains that OML4Py overloads select graphics functions. It includes a code cell for a boxplot and a corresponding boxplot visualization titled 'Distribution of IRIS Attributes' showing distributions for SEPAL_LENGTH, SEPAL_WIDTH, PETAL_LENGTH, and PETAL_WIDTH.
- histogram:** This notebook shows a code cell for a histogram and a corresponding histogram visualization titled 'Sepal Length variation in IRIS data set' showing the distribution of sepal lengths.
- Create derived variables:** This notebook shows a code cell for creating derived variables: `is_large_petal = (IRIS['PETAL_LENGTH'] > 5.0) & (IRIS['PETAL_WIDTH'] > 2.0)`.



Transparency Layer

In-database performance – indexes, query optimization, parallelism, partitioning



Leverages proxy objects for database data: `oml.DataFrame`

```
# Create table from Pandas DataFrame data
DATA = oml.create(data, table = 'CUSTOMER')

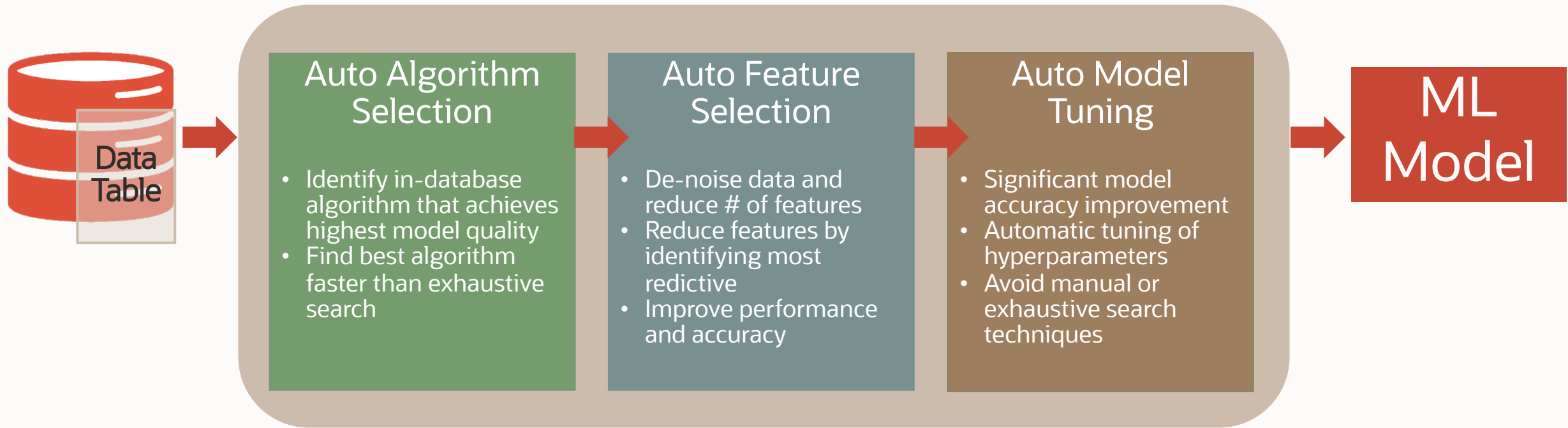
# Get proxy object to DB table boston
DATA = oml.sync(table = 'CUSTOMER')
```

Uses familiar Python syntax to manipulate database data

Overloads Python functions translating functionality to SQL

AutoML – with OML4Py

Faster and Easier Machine Learning for Data Scientists, Non-Experts and Developers



Enables non-expert users to leverage Machine Learning



OML AutoML UI

No-code AutoML-based user interface supporting automatic machine learning



Powerful, easy to use UI

- Enable non-expert users to use ML

Automates model building, tuning, and deployment

- Enhance data scientist productivity
- Support model management
- Empower non-expert users

Featuring

- Minimal user input: data, target
- Model leaderboard
- Model deployment via REST endpoints

Leader Board

Name	Algorithm	Accuracy (default)
Random Forest 1	Random Forest	89
Neural Network 1	Neural Network	87
GLMR 1	Generalized Linear Model (Ridge Regression)	86
GLM 1	Generalized Linear Model	
Decision Tree 1	Decision Tree	

Features

Name	Type
PROD_CATEGORY	VARCHAR2
PROD_CATEGORY_DESC	VARCHAR2

Insight Options

Name	Prediction Impact	Lift	ROC	Confusion Matrix
SVM Linear 1	90			
Neural Network	88			
SVM RBG 1	83			
Random Forest 1	82			
Neural Network	79			
Logistic 1	79			

Confusion Matrix

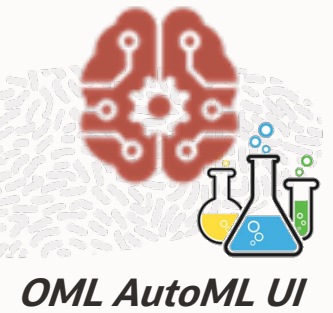
	Actual	Predicted	Impact
0	100	90	90
1	0	84	84
2	0	83	83
3	0	80	80
4	0	79	79

Features

Name	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
PROD_CATEGORY	VARCHAR2	0	5				
PROD_CATEGORY_DESC	VARCHAR2	0	5				



Create Notebook from OML AutoML UI



<- Experiments

BIKE Renting

Experiment Settings Edit

Completed Start

Attribute importance

Model Detail - svmg_d73faf48b9

Prediction Impacts Confusion Matrix

Name	Prediction Impact
YEAR	[Bar chart showing high impact]
MNTH	[Bar chart showing medium impact]
SEASON	[Bar chart showing low impact]
WEATHERSIT	[Bar chart showing very low impact]
WEEKDAY	[Bar chart showing very low impact]

Leader Board

Algorithm	Model Name	R2	Value
Support Vector Machine (Gauss...	svmg_d73faf48b9	0.7690	-458.5997
Support Vector Machine (Linear)	svml_304aa08292	0.7317	-529.5140
Generalized Linear Model	glm_c307596388	0.7134	-557.1522

Deploy Create Notebook Metrics

Features

Create Notebook for selected model

Model Detail - svmg_d73faf4...

Create Notebook

Create a notebook based on selected model and this experiment's settings. Use a generated notebook to further tune your approach using Python.

Notebook Name:

OK Cancel



Oracle Autonomous Database and OML Notebooks

Generated Notebook



ORACLE Machine Learning ML Project [ML Workspace] MLUSER Connected

BIKE_VOLUM_PRED_rg

```
Build Data
%python
import oml

columns = 'INSTANT', 'ATEMP', 'HOLIDAY', 'HUM', 'MNTN', 'SEASON', 'TEMP', 'WEATHERSIT', 'WEEKDAY', 'WINDSPEED', 'WORKINGDAY', 'YEAR', 'COUNT'
schema='MLUSER'
table='BIKE_RENTING_TRAINING'

column = ','.join(columns)
query = 'SELECT ' + column + ' FROM ' + schema + '.' + table

data_build = oml.sync(query=query)
z.show(data_build)

Create Train Data
%python
import oml

X_train = data_build[:,['INSTANT', 'ATEMP', 'HOLIDAY', 'HUM', 'MNTN', 'SEASON', 'TEMP', 'WEATHERSIT', 'WEEKDAY', 'WINDSPEED', 'WORKINGDAY', 'YEAR']]
y_train = data_build[:, 'COUNT']

Build 'SUPPORT_VECTOR_MACHINES' Model
%python
import oml

svm_settings = {
    'ALGO_NAME': 'ALGO_SUPPORT_VECTOR_MACHINES', 'SVMS_COMPLEXITY_FACTOR': '66.70000000000002', 'SVMS_EPSILON': '0.3975300444430571', 'SVMS_KERNEL_FUNCTION': 'SVMS_GAUSSIAN', 'SVMS_STD_DEV': '2.345207879911715', 'SVMS_NUM_PIVOTS': '200', 'ODMS_DETAILS': 'ODMS_DISABLE'
}

svm_mod = oml.svm('regression', **svm_settings)
svm_mod = svm_mod.fit(X_train, y_train)

Show Model Details
%python
svm_mod

Train Data
%python
```

OML4Py Model (Regression with svmg)





OML Services

REST API – deploy models outside the database

Model Management and Deployment Services

- Deploy in-database models and using ONNX format, third-party models
- Import ONNX for Tensorflow, PyTorch, MXNet, scikitlearn, etc.
- Store, version, compare ML models

Cognitive Text Services

- Extract topics and keywords
- Sentiment analysis
- Text summary and similarity



Oracle Machine Learning Services - Methods

Components with built-in Oracle Machine Learning



Admin

POST

- Token using ADB user and password

Generic

GET

- Metadata for all Versions: Version 1 Metadata
- Open API Specification

Repository

POST

- Store Model
- Update Model Namespace

GET

- Models list
- Model Info
- Model Metadata
- Model Content

DELETE

- Model

Deployment

POST

- Create Model Endpoint
- Score Model using Endpoint

GET

- Endpoints
- Endpoint Details
- Open API Specification for Endpoint

DELETE

- Endpoint

Cognitive Text

POST

- Get Most Relevant Topics
- Get Most Relevant Keywords
- Get Summaries
- Get Sentiments
- Get Semantic Similarities
- Numeric Features

GET

- Get Endpoints

Score OML Model (Single Class w Pred Details)

Examples 0

BUILD



POST

{{omlserver}}/omlmod/v1/deployment/{{omlModelURI}}/score

Send

Save

Params

Authorization

Headers (10)

Body

Pre-request Script

Tests

Settings

Cookies Code

none

form-data

x-www-form-urlencoded

raw

binary

GraphQL

JSON

Beautify

```
1 {
2   "topNdetails":5,
3   "inputRecords": [
4     {
5       "AGE":41,
6       "BOOKKEEPING_APPLICATION": 1,
7       "CUST_GENDER":"M",
8       "CUST_MARITAL_STATUS":"NeverM",
9       "EDUCATION":"HS-grad",
10      "HOME_THEATER_PACKAGE":1,
11      "HOUSEHOLD_SIZE":"4",
12      "OCCUPATION":"Crafts",
13      "YRS_RESIDENCE":6,
14      "Y_BOX_GAMES":1
15    }
16  ]
17 }
```

Response

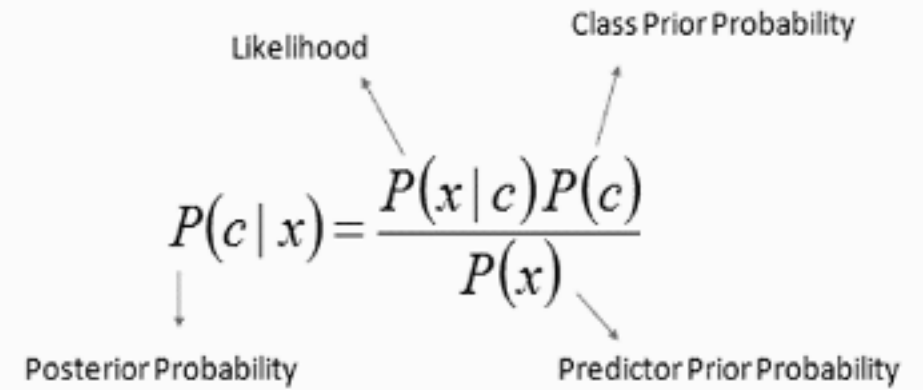
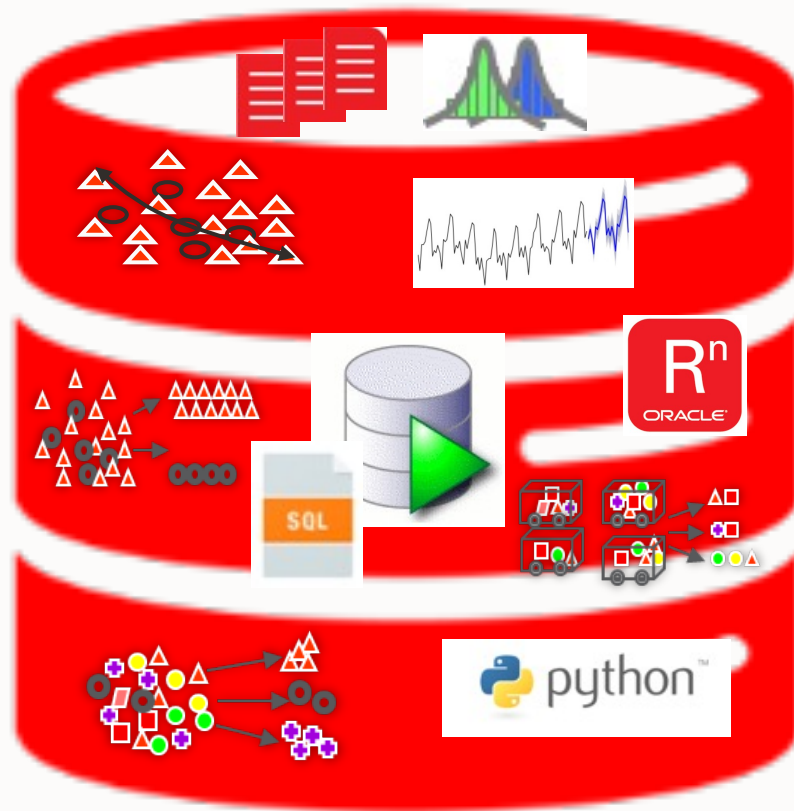


Hit Send to get a response



Why Oracle for Machine Learning?

Move the Algorithms, Not the Data!



$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$



Helpful Links



ORACLE MACHINE LEARNING ON O.COM

<https://www.oracle.com/machine-learning>

OML TUTORIALS

OML LiveLab: https://apexapps.oracle.com/pls/apex/dbpm/r/livelabs/view-workshop?p180_id=560

OML4Py LiveLab: <https://apexapps.oracle.com/pls/apex/dbpm/r/livelabs/view-workshop?wid=786>

Interactive tour: <https://docs.oracle.com/en/cloud/paas/autonomous-database/oml-tour>

Picking a Good Wine Using ADW, OML and OAC Workshop: <https://go.oracle.com/LP=109257?elqCampaignId=293121>

OML OFFICE HOURS

<https://asktom.oracle.com/pls/apex/asktom.search?office=6801#sessions>

ORACLE ANALYTICS CLOUD

<https://www.oracle.com/solutions/business-analytics/data-visualization/examples.html>

OML4PY

[OML4Py](#) (2m video)

[OML4Py Introduction](#) (17m video)

[OML4Py Technical Brief](#)

[OML4Py User's Guide](#)

[Blog: Introducing OML4Py](#)

[GitHub Repository with Python notebooks](#)

ORACLE AUTOML UI

[Oracle Machine Learning AutoML UI](#) (2m video)

[Oracle Machine Learning Demonstration](#) (6m video)

[OML AutoML UI Technical Brief](#)

[Blog: Introducing Oracle Machine Learning AutoML UI](#)

OML SERVICES

[Oracle Machine Learning Services](#) (2m video)

[OML Services Technical Brief](#)

[Oracle Machine Learning Services Documentation](#)

[Blog: Introducing Oracle Machine Learning Services](#)

[GitHub Repository with OML Services examples](#)



<https://www.oracle.com/cloud/free/>

Thank you

renee.wikestad@oracle.com