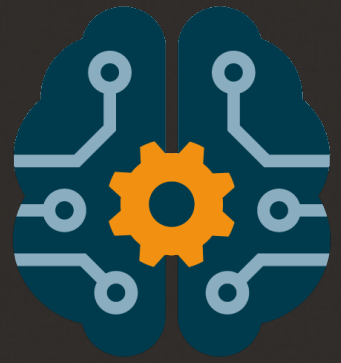
Oracle Machine Learning



Introduction

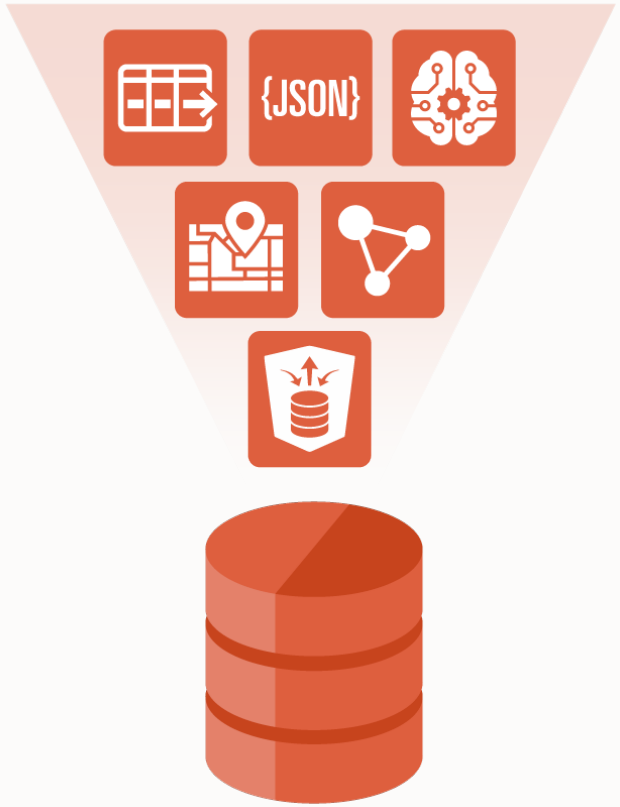
Move the Algorithms – Not the Data

Renée Wikestad
Cloud Specialist Engineer,
Oracle Norge

April 2021

Building on a converged database

In cloud and on premises – integrated, not fragmented



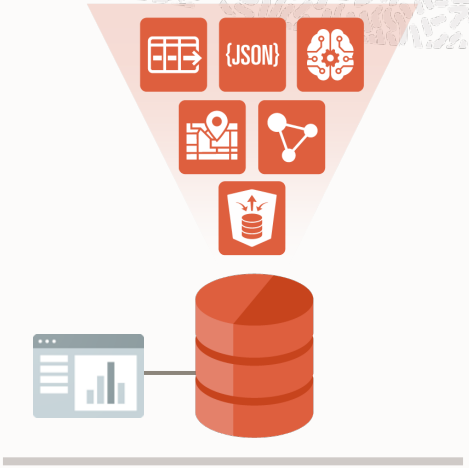
Supports any Data
Relational, JSON, graph, spatial, text, OLAP, XML, multimedia

Supports any Workload
Transactions, analytics, ML, IoT, streaming, blockchain

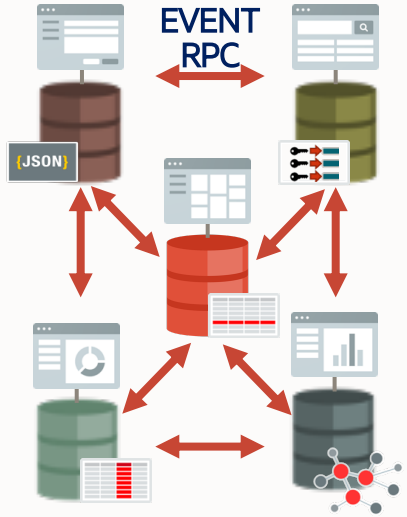
Most Productive for Developers and Analysts
Integrated microservices, events, REST, SaaS, ML, CI/CD, Low-Code

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

ORACLE Fusion Applications · ORACLE Siebel · ORACLE PeopleSoft · Oracle EBS · Oracle NetSuite · SAP · salesforce

Databases

NoSQL

Files

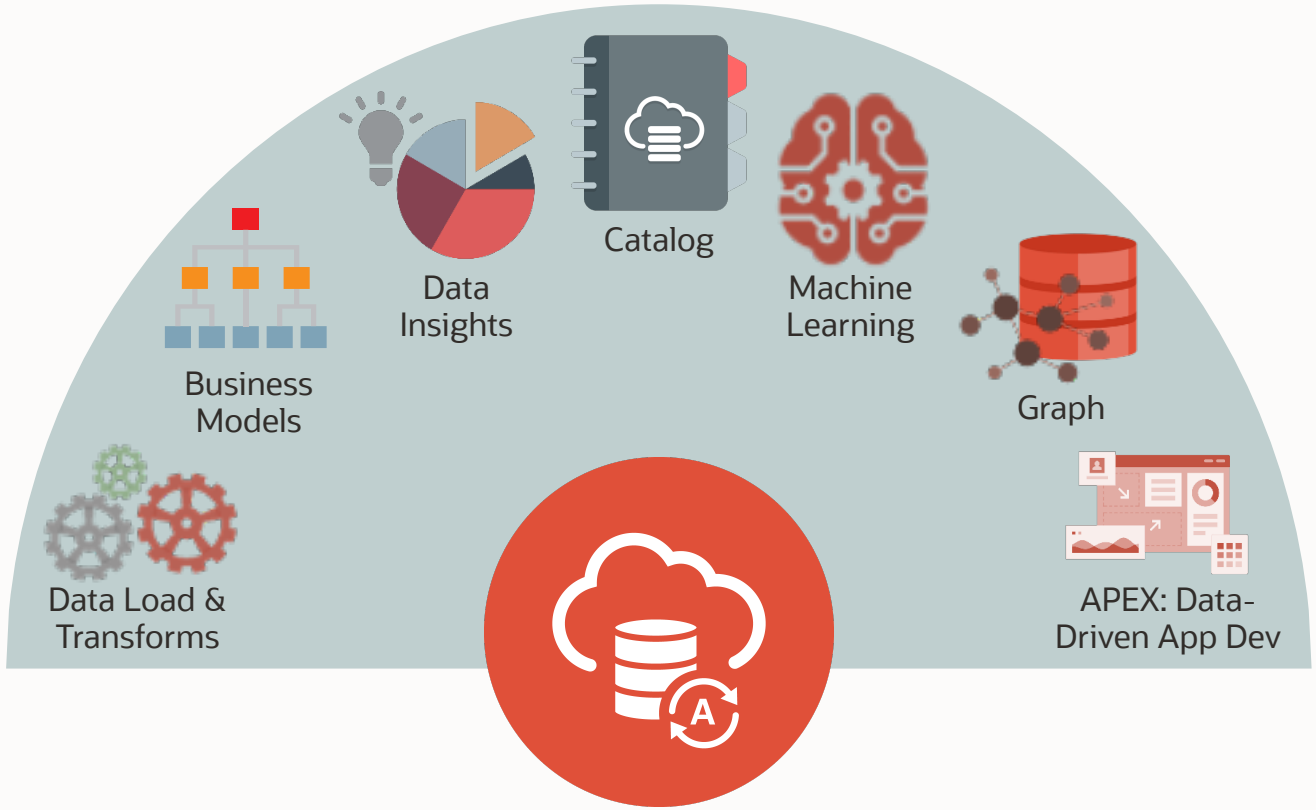
{...} · CSV

Streaming Services

kafka

Data Lake

OCI Object Storage · Amazon S3 · Microsoft Azure Blob Storage



Visualization & Development

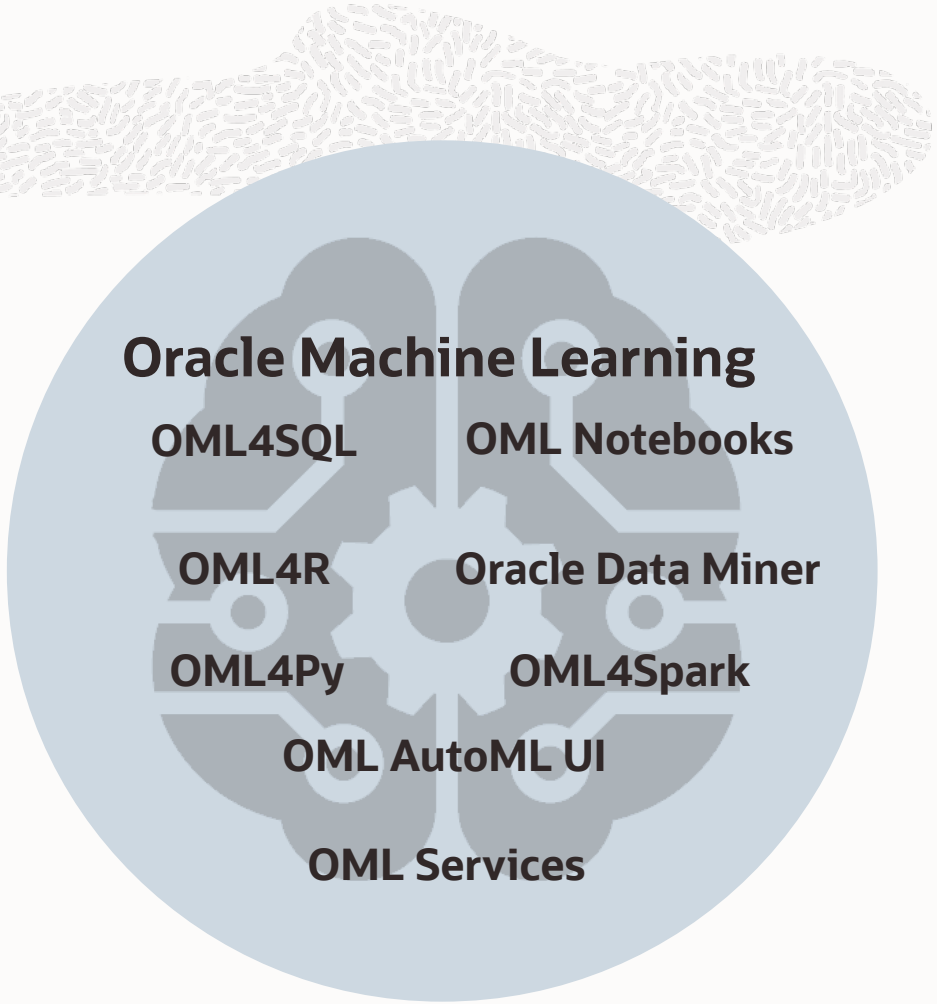
- Application Express
- Oracle ML Notebook
- Oracle Analytics Cloud
- IDEs
- Spreadsheets
- 3rd Party BI Tools

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 & Al

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

** New in 21c*

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;
```

SQL | All Rows Fetched: 1 in 0.043 seconds

	PREDICTION_PROBABILITY(BUY_INSUR,1,'YES'USING3500ASBANK_FUNDS,825ASCHECKING_AMOUNT,400ASCREDIT_BALANCE
1	0.9276956709910801

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

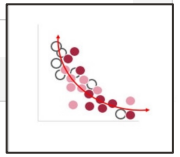
This notebook shows how to predict numerical values using Support Vector Machine (SVM) Regression. Given demographic information when renting bikes to predict the volume of bikes rented out in column Volume.

– 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:  Xnom=(X-Xmin)/(Xmax-Xmin)  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

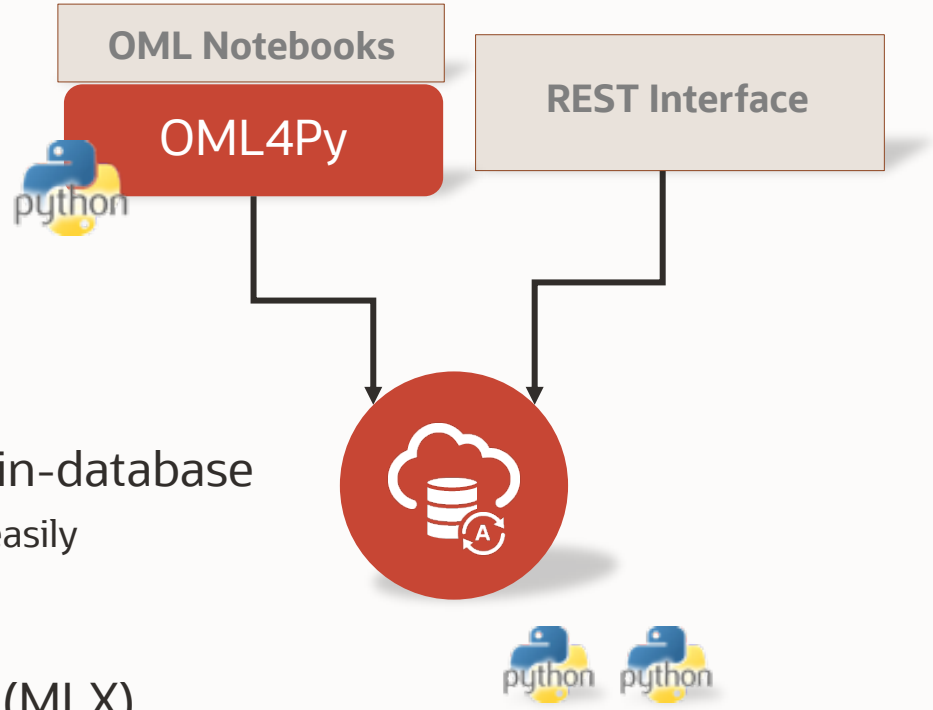
Supported in Oracle Autonomous Database with OML Notebooks

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

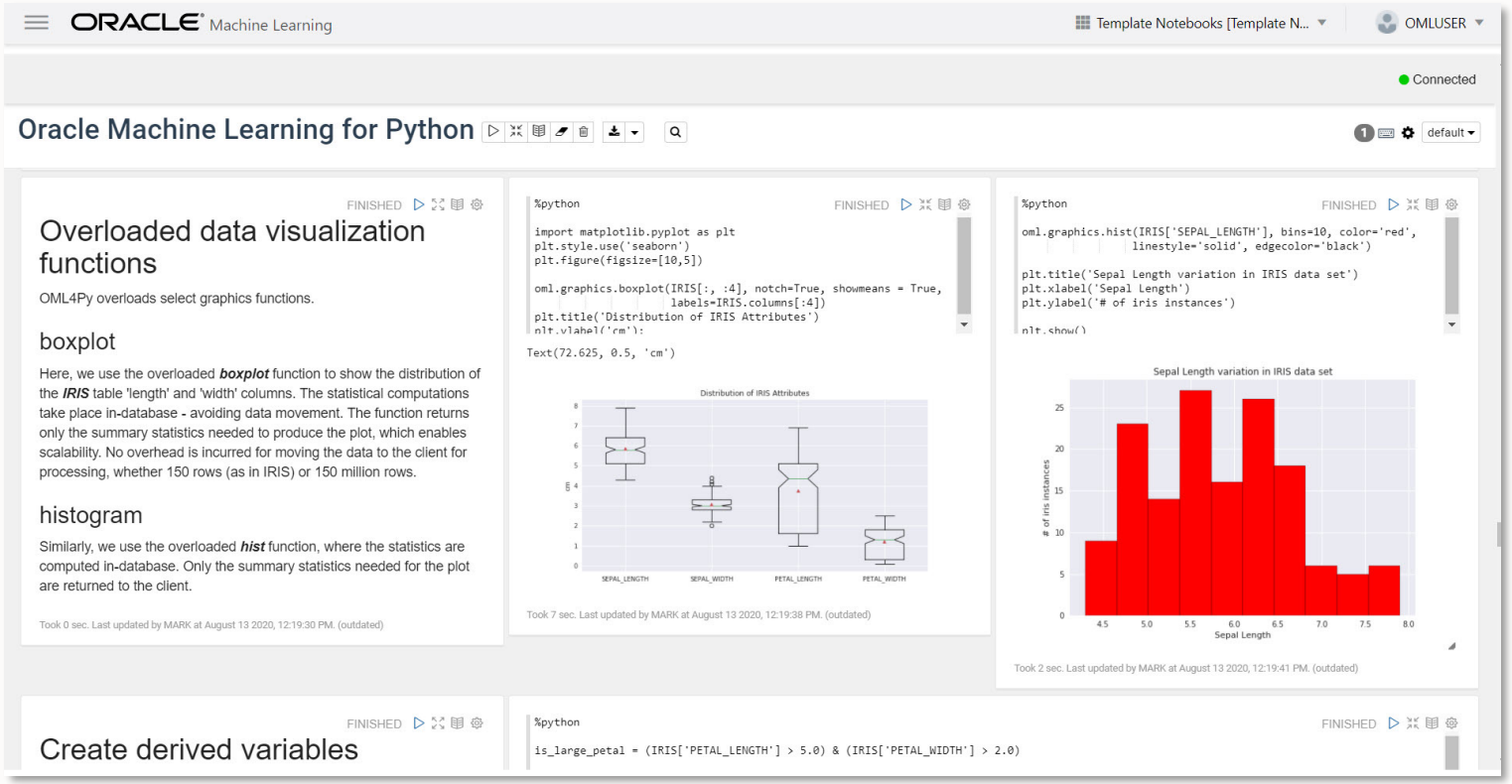
Autonomous Database as a Data Science Platform

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



Transparency Layer

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

Leverages proxy objects for database data: oml.DataFrame

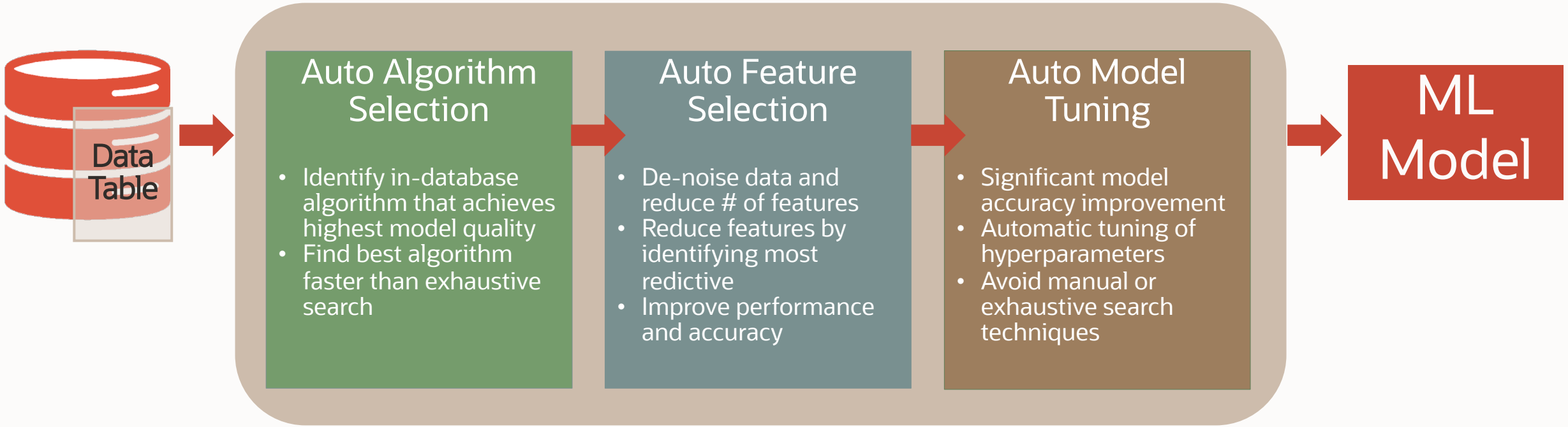
```
# Create table from Pandas DataFrame data
DATA = oml.create(data, table = 'CUSTOMER')

# 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

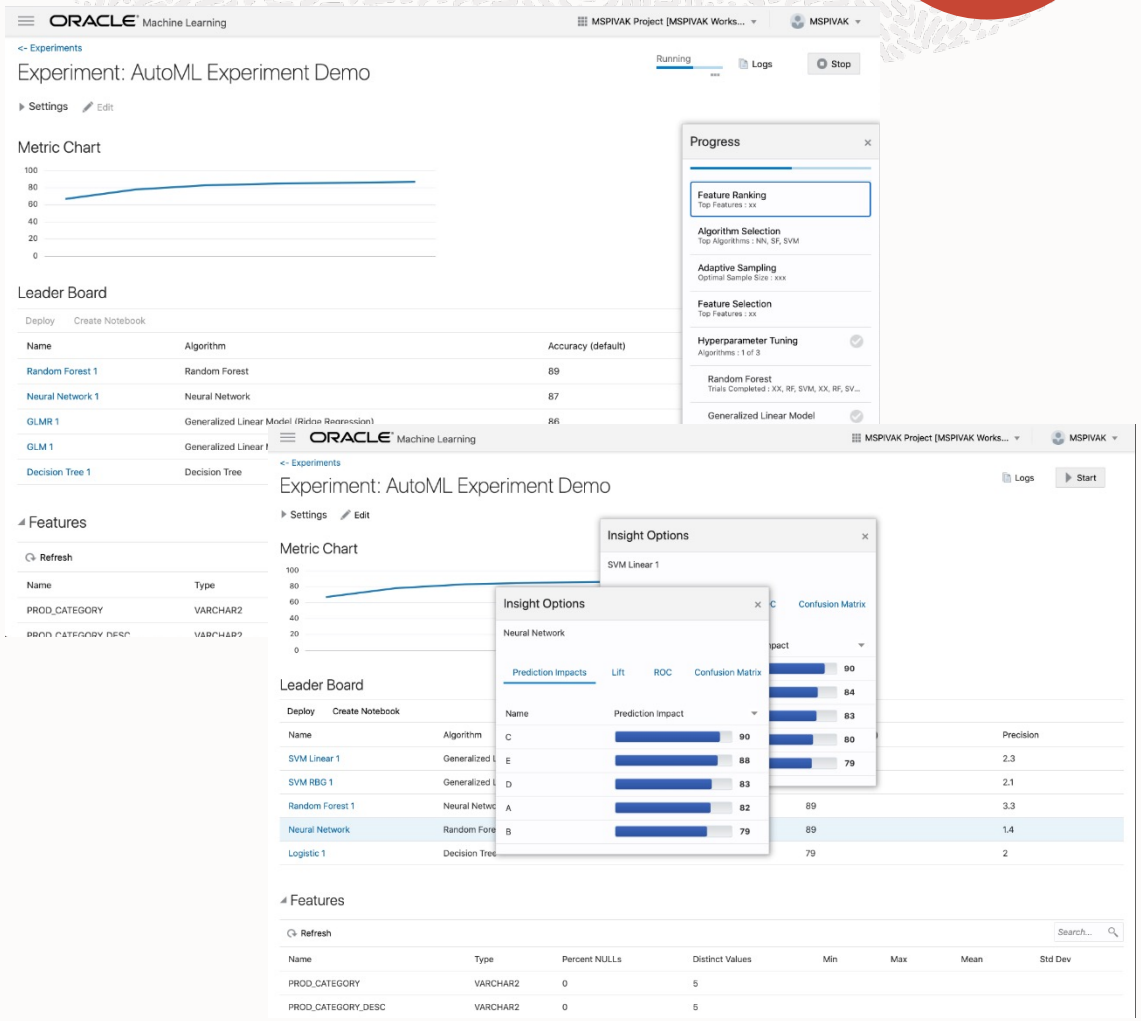
- Enable non-expert users to use ML

Automates model building, tuning, and deployment

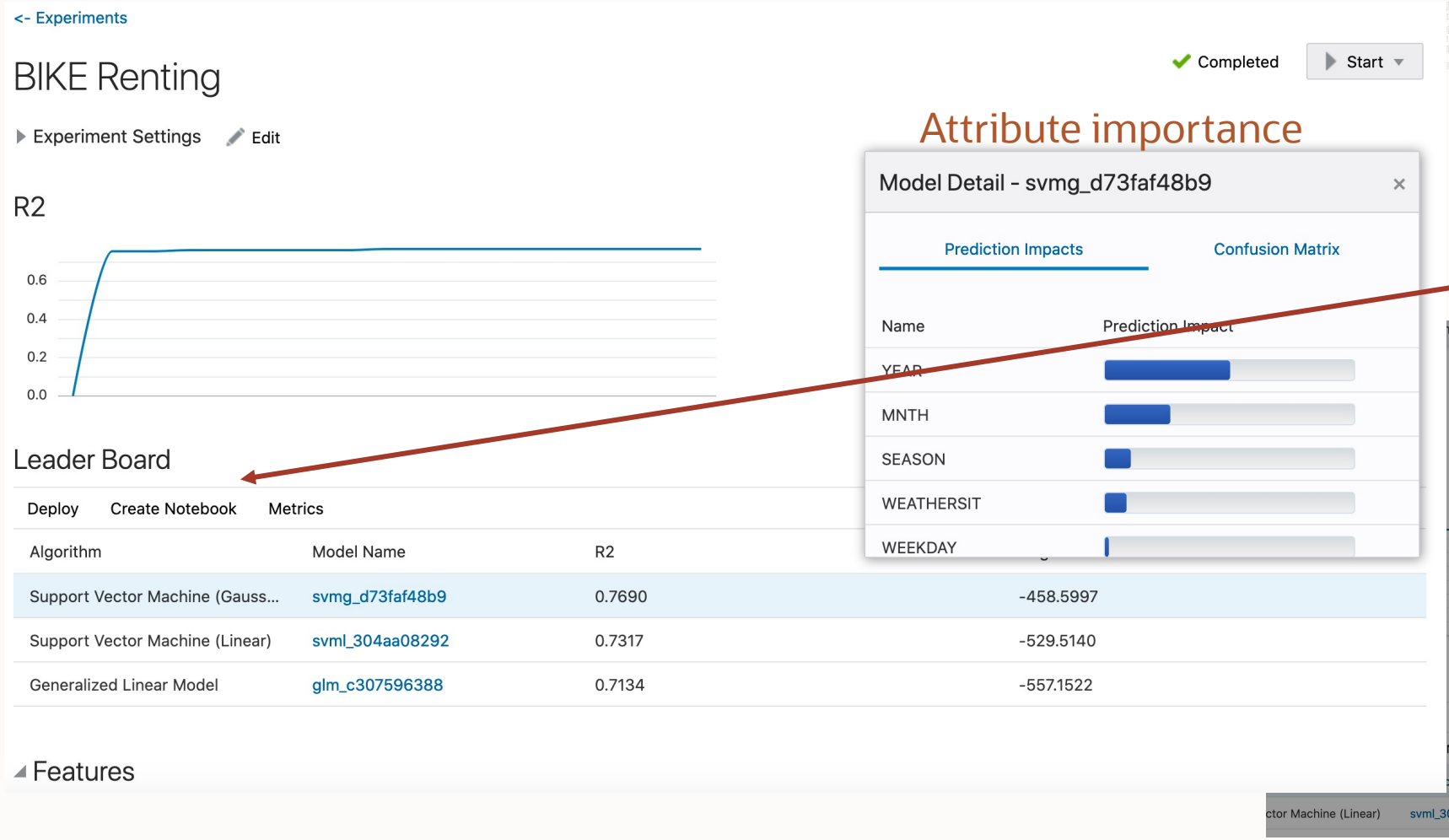
- Enhance data scientist productivity
- Support model management
- Empower non-expert users

Featuring

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

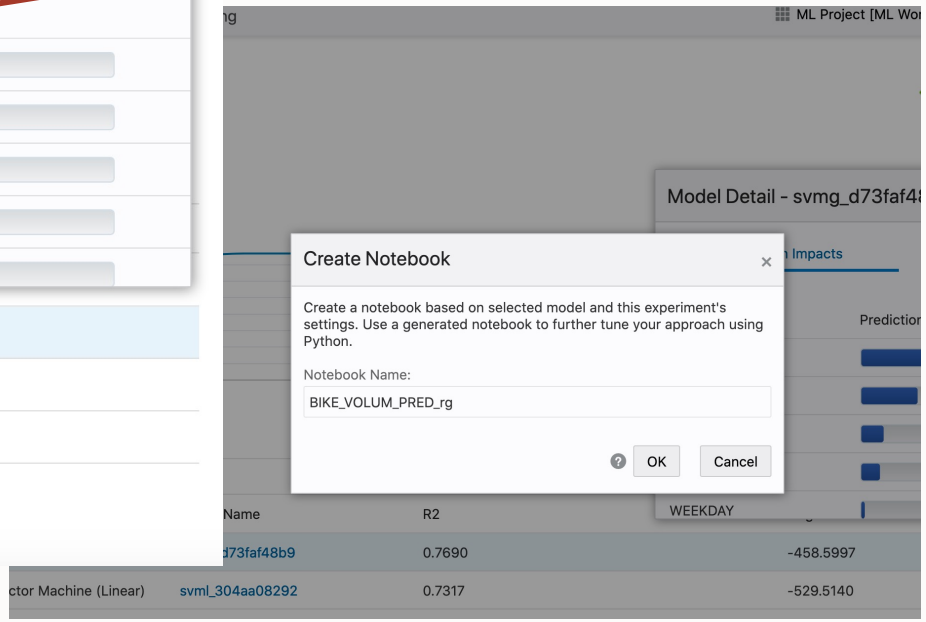


Create Notebook from OML AutoML UI



Attribute importance

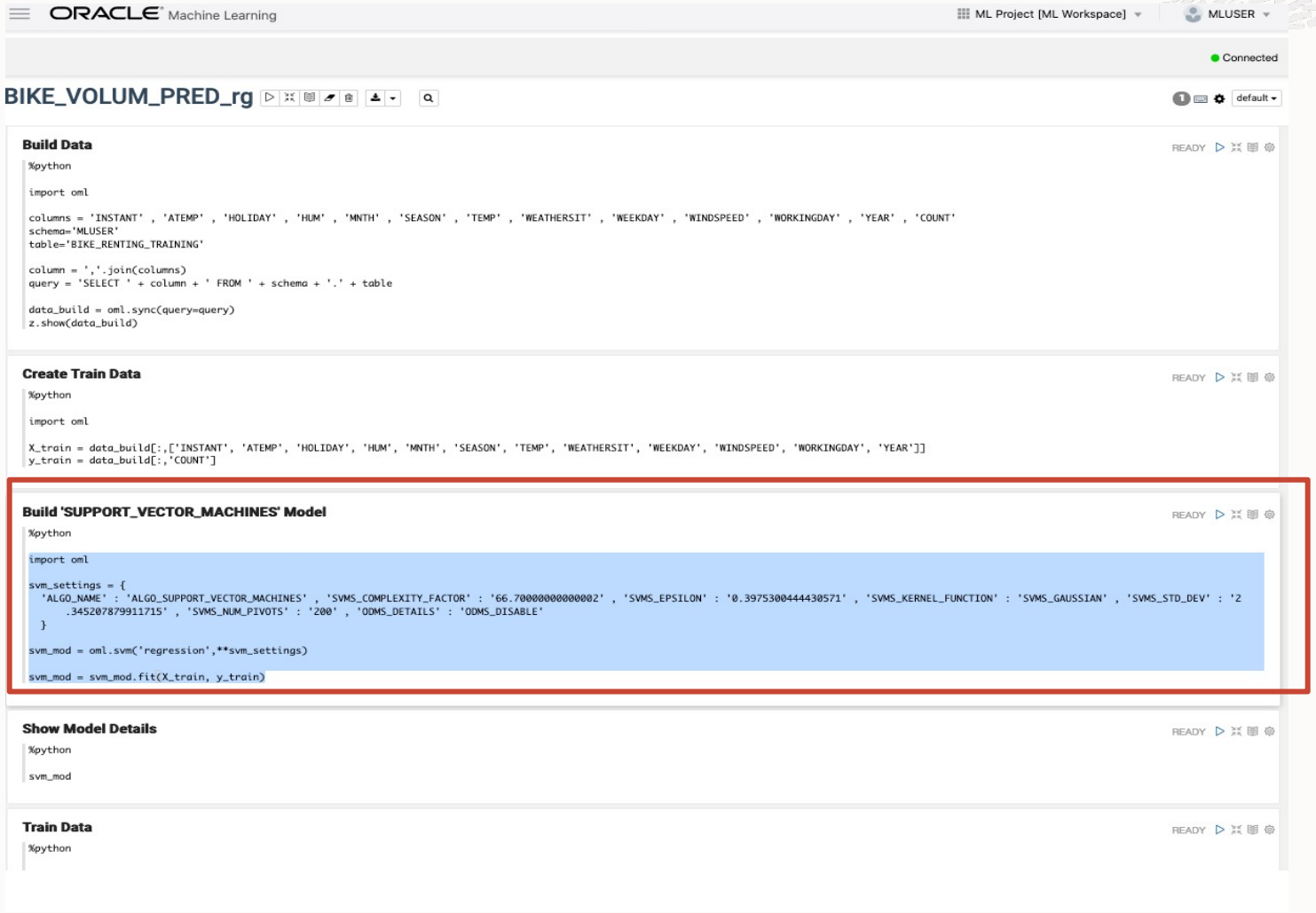
Create Notebook for selected model



OML AutoML UI

Oracle Autonomous Database and OML Notebooks

Generated Notebook



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

{REST:API}

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

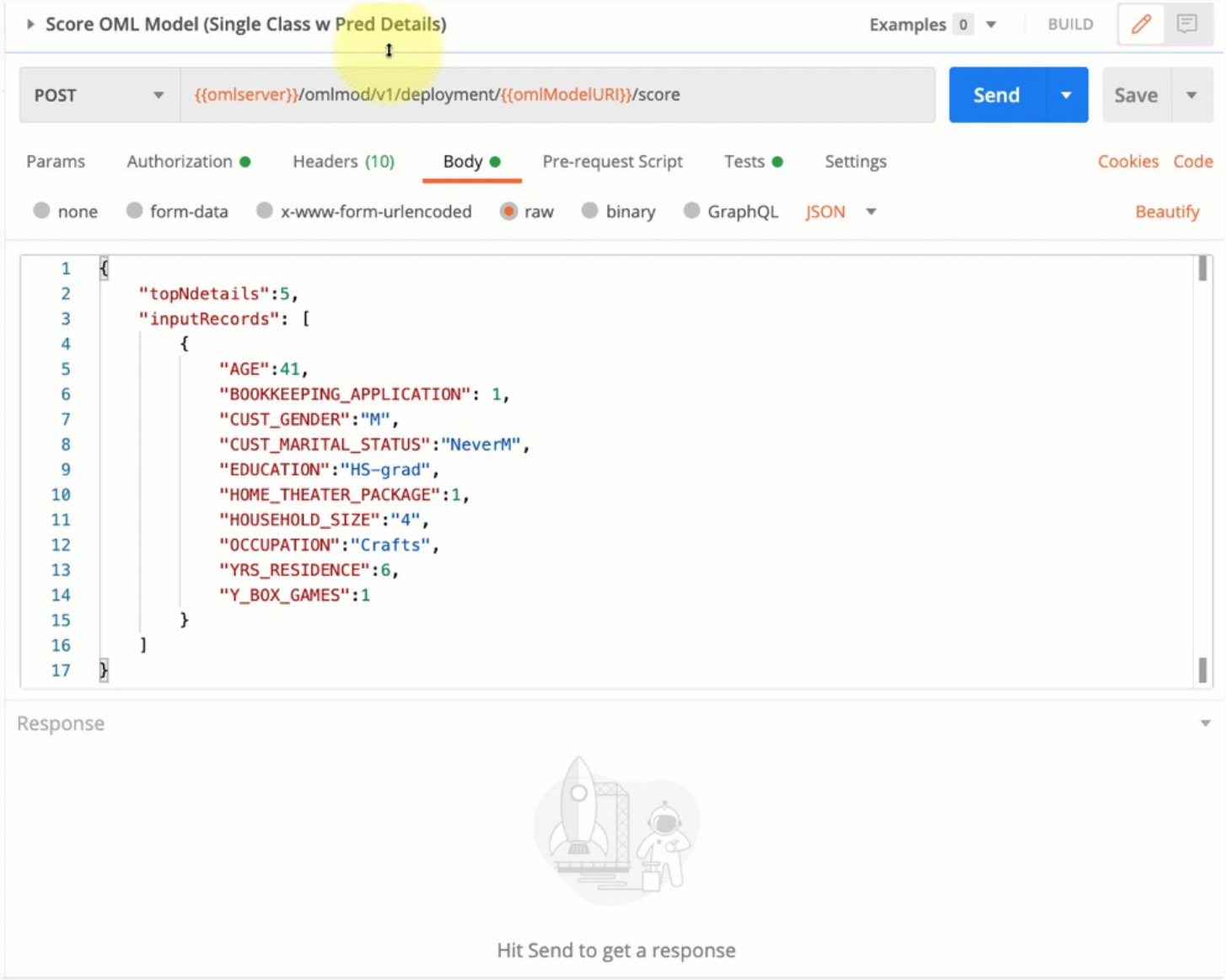
Cognitive Text

POST

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

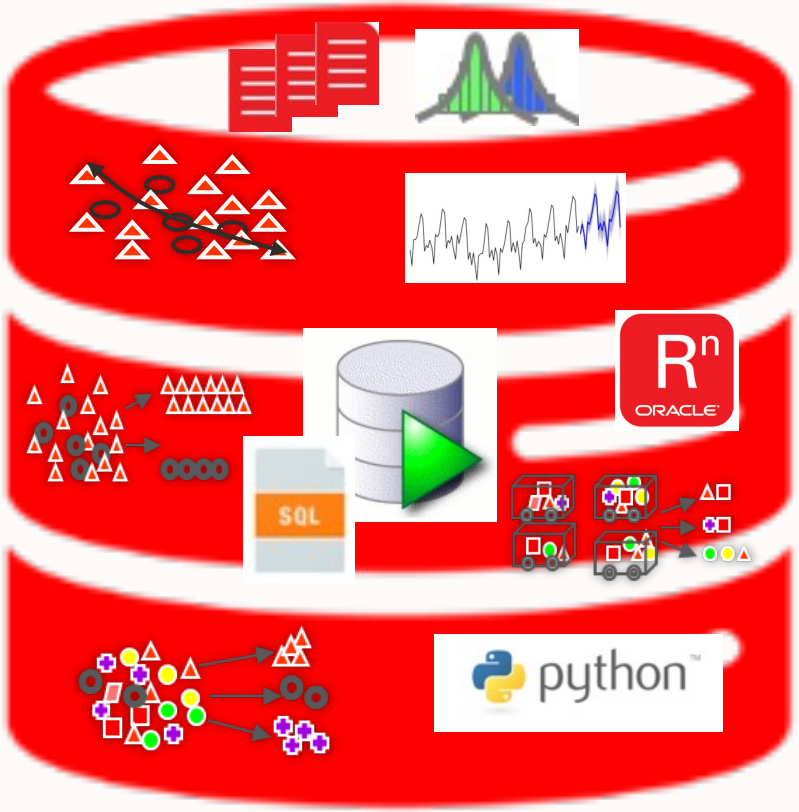
GET

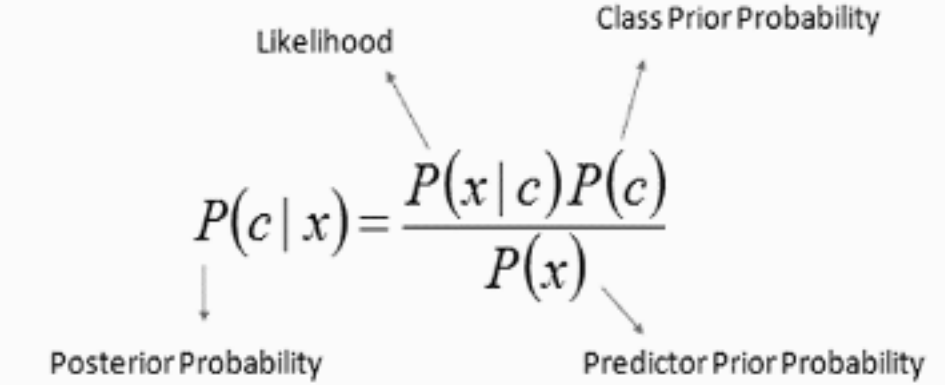
- Get Endpoints



Why Oracle for Machine Learning?

Move the Algorithms, Not the Data!





$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

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#sessionss

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