

ORACLE

Move the Algorithms; Not the Data!

Oracle Machine Learning

What's New?

Charlie Berger

Sr. Director Product Management, Machine Learning, AI and Cognitive Analytics

 charlie.berger@oracle.com  [@CharlieDataMine](https://twitter.com/CharlieDataMine)

September, 2020

Our mission is to help people
see data in new ways, discover insights,
unlock endless possibilities.



Oracle Database is a Converged Database

Key Benefits for Developers, Analysts, and Ops

- The richness of Data Driven Apps can make them ultra **complex** to implement
- By integrating the new data types and paradigms, Oracle makes it **dramatically easier and faster** to develop Data Driven Apps
 - Open Standard SQL across all data plus popular APIs: REST
 - Transactions across all data
 - Consistent, Queryable View of all data
 - Common Reliability, Scalability, Security
 - Common Management
 - **Harvest more information, discover new insights and make predictions**



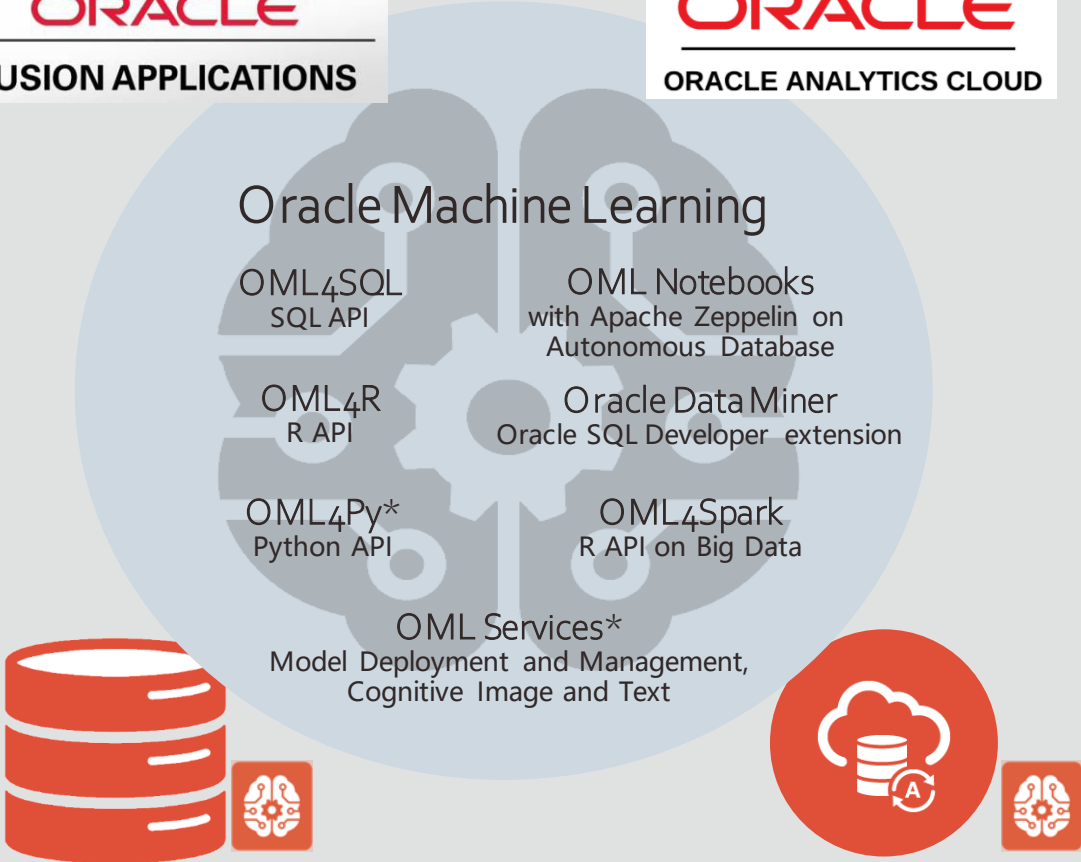
Oracle Machine Learning

Oracle Machine Learning *extends* Oracle Database(s) and enables users to build "AI" applications and analytics dashboards

OML delivers powerful in-database machine learning algorithms, automated ML functionality via SQL APIs and integration with open source Python* and R

ORACLE®
FUSION APPLICATIONS

ORACLE®
ORACLE ANALYTICS CLOUD

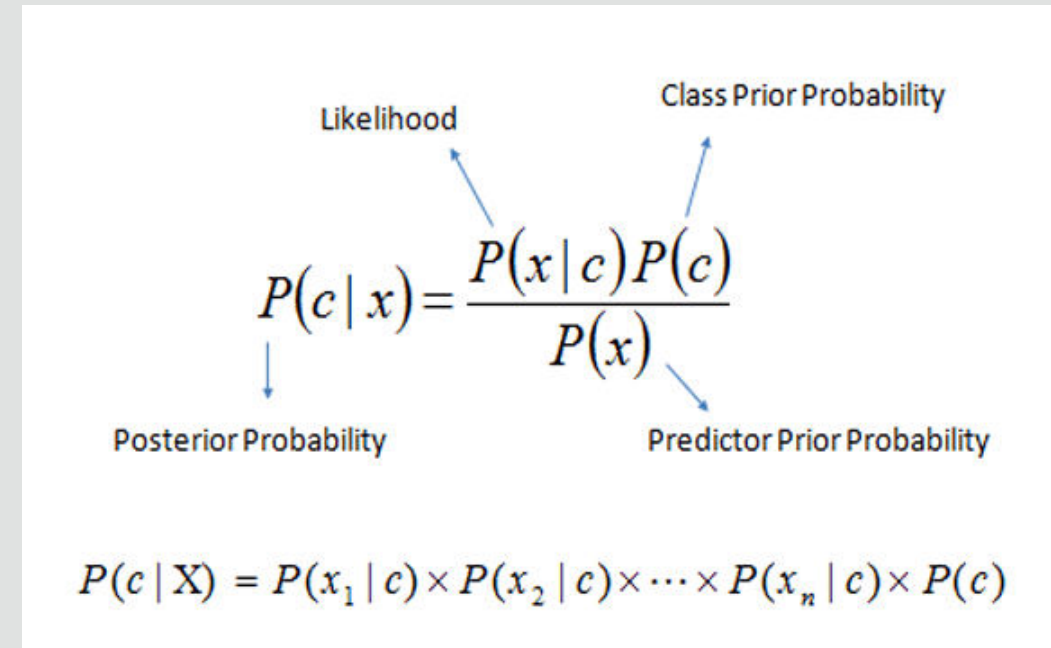
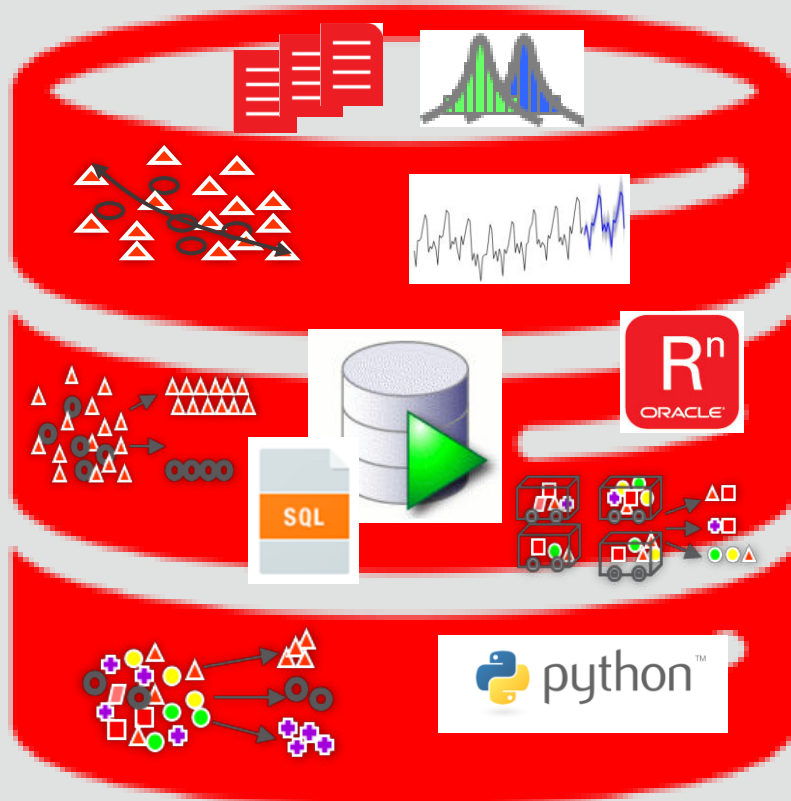


Oracle Makes Machine Learning Simple



Move the Algorithms, Not the Data!

In-Database Processing



An "AI Database"? A "Thinking Database"? It Changes *Everything!*

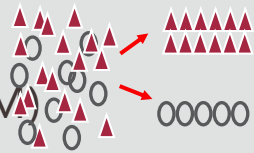


Oracle Machine Learning Algorithms



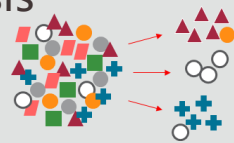
CLASSIFICATION

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



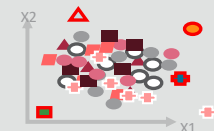
CLUSTERING

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



ANOMALY DETECTION

One-Class SVM
MSET-SPRT*



TIME SERIES

Forecasting - Exponential Smoothing

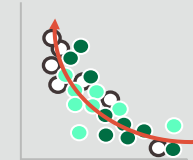
Includes popular models

e.g. Holt-Winters with trends, seasonality, irregularity, missing data



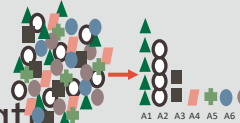
REGRESSION

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



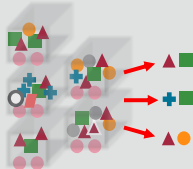
ATTRIBUTE IMPORTANCE

Minimum Description Length
Principal Comp Analysis (PCA)
Unsupervised Pair-wise KL Div
CUR decomposition for row & AI



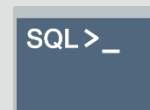
ASSOCIATION RULES

A priori/ market basket



PREDICTIVE QUERIES

Predict, cluster, detect, features



SQL ANALYTICS

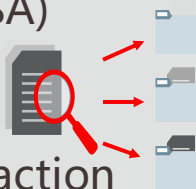
SQL Windows
SQL Patterns
SQL Aggregates

FEATURE EXTRACTION

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

TEXT MINING SUPPORT

Algorithms support text Tokenization and theme extraction
Explicit Semantic Analysis (ESA) for document similarity



STATISTICAL FUNCTIONS

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



R & PYTHON * Coming this Fall

Third-party R & Python Packages through Embedded Execution
Spark MLib algorithm integration



MODEL DEPLOYMENT & MONITORING

SQL—1st Class **Objects** * Coming this Fall
Oracle RESTful API (ORDS)
OML Services

Includes support for Partitioned Models, Transactional data and aggregations, Unstructured data, Geo-spatial data, Graph data. etc.

* New in 20c



Oracle Machine Learning

Oracle Machine Learning Notebooks included in Autonomous Databases



Key Features:

- Collaborative UI for data scientists & analysts
- Packaged with Autonomous Databases
- Easy access to shared notebooks, templates, permissions, scheduler, etc.
- Example notebooks
- OML4SQL
- OML4Py *coming this Fall*

The screenshot displays the Oracle Machine Learning interface with several notebooks open. The top notebook, 'Targeting Customers Who Buy Insurance', contains a table of customer data and a bar chart. The table data is as follows:

CUST_ID	LAST	FIRST	STATE
CU7115	TYLER	MARSH	MN
CU7117	CARISSA	POLLOCK	MI
CU7118	BUCK	DOE	CA
CU7137	WALLACE	CHRISTOPHER	NV
CU13368	BOYCE	LEMONS	NY
CU13378	LOWELL	BILLINGS	OR
CU9696	CHARIS	MCCLELLAN	CA
CU4366	MARCENE	BARTON	NV

Other notebooks visible include 'Build a Random Forest model to predicts who will purchase insurance' with SQL code, and 'Compare the models' with a performance graph. The interface also shows a 'Let's explore the CUST_INSUR_LTV data' notebook with a scatter plot and another with a person walking on a tightrope illustration.



OML to Predict Customer Behavior

Intuitive SQL API—OML4SQL

-- Build and train ML model to determine which **customers will 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;
```



OML to Predict Customer Behavior

Intuitive SQL API—OML4SQL

-- Build and train ML model to determine which **customers will 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;
```

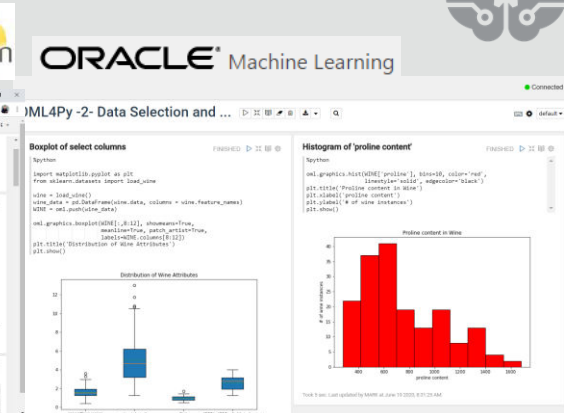
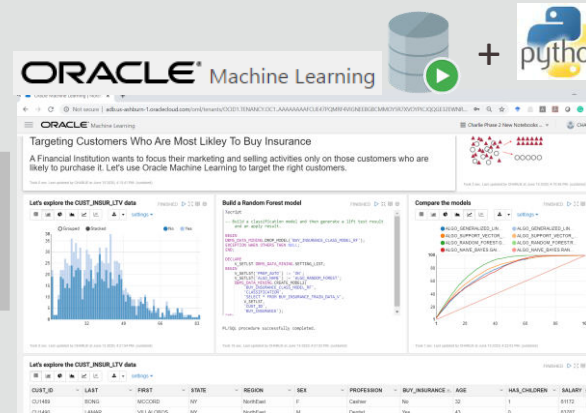
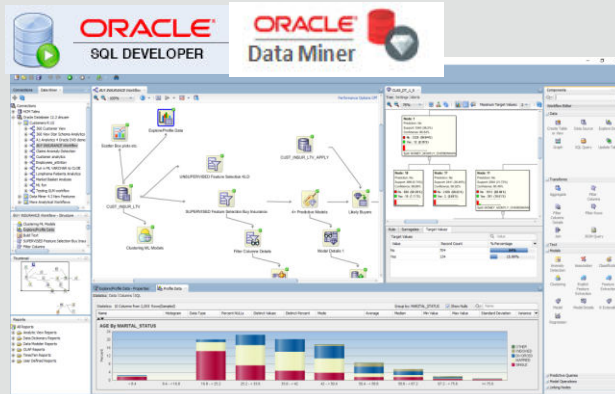
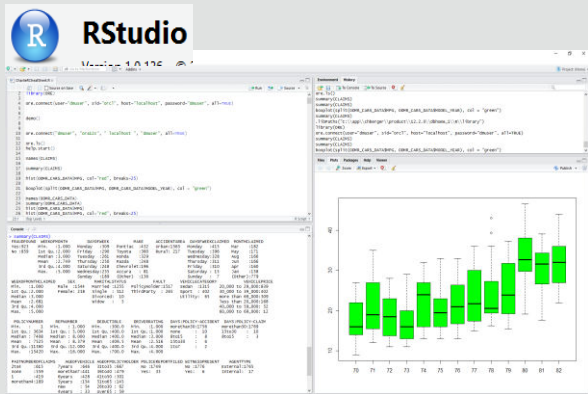


-- Apply ML model to sales to predict which **customers are likely to buy**

```
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_INSUR1,'YES'USING3500ASBANK_FUNDS,825ASCHECKING_AMOUNT,400ASCREDIT_BALANCE	
1	0.9276956709910801

Goal: Manage and Analyze All Your Data



SQL / R / Python

Boil down the Data Lake

Big Data SQL



“Engineered Features”

– Derived attributes that reflect domain knowledge—key to best models e.g.:

- Counts
- Totals
- Changes over time



Coming Soon!

Oracle Machine Learning



What's Coming....

Autonomous Database

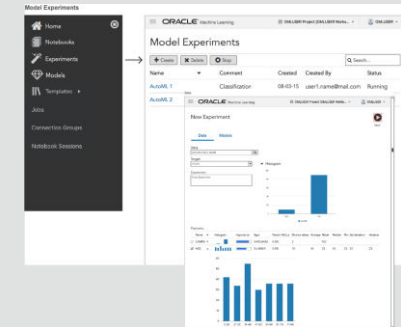
Oracle Machine Learning for Python (OML4Py) – *Target Fall 2020*



- OML4Py, Embedded Python execution, Python & REST APIs, SQL API afterwards
- AutoML functionality
- Support for 3rd party Python packages - *H2 FY21*

AutoML UI – *Target Fall 2020*

- Code-free interface with OML Notebooks
- Time series, anomaly detection – *Target H2 FY21*



OML Services – *Target Fall 2020*

- REST API for model management, deployment, ONNX and cognitive text



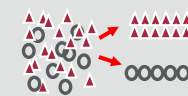
Oracle Machine Learning for R (OML4R) – *Target CY2020*

- OML4R, Embedded R execution will also support a REST API followed by a SQL API



Oracle Database 20c+ DBCS

20c: New Algorithms: XGBoost, MSET-SPRT



OML4Py – *Target Fall 2020* (18c >)

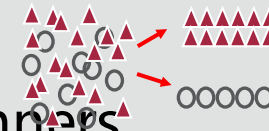


Coming in 20c | New OML Algorithms

New Algorithms

Gradient Boosted Trees (XGBoost)

Highly popular and powerful algorithm – Kaggle winners
Classification, regression, ranking, survival analysis



MSET-SPRT

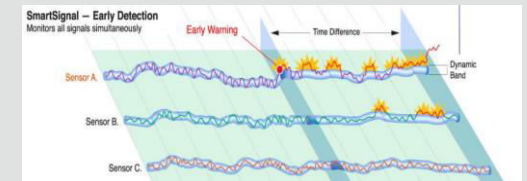
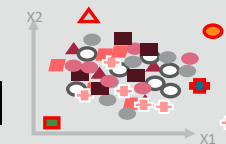
Multivariate State Estimation Technique - Sequential Probability Ratio Test (MSET-SPRT)

Nonlinear, nonparametric anomaly detection algorithm designed to monitor critical processes.

Detects subtle anomalies while also producing minimal false alarms

Calibrates expected behavior from historical normal operational sequence of monitored signals

Re-implemented and sped up in-DB and based on original Oracle Labs algorithm



Oracle Machine Learning for R / Python * Coming soon

Integration with Open Source ML

Transparency layer

- Leverage proxy objects so data remain in database
- Overload native functions translating functionality to
- Use familiar R/Python syntax to manipulate database

Parallel, distributed algorithms

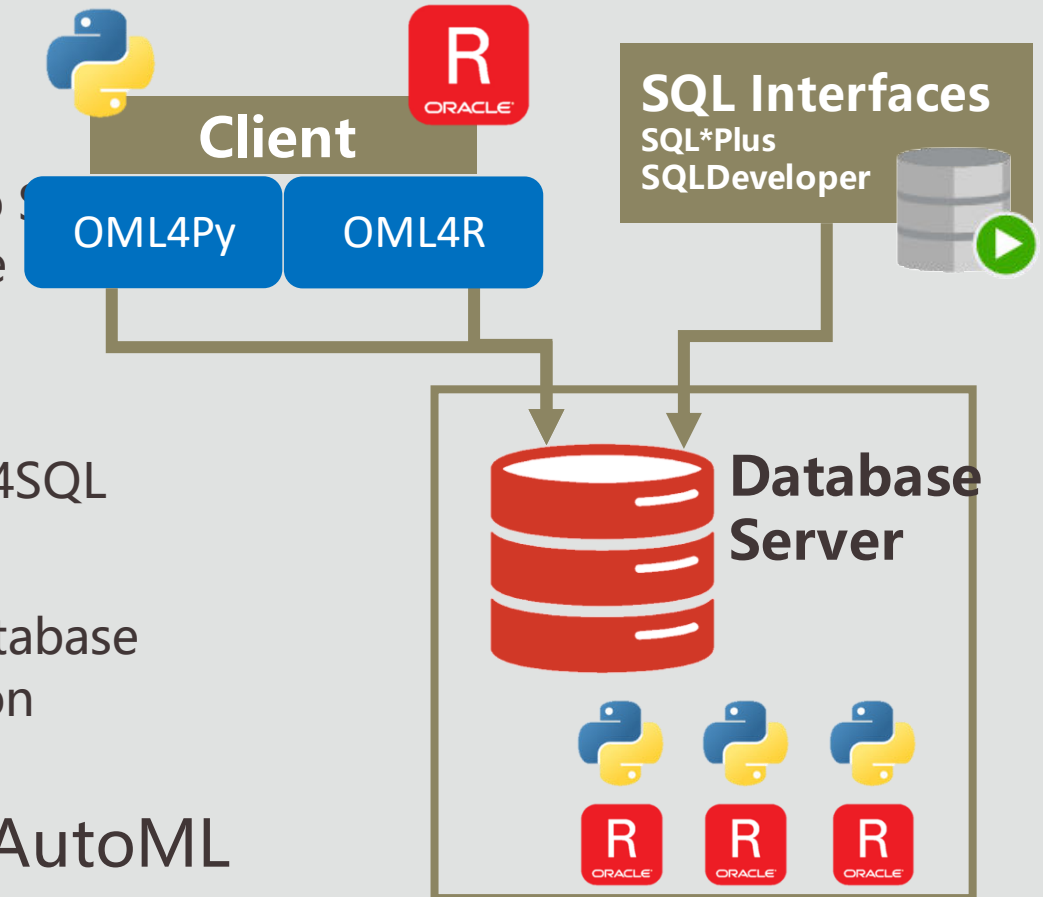
- Scalability and performance
- Exposes in-database algorithms available from OML4SQL

Embedded execution

- Manage and invoke R or Python scripts in Oracle Database
- Data-parallel, task-parallel, and non-parallel execution
- Use open source packages to augment functionality

OML4Py, Automated Machine Learning - AutoML

- Algorithm selection, feature selection, auto tune hyperparameters

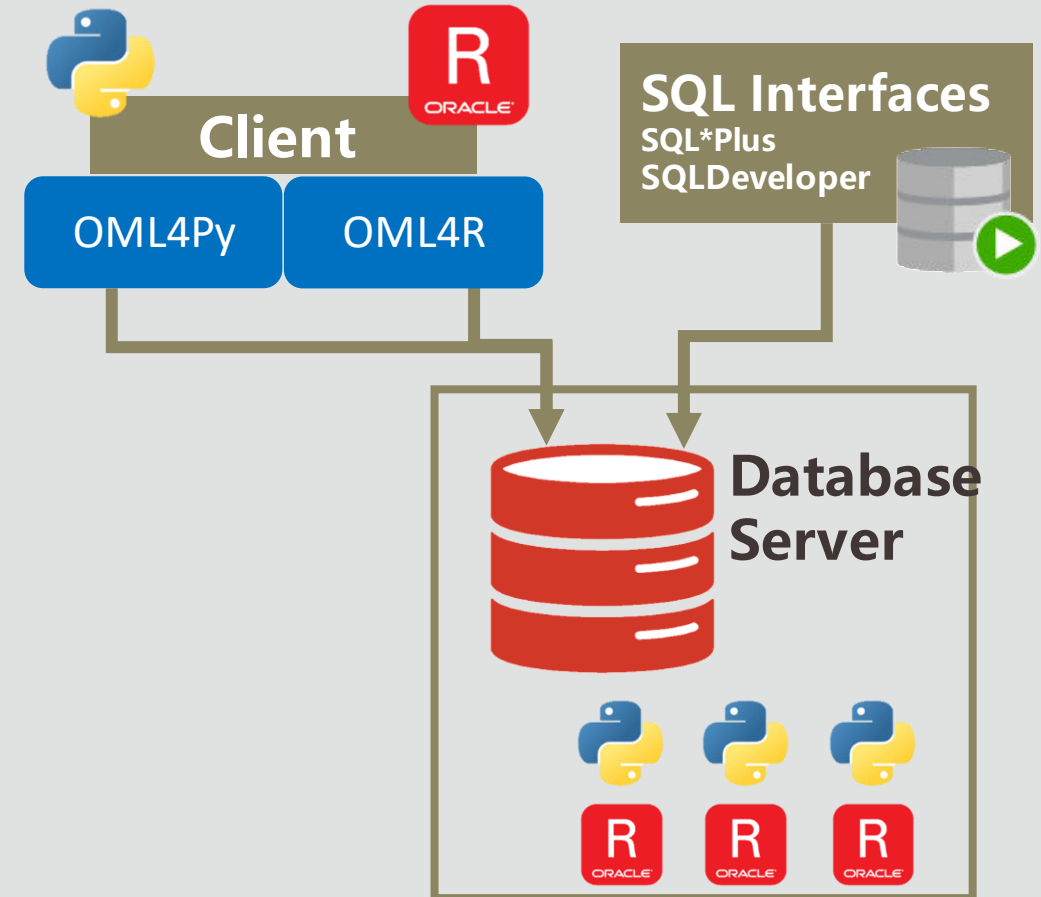


Oracle Machine Learning for R / Python * Coming soon

Preinstalled and supported Python Packages

OML4Py details and supported packages

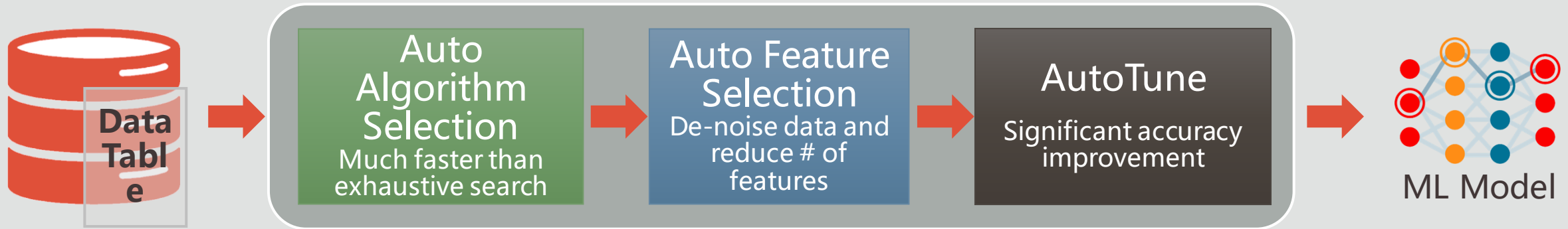
- Python 3.8.2
- cx_Oracle 7.3.0
- cyclcr 0.10.0
- joblib 0.14.0
- kiwisolver 1.1.0
- matplotlib 3.1.2
- numpy 1.18.1
- pandas 0.25.3
- pyparsing 2.4.0
- python_dateutil 2.8.1
- pytz 2019.3
- scikit-learn 0.22.1
- scipy 1.4.1
- six 1.13.0
- pyreadline 2.1* - windows only
- setuptools 41.2.0





Coming Soon! | AutoML – *new* with OML4Py

Increase data scientist productivity – reduce overall compute time



Auto Algorithm Selection

- Identify in-database algorithm that achieves highest model quality
- Find best algorithm faster than with exhaustive search

Auto Feature Selection

- Reduce # of features by identifying most predictive
- Improve performance and accuracy

Auto Tune Hyperparameters

- Significantly improve model accuracy
- Avoid manual or exhaustive search techniques

Enables non-expert users to leverage Machine Learning

Oracle Machine Learning

Oracle Machine Learning Notebooks included in Autonomous Databases



Key Features:

- Collaborative UI for data scientists & analysts
- Packaged with Autonomous Databases
- Easy access to shared notebooks, templates, permissions, scheduler, etc.
- Example notebooks
- OML4SQL
- **OML4Py** coming this Fall

Compute pairwise correlation
Using only numeric columns, we can compute a correlation matrix on the proxy object for the database table IRIS using the overloaded `corr` function. Here we see that petal length and width are highly correlated.

```
Python
corr_res = IRIS.corr()
corr_res
```

	SEPAL_LENGTH	SEPAL_WIDTH	PETAL_LENGTH	PETAL_WIDTH
SEPAL_LENGTH	1.000000	-0.117570	0.871754	0.817941
SEPAL_WIDTH	-0.117570	1.000000	-0.428440	-0.366126
PETAL_LENGTH	0.871754	-0.428440	1.000000	0.962865
PETAL_WIDTH	0.817941	-0.366126	0.962865	1.000000

Overloaded data visualization functions
OML4Py overloads select graphics functions.

boxplot
Here, we use the overloaded `boxplot` function to show the distribution of the `IRIS` table 'length' and 'width' columns. The statistical computations take place in-database - avoiding data movement. The function returns only the summary statistics needed to produce the plot, which enables scalability. No overhead is incurred for moving the data to the client for processing, whether 150 rows (as in `IRIS`) or 150 million rows.

histogram
Similarly, we use the overloaded `hist` function, where the statistics are computed in-database. Only the summary statistics needed for the plot are returned to the client.

Boxplot: Distribution of IRIS Attributes

Histogram: Sepal Length variation in IRIS data set



Oracle Machine Learning

OML4Py Python Integration Added to OML on Autonomous Database



ORACLE Machine Learning | Template Notebooks [Template N... | CHARLIE | Connected

Targeting Customers OML4Py Python

```
%python
z.show(DEMO_DF.head())
```

CUST_ID	AFFINITY_CARD	BOOKKEEPING_APPLICATION...	BULK_PACK_DISKETTE...	EDUCATION	FLAT_PANEL_MONITO...	HOME_THEATER_PACKA...	HOUSEHOL
102547	0	0	1	10th	1	0	1
101050	0	0	1	10th	1	0	1
100040	0	0	1	11th	1	0	1

Took 0 sec. Last updated by OMLUSER at August 20 2020, 4:31:31 PM. (outdated)

Graph HOUSEHOLD_SIZE grouped by AFFINITY_CARD responders

```
%python
z.show(DEMO.crosstab(['HOUSEHOLD_SIZE', 'AFFINITY_CARD']))
```

Build a Decision Tree Model

```
%python
setting = dict()
dt_mod = oml.dt(**setting)
dt_mod.fit(TRAIN_X, TRAIN_Y, case_id = 'CUST_ID')
```

Algorithm Name: Decision Tree
Mining Function: CLASSIFICATION
Target: AFFINITY_CARD

Settings:

setting name	setting value
0 ALGO_NAME	ALGO_DECISION_TREE
1 CLAS_MAX_SUP_BINS	32
2 CLAS_WEIGHTS_BALANCED	OFF
3 ODMS_DETAILS	ODMS_ENABLE
4 ODMS_MISSING_VALUE_TREATMENT	ODMS_MISSING_VALUE_AUTO
5 ODMS_SAMPLING	ODMS_SAMPLING_DISABLE
6 PREP_AUTO	ON
7 TREE_IMPURITY_METRIC	TREE_IMPURITY_GINI
8 TREE_DEPTH_MAX_DEPTH	7

Took 2 sec. Last updated by OMLUSER at August 20 2020, 4:31:35 PM. (outdated)

Display confusion matrix, lift chart and ROC curve

	PREDICTED 0	PREDICTED 1
ACTUAL 0	True Negative: 1288 (72.56%)	False Positive: 52 (2.93%)
ACTUAL 1	False Negative: 259 (14.59%)	True Positive: 176 (9.92%)
Accuracy: 82.4789%		

Model Evaluation for Decision Tree

- Algorithm: Decision Tree
- Algorithm_setting:
 - TP: 1288
 - FP: 52
 - FN: 259
 - TPR: 0.4046
 - FR: 0.0386
 - THR: 0.9812
 - NR: 0.5954
 - Precision: 0.7719
 - Accuracy: 0.8248
 - NPV: 0.8326
 - DetectionRate: 0.8326
 - BalancedAccuracy: 0.6829
 - AUC: 0.8326
 - F1Score: 0.5309
 - MatthewsCorrCoeff: 0.4702

Took 1 min 22 sec. Last updated by OMLUSER at August 20 2020, 4:32:59 PM. (outdated)



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

Enables Citizen Data Scientists

Automate production and deployment of ML models

Enhance Data Scientist productivity and user-experience

Enable non-expert users to leverage ML

Unify model deployment and monitoring

Support model management

Features

Minimal user input: data, target

Model leaderboard

Model deployment via REST

Model monitoring

Cognitive features for image and text

The screenshot displays the Oracle Machine Learning AutoML user interface. The main section is titled 'Create Experiment' and includes a 'Start' button. The form contains the following fields:

- Name: Targeting Customers
- Comments: (empty)
- Data Source: (empty)
- Prediction Type: (empty)
- Predict: (empty)
- Case ID: (empty)
- Case ID: (empty)

Below the form is a 'Features' table with the following data:

Name	Importance	Type
OCCUPATION	High	VARCHAR2
YRS_RESIDENCE	High	NUMBER
BOOKKEEPING_APPLICATION	Low	NUMBER
COMMENTS	Low	VARCHAR2
AFFINITY_CARD	Low	NUMBER
HOME_THEATER_PACKAGE	Low	NUMBER
Y_BOX_GAMES	Low	NUMBER
FLAT_PANEL_MONITOR	Low	NUMBER
BULK_PACK_DISKETTES	Low	NUMBER
CLUST_ID	Low	NUMBER

To the right of the features table is a 'Metric Chart' showing a line graph with a blue line representing accuracy over time. Below the chart is a 'Leader Board' table with the following data:

Algorithm	Name	Accuracy
Naive Bayes	nb_f8637b9359	0.4282
Random Forest	rf_aed7a58923	0.3624
Decision Tree	dt_6cd748831e	0.3624



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

The screenshot displays the Oracle Machine Learning user interface. At the top, the Oracle logo and 'Machine Learning' text are on the left, and the project name 'MSPIVAK Project [MSPIVAK Works...]' and user 'MSPIVAK' are on the right. The main content area is divided into three sections: 'How Do I?', 'Quick Actions', and 'Recent Activities'. The 'How Do I?' section contains six informational cards: 'Use AutoML' (How to create AutoML Experiments), 'Get Started' (Get started with Oracle Machine Learning), 'Create Notebooks' (How to create a notebook), 'Create Jobs' (How to create a job), 'Manage Permissions' (How to manage collaborative permissions in workspaces), and 'Try It' (Follow along with a hands on workshop). The 'Quick Actions' section contains five cards: 'AutoML' (Create and run AutoML Experiments), 'Scratchpad' (Run Scratchpad), 'Notebooks' (The place for data discovery and analytics), 'Jobs' (Schedule notebooks to run at certain times), and 'Examples' (Check out some examples). A red arrow points from the 'AutoML' card to the 'AutoML' text in the 'Recent Activities' section. The 'Recent Activities' section is titled 'Recent Activities' and shows a list of activities under the heading 'last Monday'. The activities are: 'MSPIVAK created test1_2 notebook in MSPIVAK Project [MSPIVAK Workspace]' at 8/17/20 4:54 PM, 'MSPIVAK created test1_1 notebook in MSPIVAK Project [MSPIVAK Workspace]' at 8/17/20 1:16 PM, and 'MSPIVAK created test1 notebook in MSPIVAK Project [MSPIVAK Workspace]' at 8/17/20 1:15 PM. A large 'AutoML' text is overlaid on the 'Recent Activities' section, with a red arrow pointing to it from the 'AutoML' card in the 'Quick Actions' section.

ORACLE Machine Learning

MSPIVAK Project [MSPIVAK Works...]

MSPIVAK

How Do I?

- Use AutoML**
How to create AutoML Experiments
- Get Started**
Get started with Oracle Machine Learning
- Create Notebooks**
How to create a notebook
- Create Jobs**
How to create a job
- Manage Permissions**
How to manage collaborative permissions in workspaces
- Try It**
Follow along with a hands on workshop

Quick Actions

- AutoML**
Create and run AutoML Experiments
- Scratchpad**
Run Scratchpad
- Notebooks**
The place for data discovery and analytics
- Jobs**
Schedule notebooks to run at certain times
- Examples**
Check out some examples

Recent Activities

last Monday

- MSPIVAK **created** test1_2 notebook in MSPIVAK Project [MSPIVAK Workspace]
8/17/20 4:54 PM
- MSPIVAK **created** test1_1 notebook in MSPIVAK Project [MSPIVAK Workspace]
8/17/20 1:16 PM
- MSPIVAK **created** test1 notebook in MSPIVAK Project [MSPIVAK Workspace]
8/17/20 1:15 PM

AutoML



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

ORACLE Machine Learning

MSPIVAK Project [MSPIVAK Works...]

MSPIVAK

Create Experiment

Start Save Cancel

Name *
Targeting Customers

Comments

Data Source *
[Search]

Prediction Type *
Select Prediction Type

Predict *
Select Prediction Target

Case ID
Select Case ID

Additional Settings

Features

Refresh Search...

Name	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
No data to display.							

Create an Experiment

Select Data Source



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

The screenshot displays the Oracle Machine Learning 'Create Experiment' interface. A 'Select Table' dialog box is open, showing a list of schemas and tables. The 'SH' schema is selected in the 'Schema' column, and 'SUPPLEMENTARY_DEMOGRAPHICS' is selected in the 'Table' column. A red arrow points from the text 'Select Data Source' to the 'SH' schema. The background interface includes a 'Name' field with 'Targeting Customers', a 'Data Source' field, and a 'Prediction Type' dropdown. The 'Additional Settings' and 'Features' sections are also visible.

Schema	Table
ADBSNMP	COSTS
ADMIN	PRODUCTS
GGADMIN	PROMOTIONS
JJSANCHE	CHANNELS
MSPIVAK	SALES
SH	SUPPLEMENTARY_DEMOGRAPHICS
SSB	TIMES
	COUNTRIES
	CUSTOMERS

Select Data Source



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

ORACLE Machine Learning

MSPIVAK Project [MSPIVAK Works...]

MSPIVAK

Create Experiment

Start Save Cancel

Name *
Targeting Customers

Comments

Data Source *
SH.SUPPLEMENTARY_DEMOGRAPHICS

Prediction Type *
Select Prediction Type

Additional Settings

Features

Refresh

<input type="checkbox"/>	Name	Type	Percent NULLs
<input checked="" type="checkbox"/>	AFFINITY_CARD	NUMBER	0
<input checked="" type="checkbox"/>	BOOKKEEPING_APPLICATION	NUMBER	0
<input checked="" type="checkbox"/>	BULK_PACK_DISKETTES	NUMBER	0
<input checked="" type="checkbox"/>	COMMENTS	VARCHAR2	4.772991850989523
<input checked="" type="checkbox"/>	CUST_ID	NUMBER	0

Predict *
Select Prediction Target

- CUST_ID
NUMBER
- EDUCATION
VARCHAR2
- OCCUPATION
VARCHAR2
- HOUSEHOLD_SIZE
VARCHAR2
- YRS_RESIDENCE
NUMBER
- AFFINITY_CARD
NUMBER**
- BULK_PACK_DISKETTES
NUMBER
- FLAT_PANEL_MONITOR
NUMBER
- HOME_THEATER_PACKAGE
NUMBER

Search...

Std Dev
.6
.45
.68
1306.44

Select Target Attribute



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine

ORACLE Machine Learning

MSPIVAK Project [MSPIVAK Works...]

MSPIVAK

Create Experiment

Experiment ready to run

Name *
Targeting Customers

Comments

Data Source *
SH.SUPPLEMENTARY_DEMOGRAPHICS

Prediction Type *
Classification

Predict *
AFFINITY_CARD

Case ID
CUST_ID

Additional Settings

Features

Refresh

<input type="checkbox"/>	Name	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
<input checked="" type="checkbox"/>	AFFINITY_CARD	NUMBER	0	2	0	1	.24	.6
<input checked="" type="checkbox"/>	BOOKKEEPING_APPLICATION	NUMBER	0	2	0	1	.89	.45
<input checked="" type="checkbox"/>	BULK_PACK_DISKETTES	NUMBER	0	2	0	1	.64	.68
<input checked="" type="checkbox"/>	COMMENTS	VARCHAR2	4.772991850989523	43				
<input type="checkbox"/>	CUST_ID	NUMBER	0	4500	100001	104500	102250.76	1306.44
<input type="checkbox"/>	EDUCATION	VARCHAR2	0	16				



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

Additional Settings

Reset

Maximum Top Models *
5

Limit Run Duration (Hours) *
8

Model Metric *
Accuracy

Algorithms

- Name
- Decision Tree
- Generalized Linear Model
- Generalized Linear Model (Ridge Regression)
- Neural Network
- Random Forest
- Support Vector Machine (Gaussian)

Features

Refresh

	Name	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev.
<input type="radio"/>	AFFINITY_CARD	NUMBER	0	2	0	1	.24	.6
<input checked="" type="radio"/>	BOOKKEEPING_APPLICATION	NUMBER	0	2	0	1	.89	.45

← Additional Settings



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

The screenshot displays the AutoML user interface configuration page. At the top, there are input fields for 'Targeting Customers' and 'Comments'. Below these are the 'Data Source' (SH.SUPPLEMENTARY_DEMOGRAPHICS) and 'Prediction Type' (Classification) dropdowns. To the right, there are 'Predict' (AFFINITY_CARD) and 'Case ID' (CUST_ID) dropdowns. The 'Additional Settings' section includes a 'Reset' button, 'Maximum Top Models' (5), 'Limit Run Duration (Hours)' (8), and a 'Model Metric' dropdown menu. The dropdown menu is open, showing options: Accuracy, Balanced Accuracy, ROC AUC, and F1. An orange arrow points to the 'Accuracy' option with the text 'Select model metric'. At the bottom, there are checkboxes for 'Generalized Linear Model (Ridge Regression)', 'Neural Network', 'Random Forest', and 'Support Vector Machine (Gaussian)', all of which are checked.

Targeting Customers

Comments

Data Source *
SH.SUPPLEMENTARY_DEMOGRAPHICS

Prediction Type *
Classification

Predict *
AFFINITY_CARD

Case ID
CUST_ID

Additional Settings

Reset

Maximum Top Models *
5

Limit Run Duration (Hours) *
8

Model Metric *
Accuracy

Accuracy

Balanced Accuracy

ROC AUC

F1

Generalized Linear Model (Ridge Regression)

Neural Network

Random Forest

Support Vector Machine (Gaussian)

Select model metric

Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

ORACLE Machine Learning

MSPIVAK Project [MSPIVAK Works...]

MSPIVAK

Create Experiment

Start Save Cancel

Faster Results
Better Accuracy

Name *
Targeting Customers

Comments

Data Source *
SH.SUPPLEMENTARY_DEMOGRAPHICS

Prediction Type *
Classification

Predict *
AFFINITY_CARD

Case ID
CUST_ID

Additional Settings

Reset

Maximum Top Models *
5

Limit Run Duration (Hours) *
8

Model Metric *
Accuracy

Algorithms

- Name
- Decision Tree
- Generalized Linear Model

← More settings



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

Leader Board



Leader board

Algorithm	Name	Accuracy
Naive Bayes	nb_f8637b9359	0.4282
Random Forest	rf_aed7a58923	0.3624
Decision Tree	dt_6cd748831e	0.3624
Neural Network	nn_86b0b368e1	0.3529
Support Vector Machine (Linear)	svm_b3954f73b4	0.3224

Features

Attribute importance

Refresh

Name	Importance	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
OCCUPATION	<div><div style="width: 25%;"></div></div>	VARCHAR2	0	15				
YRS_RESIDENCE	<div><div style="width: 25%;"></div></div>	NUMBER	0	15	0	14	4.02	1.97
BOOKKEEPING_APPLICATION	<div><div style="width: 10%;"></div></div>	NUMBER	0	2	0	1	.89	.45
COMMENTS	<div><div style="width: 5%;"></div></div>	VARCHAR2	4.772991850989523	43				
AFFINITY_CARD	<div><div style="width: 5%;"></div></div>	NUMBER	0	2	0	1	.24	.6
HOME_THEATER_PACKAGE	<div><div style="width: 5%;"></div></div>	NUMBER	0	2	0	1	.56	.7
Y_BOX_GAMES	<div><div style="width: 5%;"></div></div>	NUMBER	0	2	0	1	.31	.66
FLAT_PANEL_MONITOR	<div><div style="width: 5%;"></div></div>	NUMBER	0	2	0	1	.58	.7
BULK_PACK_DISKETTES	<div><div style="width: 5%;"></div></div>	NUMBER	0	2	0	1	.64	.68



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine

The screenshot displays the Oracle Machine Learning AutoML User Interface. The main area shows an experiment named 'test 4' with a 'Metric Chart' and a 'Leader Board'. A 'Run Summary' panel is open on the right, listing completed tasks: Algorithm Selection, Adaptive Sampling, Feature Selection, Hyperparameter Tuning, Naive Bayes, Random Forest, Neural Network, Decision Tree, and Support Vector Machine (Linear). Three red arrows point from the text 'Model summaries' to the 'Metric Chart', the 'Leader Board', and the 'Run Summary' panel.

ORACLE Machine Learning

MSPIVAK Project [MSPIVAK Works...]

MSPIVAK

<- Experiments

test 4

Experiment Settings Edit

Metric Chart

Leader Board

Deploy Create Notebook Metrics

Algorithm	Name	Accuracy
Naive Bayes	nb_f8637b9359	0.4282
Random Forest	rf_aed7a58923	0.3624
Decision Tree	dt_6cd748831e	0.3624
Neural Network	nn_86b0b368e1	0.3529
Support Vector Machine (Linear)	svm_b3954f73b4	0.3224
Naive Bayes	nb_f8637b9359	0.4282
Random Forest	rf_aed7a58923	0.3624

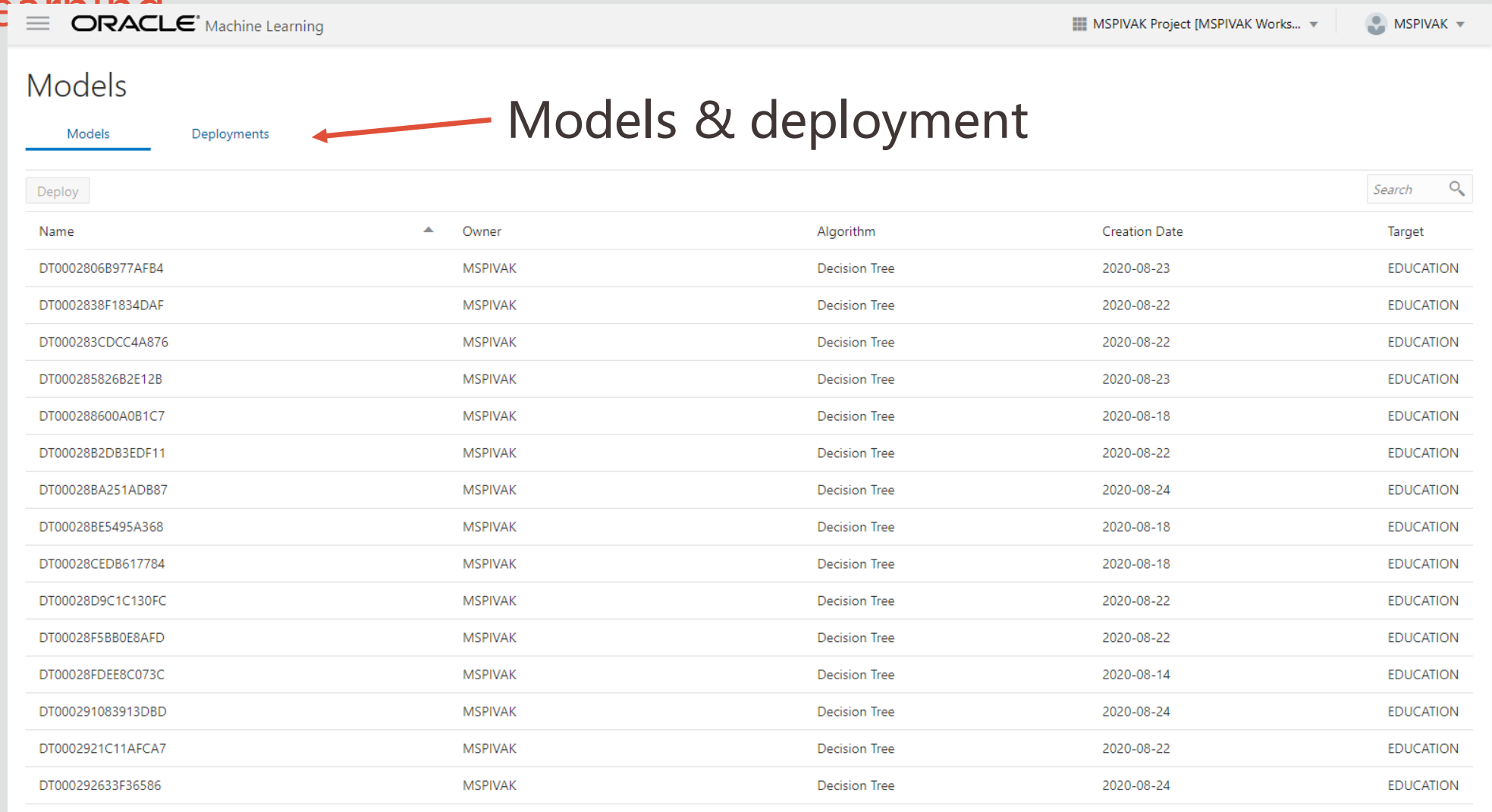
Run Summary

- Algorithm Selection Completed
- Adaptive Sampling Completed
- Feature Selection Completed
- Hyperparameter Tuning Completed
- Naive Bayes Completed
- Random Forest Completed
- Neural Network Completed
- Decision Tree Completed
- Support Vector Machine (Linear) Completed

Model summaries

Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning



The screenshot displays the Oracle Machine Learning AutoML User Interface. The top navigation bar includes the Oracle logo, the text "Machine Learning", and the project name "MSPIVAK Project [MSPIVAK Works...". The main content area is titled "Models" and features two tabs: "Models" (selected) and "Deployments". A red arrow points from the text "Models & deployment" to the "Deployments" tab. Below the tabs is a table with columns for Name, Owner, Algorithm, Creation Date, and Target. The table contains 15 rows of model data, all with "Decision Tree" as the algorithm and "EDUCATION" as the target. A "Deploy" button is visible in the top left of the table area, and a search bar is in the top right.

Name	Owner	Algorithm	Creation Date	Target
DT0002806B977AFB4	MSPIVAK	Decision Tree	2020-08-23	EDUCATION
DT0002838F1834DAF	MSPIVAK	Decision Tree	2020-08-22	EDUCATION
DT000283CDCC4A876	MSPIVAK	Decision Tree	2020-08-22	EDUCATION
DT000285826B2E12B	MSPIVAK	Decision Tree	2020-08-23	EDUCATION
DT000288600A0B1C7	MSPIVAK	Decision Tree	2020-08-18	EDUCATION
DT00028B2DB3EDF11	MSPIVAK	Decision Tree	2020-08-22	EDUCATION
DT00028BA251ADB87	MSPIVAK	Decision Tree	2020-08-24	EDUCATION
DT00028BE5495A368	MSPIVAK	Decision Tree	2020-08-18	EDUCATION
DT00028CEDB617784	MSPIVAK	Decision Tree	2020-08-18	EDUCATION
DT00028D9C1C130FC	MSPIVAK	Decision Tree	2020-08-22	EDUCATION
DT00028F5BB0E8AFD	MSPIVAK	Decision Tree	2020-08-22	EDUCATION
DT00028FDEE8C073C	MSPIVAK	Decision Tree	2020-08-14	EDUCATION
DT000291083913DBD	MSPIVAK	Decision Tree	2020-08-24	EDUCATION
DT0002921C11AFCA7	MSPIVAK	Decision Tree	2020-08-22	EDUCATION
DT000292633F36586	MSPIVAK	Decision Tree	2020-08-24	EDUCATION



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

The screenshot shows the Oracle Machine Learning web interface. A 'Deploy Model' dialog box is open in the foreground, with a red arrow pointing to the 'Name' field. The dialog box contains the following fields and options:

- Name * (with a tooltip: "Enter a unique, alphanumeric Name (max 50 characters)"): DT0002806B977AFB4
- URI *
- Version *
- Namespace
- Shared
- OK and Cancel buttons

The background shows a table of models with columns for Name, Namespace, Model Type, Creation Date, and Target. The text "Model deployment" is overlaid on the dialog box.

Name	Namespace	Model Type	Creation Date	Target
DT0002806B977AFB4				
DT0002838F1834DAF			2020-08-23	EDUCATION
DT000283CDCC4A876			2020-08-22	EDUCATION
DT000285826B2E12B			2020-08-22	EDUCATION
DT000288600A0B1C7			2020-08-23	EDUCATION
DT00028B2DB3EDF11			2020-08-18	EDUCATION
DT00028BA251ADB87			2020-08-22	EDUCATION
DT00028BE5495A368			2020-08-24	EDUCATION
DT00028CEDB617784			2020-08-18	EDUCATION
DT00028D9C1C130FC	MSPIVAK	Decision Tree	2020-08-22	EDUCATION
DT00028F5BB0E8AFD	MSPIVAK	Decision Tree	2020-08-22	EDUCATION
DT00028FDEE8C073C	MSPIVAK	Decision Tree	2020-08-14	EDUCATION



Coming Soon! | AutoML User Interface

"Code-free" user interface supporting automated end-to-end machine learning

The screenshot displays the Oracle Machine Learning Model Repository interface. The main view is titled "Model Repository" and shows a table of models. A modal window titled "Model metadata for DT000281545B55889" is open, displaying the following JSON metadata:

```
{
  "miningFunction": "CLASSIFICATION",
  "algorithm": "DECISION_TREE",
  "attributes": [
    {
      "name": "AFFINITY_CARD",
      "attributeType": "NUMERICAL"
    },
    {
      "name": "BOOKKEEPING_APPLICATION",
      "attributeType": "NUMERICAL"
    },
    {
      "name": "OCCUPATION",
      "attributeType": "CATEGORICAL"
    },
    {
      "name": "Y_BOX_GAMES",
      "attributeType": "NUMERICAL"
    }
  ],
  "output": {
    "name": "EDUCATION",
    "attributeType": "CATEGORICAL"
  },
  "labels": [
    "10th",
    "11th",
    "12th",
    "13th"
  ]
}
```

An orange arrow points from the text "Model metadata & model deployment" to the "miningFunction" field in the JSON metadata.

Name	Shared	Version	Name
DT000281545B55889		1.0	
DT0002868CBEB0E51		1.0	



Coming Soon! | OML Services

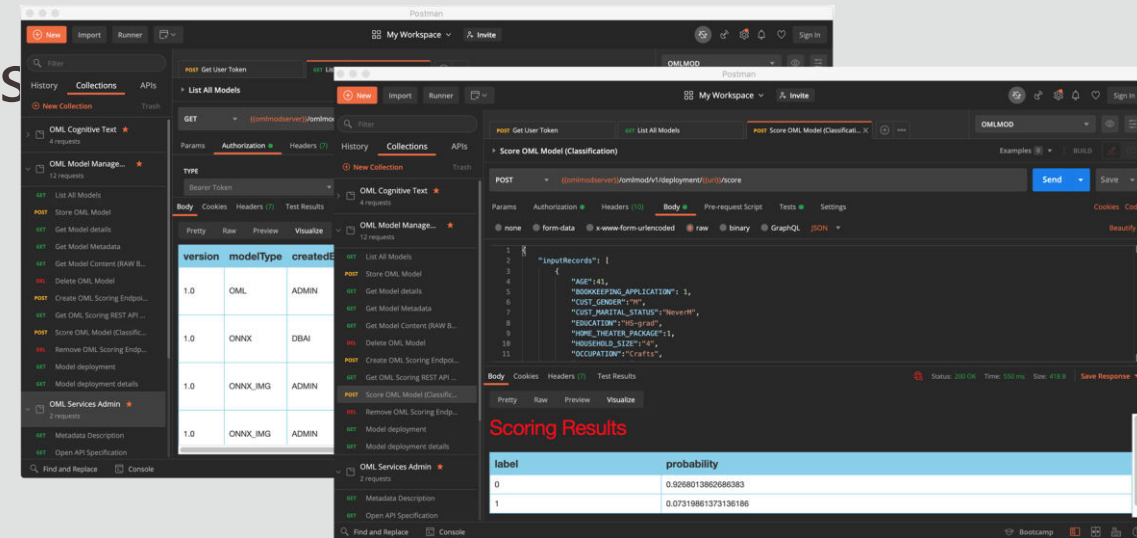
Supports model deployment and lifecycle management beyond Database

Support model deployment and model lifecycle management for both in-database OML and third-party classification for regression models via REST APIs

Supports ONNX (Open Neural Networks eXchange) format including Scikit-learn and TensorFlow and others

Supports OML cognitive text capabilities, with capabilities for topic discovery, keywords, summary, sentiment, and feature extraction (English, Spanish, and French initially).

Cognitive image supports ONNX format third-party model scoring for images or tensors



Coming Soon! | OML Services

Supports model deployment and lifecycle management beyond Database

The screenshot displays the Postman interface with two overlapping windows. The background window shows a collection named 'List All Models' with a table of model details. The foreground window shows a 'POST Score OML Model (Classification)' request with a JSON body and a 'Scoring Results' table in the response.

version	modelType	createdBy
1.0	OML	ADMIN
1.0	ONNX	DBAI
1.0	ONNX_IMG	ADMIN
1.0	ONNX_IMG	ADMIN

```
POST {{omlmodserver}}/omlmod/v1/deployment/{{uri}}/score
```

```
{
  "inputRecords": [
    {
      "AGE": 41,
      "BOOKKEEPING_APPLICATION": 1,
      "CUST_GENDER": "M",
      "CUST_MARITAL_STATUS": "NeverM",
      "EDUCATION": "HS-grad",
      "HOME_THEATER_PACKAGE": 1,
      "HOUSEHOLD_SIZE": 4,
      "OCCUPATION": "Crafts",
    }
  ]
}
```

label	probability
0	0.9268013862686383
1	0.07319861373136186



Oracle Analytics Cloud 5.7
+ OML = ~"Solutions"

Oracle Machine Learning + Oracle Analytics Cloud

ORACLE Machine Learning

Employee Attrition Prediction Model Development_Final_1

Predict Employees that will Leave the Company

We will use our employee data set to create a model that will help us predict which employees are most likely to leave the company.

Retail attrition rates are an important topic in the retail industry as leadership seeks to determine ways that will lower labor costs to increase total company revenue. Turnover in the retail industry is exceptionally high and affects company profitability by requiring companies to spend more time recruiting, hiring, and training new employees. Constant turnover also disrupts the work environment, as other employees must fill the gap left by those leaving. This can disrupt not only the working environment for the employees but also the perceptions of the customers if their needs are not met. Developing solutions to mitigate turnover intentions of employees is increasingly important and some leaders believe turnover is tied to the training and development of the employees.

Created and last updated June 2020 by Derrick Cameron

Let's explore the data - age attribute

Show the distribution of employee attrition risk - count of employees at risk of leaving by percent risk

We have five models. Which is best for predicting attrition? Lets compare the models using a cumulative gains chart.

ORACLE Machine Learning



ORACLE Analytics

Home

Search Everything

Projects and Reports Data Recent Data Sets Favorite Projects Machine Learning

Projects

View all

Employee Attrition Prediction Model Dev...

For each employee what are the factors that determine whether they stay or leave, and what weight do these factors have?

```

sql
select a.employee_number, b.attrition_prediction, a.name, a.actual_value, a.weight, a.rank
from prediction_details_v a, new_employees_v b
where a.employee_number = b.employee_number
order by a.employee_number, a.rank
    
```

EMPLOYEENUMBER	ATTRITION_PREDICTION	NAME	ACTUAL_VALUE	WEIGHT	RANK
1	Yes	OVERTIME	Yes	0.421	1
1	Yes	MARITALSTATUS	Single	0.217	2
1	Yes	NUMCOMPANIESWORKED	8	0.201	3
1	Yes	WORKLIFEBALANCE	1	0.188	4
1	Yes	RELATIONSHIPSATISFAC	1	0.103	5
2	No	AGE	49	0.013	1
2	No	RELATIONSHIPSATISFAC	4	0.008	2

ORACLE Machine Learning

ORACLE Analytics

Home

Type Model

Projects and Reports Data Recent Data Sets Favorite Projects Machine Learning

Sort by Modified

This Month

BUY_INSURANCE_CLA... WINE_CLASS_MODEL... WINE_CLASS_MODEL N1_CLASS_MODEL_DE... N1_CLASS_MODEL ATTRITION_MODEL_RF

N1_CLASS_MODEL_SVM ATTRITION_MODEL_SVM



Register an OML Model in OAC

The screenshot displays the Oracle Analytics Cloud (OAC) home page. At the top left, the Oracle Analytics logo is visible. The main header is blue with a 'Home' button and a 'Create' button. A search bar labeled 'Search Everything' is positioned above a navigation bar with tabs for 'Projects and Reports', 'Data', 'Recent Data Sets', 'Favorite Projects', and 'Machine Learning'. The central area features a large heading 'Register OML models' with a red arrow pointing to a dropdown menu. The dropdown menu includes options: 'Import Project/Flow...', 'Data Set Management', 'Open Data Modeler', 'Register ML Model' (highlighted with a red border), 'Open Classic Home', and 'Customize Home Page...'. Below the menu, a grid of project cards is shown, including 'Wine2 Project', 'DMSVLWINE_CLASS_M...', 'DMSVXWINE_CLASS...', 'Picking a Good Wine', 'WINEREVIEWS130KTEXT', 'CUST_INSUR_LTV_Like...', 'Claims Fraud', 'CLAIMS_SUSPICIOUS', 'CLAIMS', 'CUST_INSUR_LIKELY...', and 'BUY_INSURANCE_CLA...'. A 'Projects' section at the bottom shows a subset of these projects with a 'View all' link. A 'Data Sets' section is partially visible at the bottom with another 'View all' link.



Register an OML Model in OAC

The screenshot shows the Oracle Analytics Cloud (OAC) interface. At the top, there is a navigation bar with 'ORACLE Analytics', a search icon, and a user profile icon 'CB'. Below this is a 'Home' header with a 'Create' button. The main content area is titled 'Select a Model to Register'. A modal dialog box titled 'Register an OML model' is open, showing a search filter 'attrition'. The dialog contains a table of models and a detailed view of the selected model.

Type	Name	Last Modified
	ATTRITION_MODEL_DT	Aug 4, 2020
	ATTRITION_MODEL_GLM	Aug 4, 2020
	ATTRITION_MODEL_NB	Aug 4, 2020
	ATTRITION_MODEL_RF	Aug 4, 2020
	ATTRITION_MODEL_SVM	Aug 4, 2020
	ATTRITION_MODEL_SVM2	May 20, 2020

Parameters	
target	ATTRITION
ALGO_NAME	ALGO_RANDOM_FOREST
PREP_AUTO	ON
TREE_TERM_MINPCT_NODE	.05
TREE_TERM_MINREC_SPLIT	20
ODMS_RANDOM_SEED	0
TREE_IMPURITY_METRIC	TREE_IMPURITY_GINI
CLAS_MAX_SUP_BINS	32
CLAS_WEIGHTS_BALANCED	OFF



Define a Data Flow using OML in-DB Model

The screenshot displays the Oracle Analytics web interface. At the top, the header shows 'ORACLE Analytics' and a user profile 'CB'. Below the header, a blue navigation bar contains 'Untitled', 'Run Data Flow', and 'Save' buttons. A left sidebar lists 'Data Flow Steps' including 'Add Data', 'Join', 'Union Rows', 'Filter', 'Aggregate', 'Save Data Set', 'Create Essbase Cube', 'Add Columns', 'Select Columns', 'Rename Columns', 'Transform Column', 'Merge Columns', 'Split Columns', 'Bin', 'Group', and 'Branch'. The main workspace is titled 'Pick an OML model' and features a grid of icons for various data processing tasks. A red arrow points to the 'Apply Model' icon, which is also enclosed in a red box. Below the grid, there are checkboxes for 'DAILYRATE' and 'DEPARTMENT'. At the bottom, a 'Query Error' message is visible with a 'More Detail' link.



Define a Data Flow using OML in-DB Model

The screenshot shows the Oracle Analytics interface. At the top, there's a navigation bar with 'ORACLE Analytics' and a user profile 'CB'. Below that, a blue header bar contains 'Untitled', 'Run Data Flow', and 'Save'. A left sidebar lists 'Data Flow Steps' including 'Add Data', 'Join', 'Union Rows', 'Filter', 'Aggregate', 'Save Data Set', 'Create Essbase Cube', 'Add Columns', 'Select Columns', 'Rename Columns', 'Transform Column', 'Merge Columns', 'Split Columns', 'Bin', and 'Group'. The main workspace shows a data flow diagram with three steps: 'EMPLOY...', 'Apply Model', and 'Save Data'. A red box highlights these steps, and a red arrow points to the 'Apply Model' step with the text 'OAC runs OML model in ADW'. Below the diagram is the 'Save Data Set' configuration panel. It shows 'Data Set' as 'At_Risk_Employees' and 'Description' as 'At_Risk_Employees'. The 'Save data to' section is set to 'Database Connection' with 'Table' 'At_Risk_Employees' and 'Connection' 'adwc4pm_3'. The 'When run' section is set to 'Replace existing data'. A table of columns is displayed with their respective database names, treatments, and aggregations.

Name	Database Name	Treat As	Default Aggregation
AGE	AGE	Measure	Sum
ATTRITION	ATTRITION	Attribute	
BUSINESSSTRAVEL	BUSINESSSTRAVEL	Attribute	
DAILYRATE	DAILYRATE	Measure	Sum
DEPARTMENT	DEPARTMENT	Attribute	
DISTANCEFROMHOME	DISTANCEFROMHOME	Measure	Sum

99 AGE	ab ATTRITION	ab BUSINESSSTRA...	99 DAILYRATE	ab DEPARTMENT	99 DISTANCEFR...	99 EDUCATION	ab EDUCATIONF
26	Yes	Travel_Rarely	1357	Research & Development	25	3	Life Sciences
27	No	Travel_Frequently	994	Sales	8	3	Life Sciences
30	No	Travel_Frequently	721	Research & Development	1	2	Medical
41	Yes	Travel_Rarely	1360	Research & Development	12	3	Technical Degree
34	No	Non-Travel	1065	Sales	23	4	Marketing
27	No	Travel_Rarely	408	Research & Development	10	2	Life Sciences

Investigate OML Insights and Predictions

The screenshot displays the Oracle Analytics interface. At the top, the navigation bar includes the Oracle Analytics logo, a search bar, and a 'Create' button. Below the navigation bar, there are tabs for 'Data Sets', 'Connections', 'Data Flows', 'Sequences', and 'Data Replications'. A search bar labeled 'Search Data' and a 'Sort By' dropdown menu are also present. The main content area is titled 'Today OML's Predictions' and is divided into three sections: 'Today', 'Yesterday', and 'This Month'. Each section contains a grid of data set cards, each with a cloud icon and a label. The 'Today' section has three cards: 'At_Risk_Employees', 'EMPLOYEE_DATA', and 'DMSVLWINE_CLASS_M...'. The 'Yesterday' section has two cards: 'DMSVXWINE_CLASS_...' and 'DMSVLWINE_CLASS_M...'. The 'This Month' section has six cards: 'DMSVAGOOD_WINE_AI', 'WINEREVIEWS130KTEXT', 'price_bin_UnitedStates', 'Best_Wines', 'CLAIMS_SUSPICIOUS', and 'CLAIMS'. The 'At_Risk_Employees' card is highlighted with a red border.

ORACLE Analytics

Data

Data Sets Connections Data Flows Sequences Data Replications

Search Data Sort By Modified

Today **OML's Predictions**

At_Risk_Employees EMPLOYEE_DATA DMSVLWINE_CLASS_M...

Yesterday

DMSVXWINE_CLASS_... DMSVLWINE_CLASS_M...

This Month

DMSVAGOOD_WINE_AI WINEREVIEWS130KTEXT price_bin_UnitedStates Best_Wines CLAIMS_SUSPICIOUS CLAIMS



Investigate OML Insights and Predictions

The screenshot shows the Oracle Analytics interface for a dashboard titled 'At_Risk_Employees'. At the top, there are navigation tabs: 'Prepare', 'Visualize', and 'Narrate'. Below these are filter controls for 'DEPARTMENT', 'BUSINESS TRAVEL', 'JOB LEVEL', 'YEARS IN CURRENT ROLE', and 'PERCENTSALARYHIKE', all currently set to 'All' or 'Full Range'. A red arrow labeled 'Interactive Filters' points to these controls. On the left, a 'Data' sidebar lists various attributes like EDUCATION, MONTHLYRATE, and PERCENTSALARYHIKE. The main area features a table with columns for EMPLOYEE NUMBER, DEPARTMENT, GENDER, and PREDICTION. A red box highlights the table and a 'Range' filter dialog. To the right, a bar chart titled 'Value Distribution for PERCENTSALARYHIKE by BUSINESS TRAVEL, DE...' is visible. At the bottom, the status bar shows 'Canvas 1' and '1001+ Rows, 7 Colu...'. The interface is branded with 'ORACLE Analytics' in the top left.

Predictions



Investigate OML Insights and Predictions

ORACLE Analytics

At_Risk_Employees

Prepare Visualize Narrate Save

PREDICTION DEPARTMENT (3) STOCKOPTIONLEVEL

Yes Human Resources, Research & Developme... Full Range

Bar

Trellis Columns

Trellis Rows

Values (Y-Axis)

PREDICTION...

Category (X-Axis)

YEARSATCO...

Color

DEPARTMENT

Size (Width)

Tooltip

Detail

Filters

PREDICTIONPROBABILITY by YEARSATCOMPANY (Grouped), DEPARTMENT

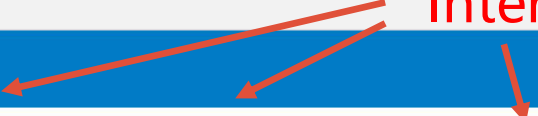
YEARSATCOMPANY (Grouped)	Research & Development	Sales	Human Resources
6.4 - 8	0.66	0.66	0.00
11.2 - 12.8	0.52	0.74	0.00
4.8 - 6.4	0.62	0.61	0.57
0 - 1.6	0.60	0.64	0.00
1.6 - 3.2	0.61	0.59	0.00
9.6 - 11.2	0.61	0.55	0.00
12.8 - 14.4	0.67	0.52	0.00
8 - 9.6	0.59	0.55	0.00
14.4 - 16	0.56	0.00	0.00
17.6 - Max	0.58	0.00	0.51
3.2 - 4.8	0.58	0.53	0.51

DEPARTMENT Research & Development Sales Human Resources

Canvas 1 Canvas 2

11 Groups, 3 Bars

Interactive Filters



Investigate OML Insights and Predictions

ORACLE Analytics

At_Risk_Employees

Prepare Visualize Narrate Save

Click here or drag data to add a filter

Wide Range of Graphs Available

Chord Diagram

PREDICTION, BUSINESS TRAVEL, PREDICTIONPROBABILITY

Yes No

Travel_Rarely

Travel_Frequently

Non-Travel

PREDICTIONPROBABILITY 0.6 0.9

Canvas 1 Canvas 2 Canvas 3

5 Nodes



Wait, there is more!



OML + APEX

Interactively Explore Data and OML Insights and Predictions

The screenshot displays an Oracle APEX application interface. At the top, a blue header bar contains the text "Employee_Attrition" on the left and a user profile icon labeled "charlie_apex" on the right. A dark grey sidebar on the left lists navigation options: Home, Key Factors, Employee Attrition, More Employee Attrition Grap..., and Administration. The main content area features a purple icon of a bar chart with an upward arrow, followed by the title "Employee_Attrition". Below this is a large blue button with a white grid icon and the text "Employee Attrition". At the bottom of the main area, a white box contains the text "Created By: Charlie Berger, Dhvani Seth, Siddesh Ujjini, Rosy Parmar". The footer of the application shows a row of utility icons: Home, Application 104, Edit Page 1, Session, View Debug, Debug, Page Info, Quick Edit, Theme Roller, and a settings gear icon.



OML + APEX

Interactively Explore Data and OML Insights and Predictions

Employee_Attrition

Home Application 104 Edit Page 6 Session View Debug Debug Page Info Quick Edit Theme Roller charlie_apex

Employee_Attrition_From_Table

Search... Go

DEPARTMENT

- Research & Development (961)
- Sales (446)
- Human Resources (63)

Attrition Prediction

- No (1333)
- Yes (137)

ATTRITION_PRED_PCNT

- 20-40 (189)
- 40-60 (99)
- 60-80 (64)
- 80-100 (31)
- <20 (1087)

to Go

JOBLEVEL

- 1 (543)
- 2 (534)
- 3 (218)
- 4 (106)
- 5 (69)

BUSINESSSTRAVEL

- Travel_Rarely (1043)
- Travel_Frequently (277)
- Non-Travel (150)

Business Travel Frequency by Department

Travel Frequency	Research & Development	Sales	Human Resources
Non-Travel	~100	~50	~10
Travel_Frequently	~180	~80	~10
Travel_Rarely	~680	~320	~40

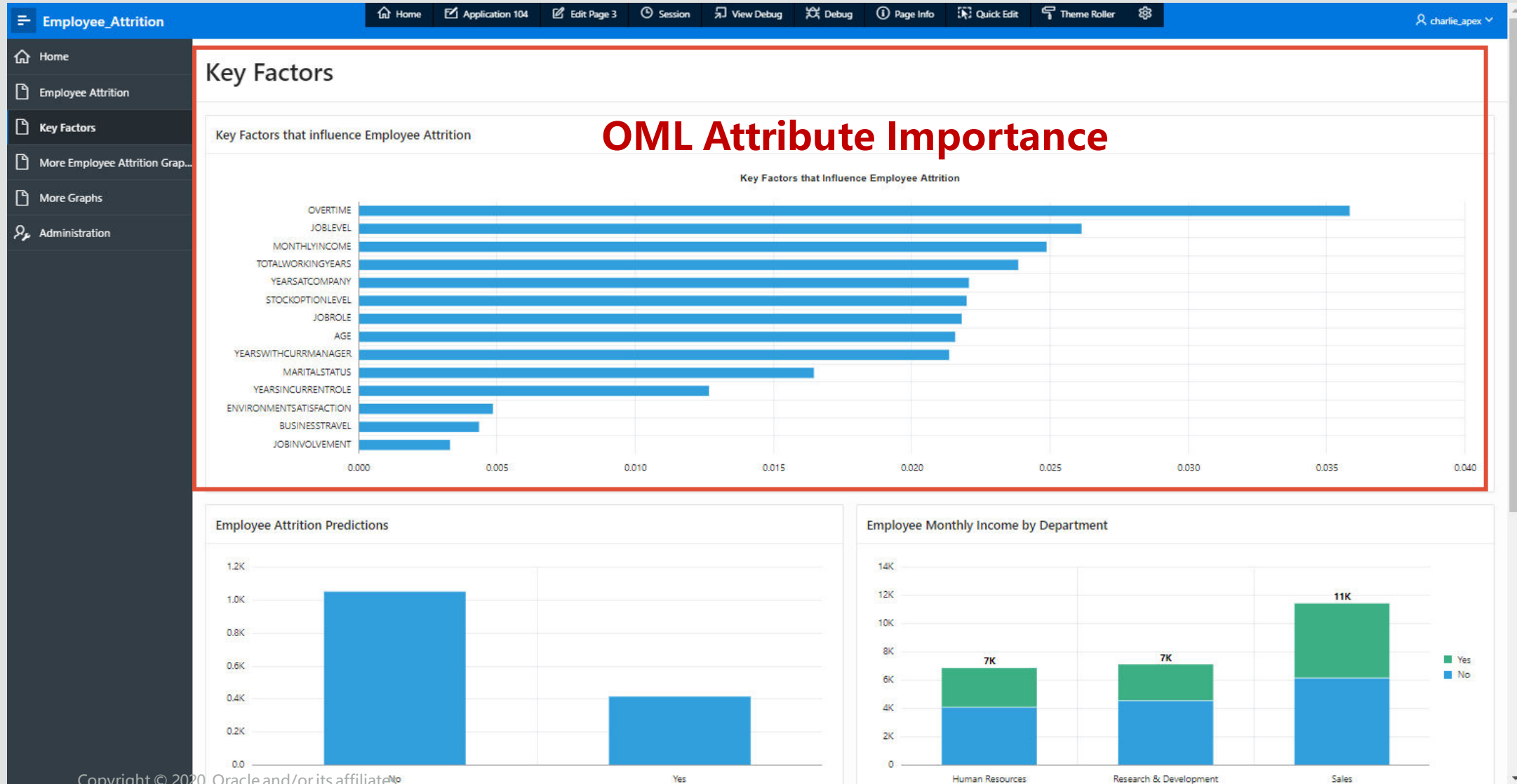
Predictions

Age	Employee number	Attrition Prediction Pcnt	Attrition Prediction	Overtime	Businesstravel	Dailyrate	Department	Distancefromhome	Education	Educationfield	Employeecount	Environmentsatisfaction	Gender	Hou
26	622	96.3	Yes	Yes	Travel_Rarely	471	Research & Development	24	3	Technical Degree	1	3	Male	
24	1,494	95.8	Yes	Yes	Travel_Frequently	381	Research & Development	9	3	Medical	1	2	Male	
21	478	95	Yes	Yes	Travel_Frequently	756	Sales	1	1	Technical Degree	1	1	Female	
18	614	94	Yes	Yes	Travel_Frequently	1,306	Sales	5	3	Marketing	1	2	Male	
32	911	94	Yes	Yes	Travel_Rarely	374	Research & Development	25	4	Life Sciences	1	1	Male	
32	33	92.6	Yes	Yes	Travel_Frequently	1,125	Research & Development	16	1	Life Sciences	1	2	Female	
25	1,273	92.1	Yes	Yes	Travel_Frequently	599	Sales	24	1	Life Sciences	1	3	Male	
32	1,862	92.1	Yes	Yes	Travel_Rarely	414	Sales	2	4	Marketing	1	3	Male	



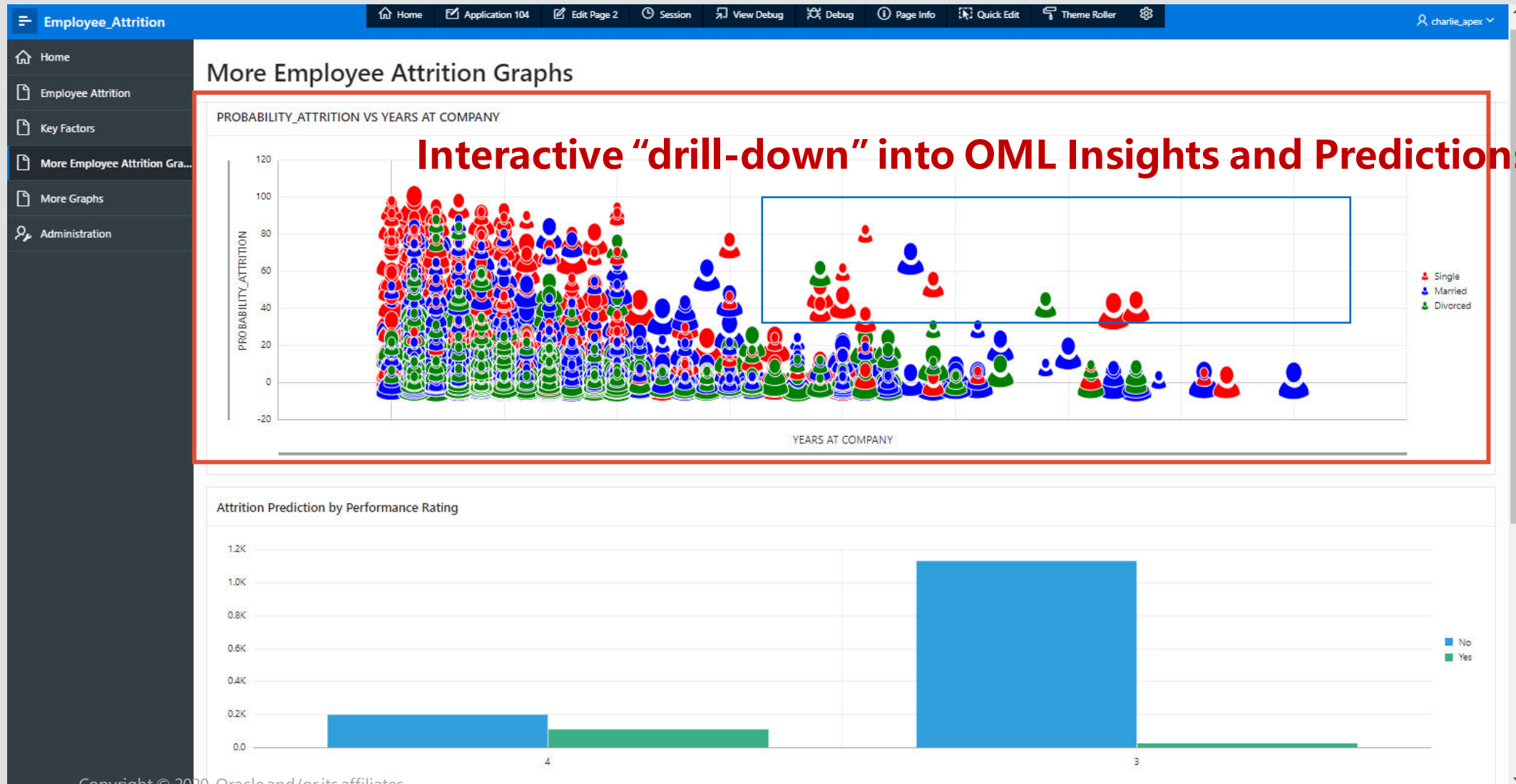
OML + APEX

Interactively Explore Data and OML Insights and Predictions



OML + APEX

Interactively Explore Data and OML Insights and Predictions



The Changing Role of the DBA:

From Oracle Data Professional to Oracle Data Scientist in 6 Weeks!

<https://www.datacamp.com/community/blog/data-scientist-vs-data-engineer>

<https://www.kdnuggets.com/2020/02/poll-automl-replace-data-scientists.html>

Copyright © 2020 Oracle and/or its affiliates.



Data Professional to Data Scientist Journey

You are Likely Already Doing Much of The Work!

Data extraction

Data wrangling

Deriving new attributes
("feature engineering")

...
...]

Where the Machine Learning "Magic" Happens

Import predictions & insights

Translate and deploy ML models
Oracle

Automate

Typically 80% of the work!

Most data scientists spend only 20 percent of their time on actual data analysis and 80 percent of their time finding, cleaning, and reorganizing huge amounts of data, which is an inefficient data strategy¹



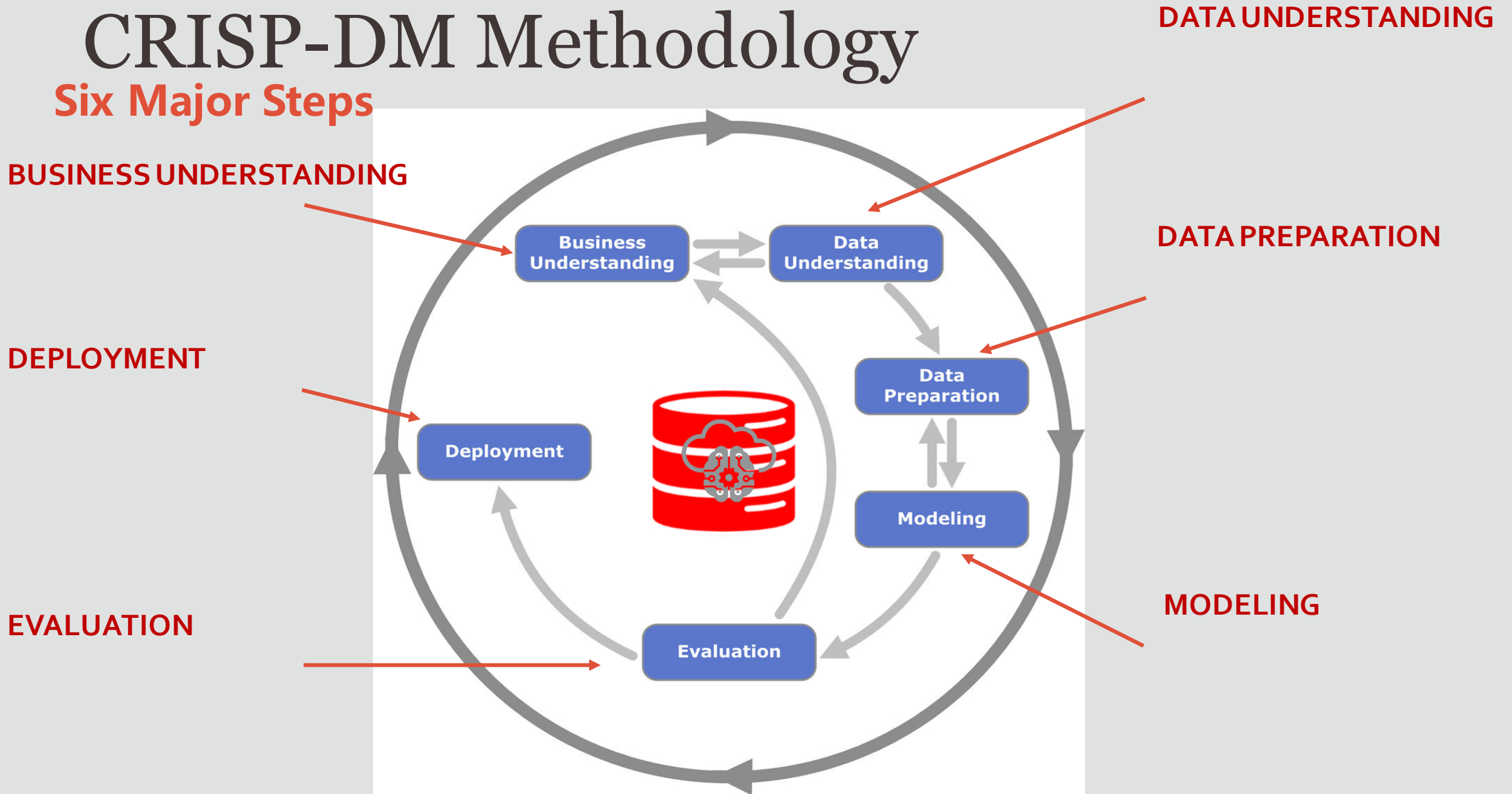
Eliminated or minimized w/

Data Management platform becomes
combined/hybrid data management + machine
learning platform

¹ - <https://www.infoworld.com/article/3228245/data-science/the-80-20-data-science-dilemma.html>

CRISP-DM Methodology

Six Major Steps



CRISP-DM Methodology

Six Major Steps

BUSINESS UNDERSTANDING

Well-defined business problem

DEPLOYMENT

In-DB ML model apply

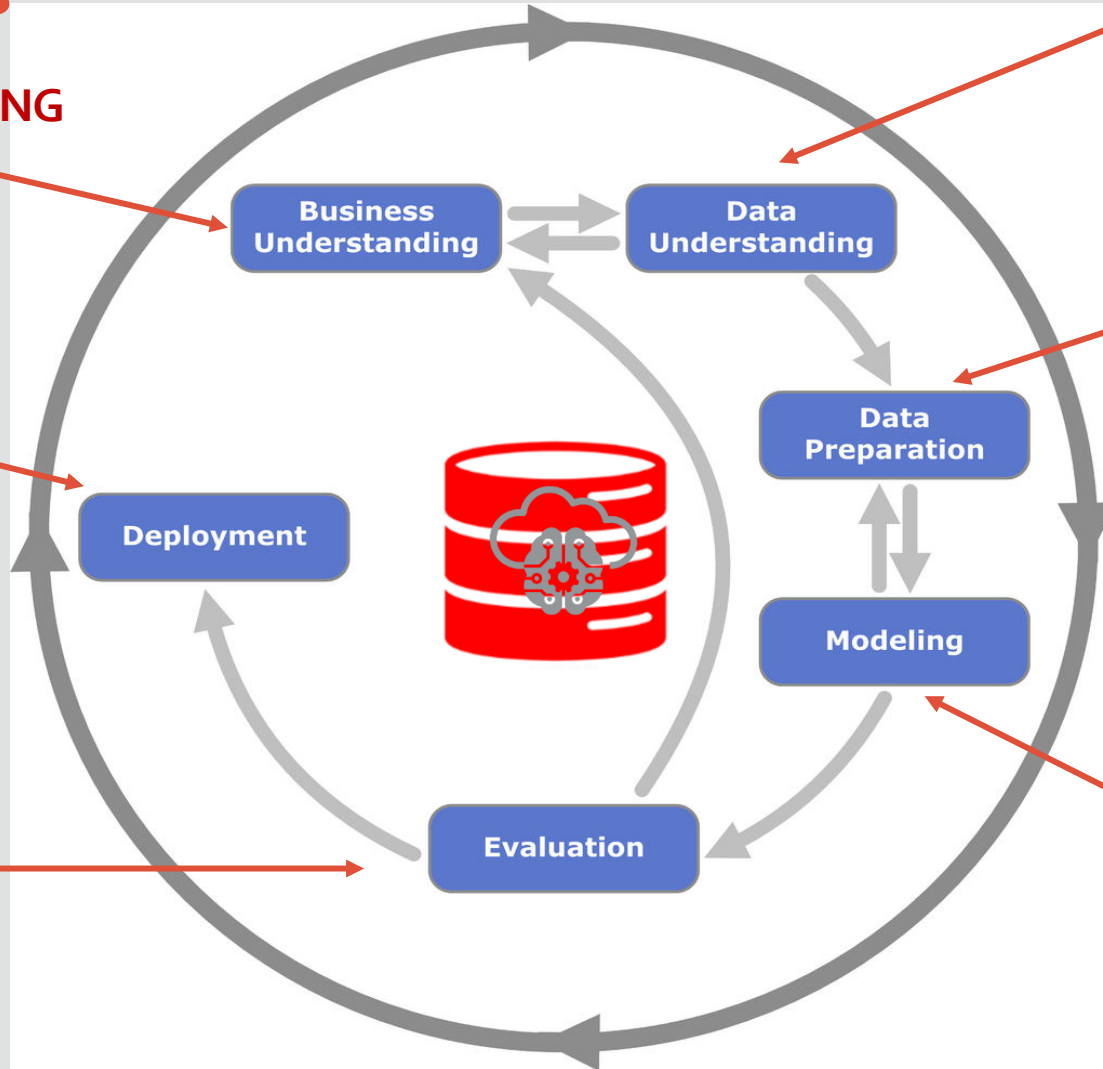
- Real-time ML apply
- In-database, REST

Embed methodology

- Applications
- Dashboards

EVALUATION

Model evaluation
Model comparison
Model selection



DATA UNDERSTANDING

Assemble the “right data”

Data profiling

- Data visualization
- Univariate statistics/group by
- Bi-variate statistics

DATA PREPARATION

Sampling/Stratified

Algorithm req'd transforms

- Auto Data Preparation
- Missing Values, Binning, Normalization, etc.
- Unstructured data
- Aggregations

Domain specific transforms

- “Engineered Features”

Features Selection

MODELING

Algorithm settings/defaults

- Stratified sampling
- Feature selection
- Build model(s)



CRISP-DM Methodology

Six Major Steps

BUSINESS UNDERSTANDING

Well-defined business problem

DEPLOYMENT

In-DB ML model apply

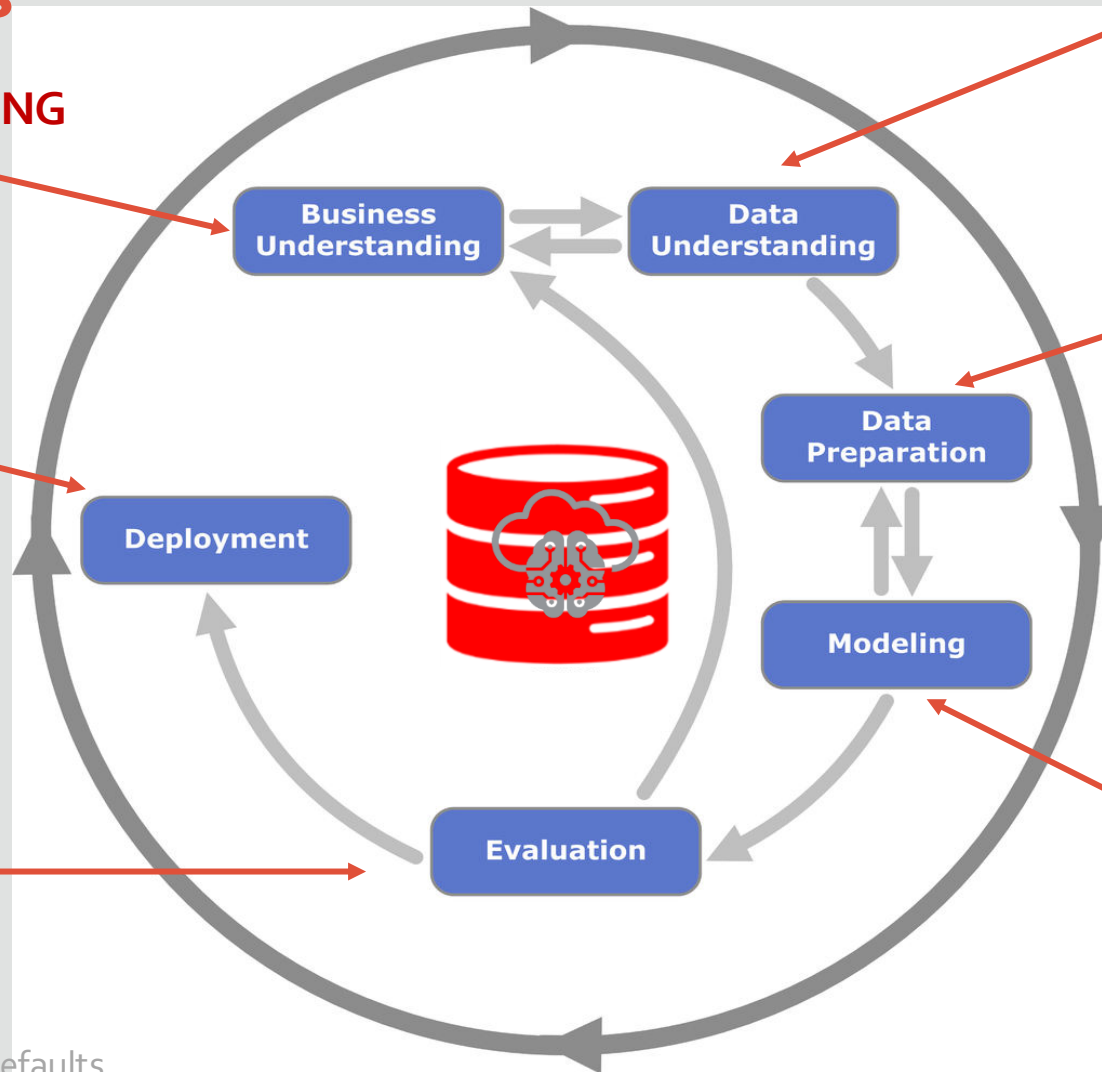
- Real-time ML apply
- In-database, REST

Embed methodology

- Applications
- Dashboards

EVALUATION

Model evaluation
Model comparison
Model selection



DATA UNDERSTANDING

Assemble the "right data"

Data profiling

- Data visualization
- Univariate statistics/group by
- Bi-variate statistics

DATA PREPARATION

Sampling/Stratified

Algorithm req'd transforms

- Auto Data Preparation
- Missing Values, Binning, Normalization, etc.
- Unstructured data
- Aggregations

Domain specific transforms

- "Engineered Features"

Features Selection

MODELING

Algorithm settings/defaults

- Stratified sampling
- Feature selection
- Build model(s)

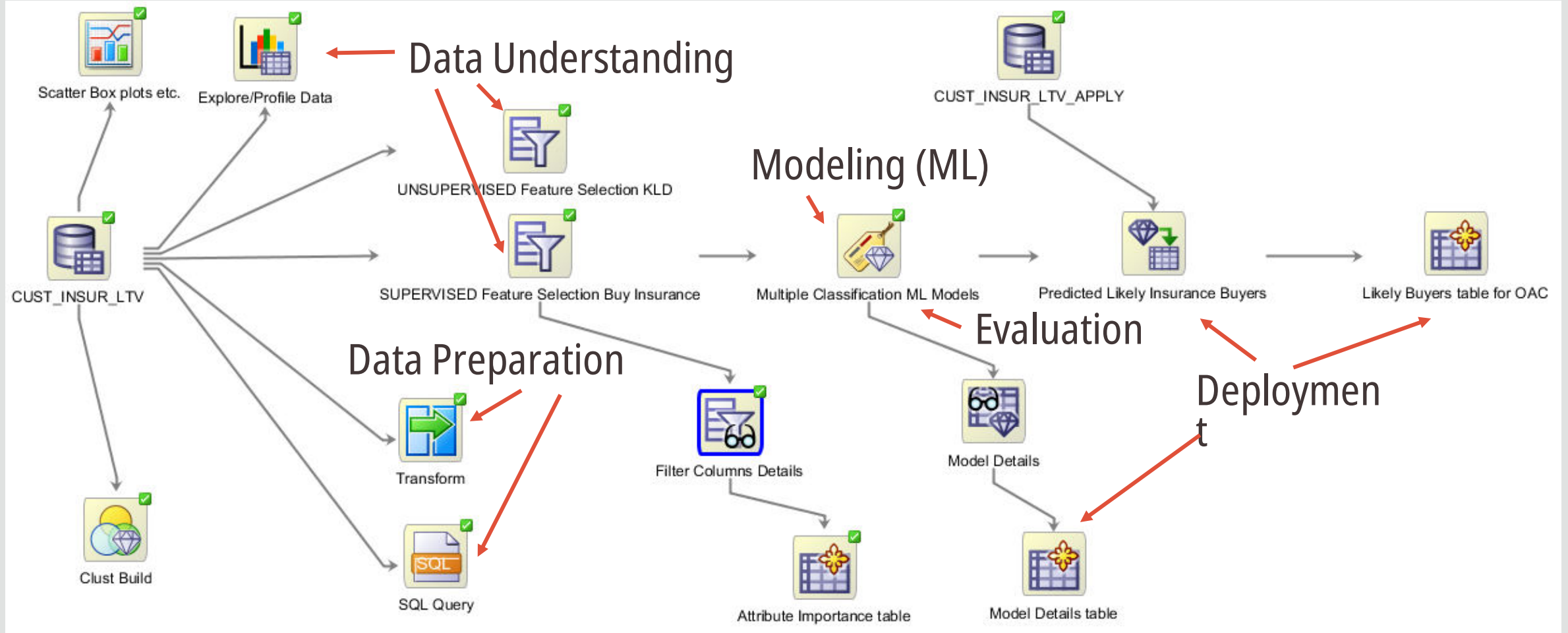
* Automated and/or system defaults

Oracle Machine Learning

SQL Developer Extension: Oracle Data Miner UI



Business Understanding: Target customers most likely to Buy Insurance





Example Templates

+ Create Notebook

Search by...

Anomaly Detection

This notebook shows how to detect...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Anomaly Detection' 'Machine...

★ 5 Likes 🔍 1601 📄 76

Association Rules

Notebook to show the use of Assoc...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'SQL' 'Associations' 'Rules' 'M...

★ 2 Likes 🔍 772 📄 106

Attribute Importance

Notebook to identify key attributes...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'SQL' 'Attribute Importance' 'K...

★ 2 Likes 🔍 530 📄 29

Classification Prediction M...

Example notebook to predict custo...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Classification' 'Prediction' 'De...

★ 3 Likes 🔍 1024 📄 97

Clustering

This notebook shows how to identi...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Clustering' 'K-Means' 'Expect...

★ 1 Likes 🔍 598 📄 35

My First Notebook

Oracle Machine Learning example ...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'SQL' 'Data' 'Graph'

★ 4 Likes 🔍 914 📄 27

Regression

This notebook shows how to predic...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Regression' 'SVM' 'GLM' 'Logi...

★ 1 Likes 🔍 838 📄 31

Statistical Function

Oracle Machine Learning example ...

Author:

Date Added: 2/13/18 11:16 PM

Tags: 'Statistics' 'ANOVA' 'T-test' 'F-...

★ 2 Likes 🔍 356 📄 11

Time Series Forecasting

Oracle Machine Learning supports ...

Author:

Date Added: 9/5/19 4:14 AM

Tags: 'Prediction' 'Time Series' 'ESM'

★ 0 Likes 🔍 1 📄 0

More Info, Quick Starts, HOLs, Docs and Tutorials

www.oracle.com/MachineLearning

<https://twitter.com/CharlieDataMine>

OML demos on YouTube Channel:

<https://www.youtube.com/channel/UCtjWZqDAssJJ5sMi48AI85Q>

Customer References:

<https://www.oracle.com/database/technologies/datawarehouse-bigdata/oml-customers.html>

Changing Role of the DBA: From Oracle Database Professional to Data Scientist in 6 Weeks!

• **YouTube:** <https://youtu.be/jFBMhOapGL8>

• **Blog Series:** <https://blogs.oracle.com/machinelearning/the-changing-role-of-the-dba%3a-from-database-developer-to-data-scientist-in-six-weeks>

Hands-on Labs

OML + Wine + ADW + OAC + APEX + REST HOL Workshop

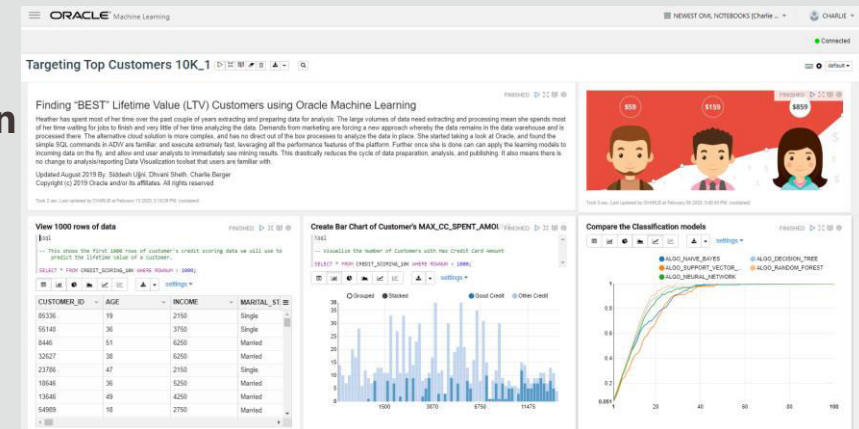
<https://blogs.oracle.com/machinelearning/hands-on-lab%3a-how-to-pick-a-good-wine-for-30%3c-using-oracle-autonomous-database%2c-oracle-machine-learning%2c-apex%2c-oracle-analytics-cloud-and-rest-services>

OML Oracle Data Miner HOL Workshop

<https://blogs.oracle.com/machinelearning/learn-how-to-use-oracle-data-miner-ui-in-45-minutes>

Oracle Machine Learning Notebooks + ADW HOL Workshop

<https://blogs.oracle.com/machinelearning/learn-to-use-oracle-machine-learning-notebooks>



Congratulations!

You are an Oracle Data Scientist!



A diploma template with a gold border and a light gray background with a diagonal line pattern. At the top, the word "Diploma" is written in a large, gold, gothic-style font. Below it, a gold ribbon contains the text "THIS CERTIFICATE IS PRESENTED TO". The Oracle logo is centered below the ribbon. Underneath the logo, the title "Data Scientist" is written in a large, black, sans-serif font. Below the title, the text "LOREM IPSUM DOLOR SIT AMET" is written in a smaller, gold, sans-serif font. A paragraph of Lorem Ipsum text follows: "Obtain a signed certificate. Obtaining a signed certificate involves creating a certificate signing request (CSR) and sending it to a CA in accordance with the CA's enrollment process. After conducting some checks on your company, the CA signs your request, encrypts it with a private key, and sends you a validated certificate. See the instructions provided by the CA for more information." At the bottom, there are two fields: "DATE" on the left and "SIGNATURE" on the right, both with dashed lines above them. In the center, there is a gold seal with an eagle and the text "PREMIUM QUALITY". The entire diploma is surrounded by faint, repeating "ORACLE" watermarks.



Thank You



Charlie Berger

Senior Director, Product Management
Machine Learning, AI and Cognitive Analytics
Charlie.berger@oracle.com 
CharlieDataMine 