

Oracle Tools for Machine Learning

Heli Helskyaho, Matias Helskyaho
@helifromfinland

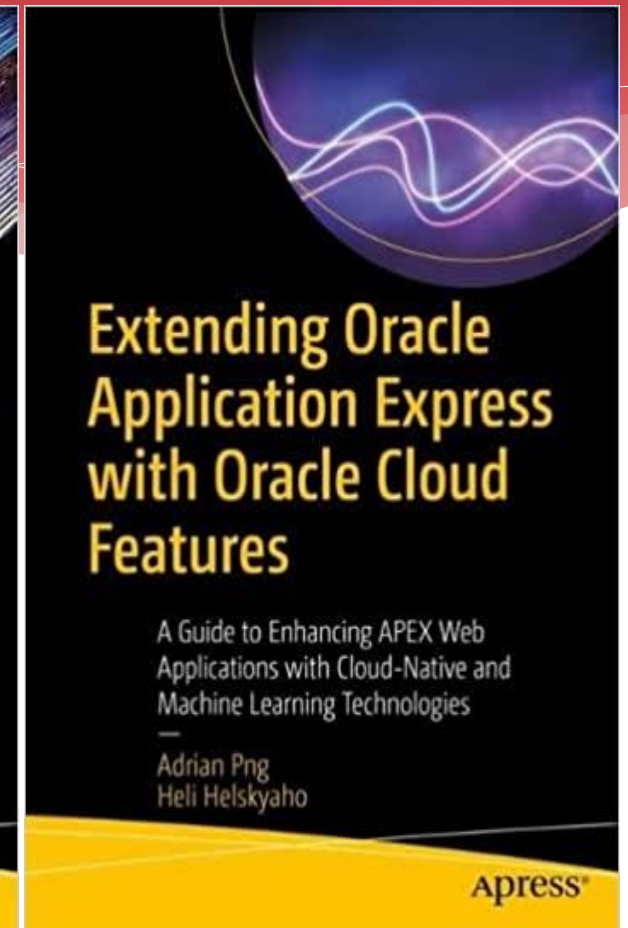
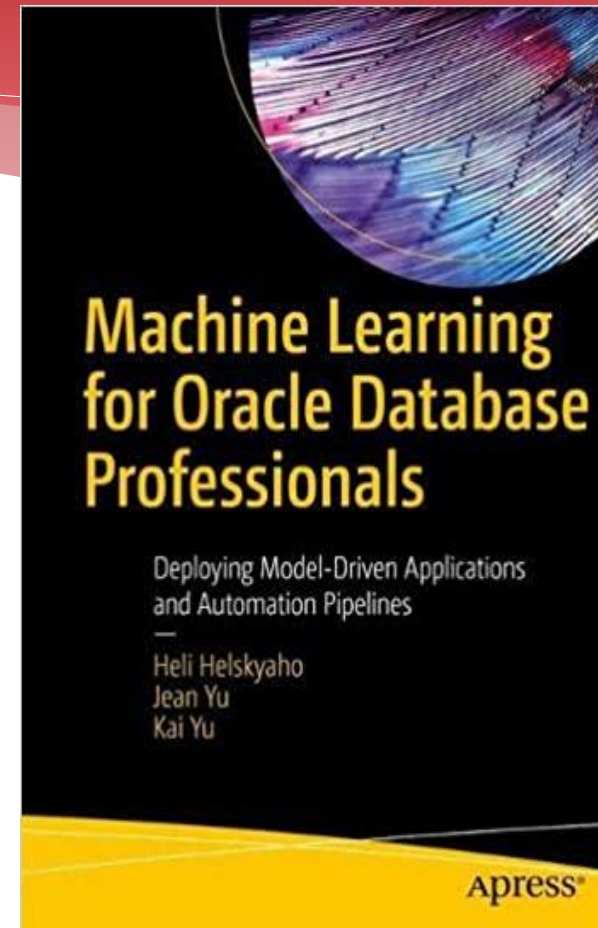
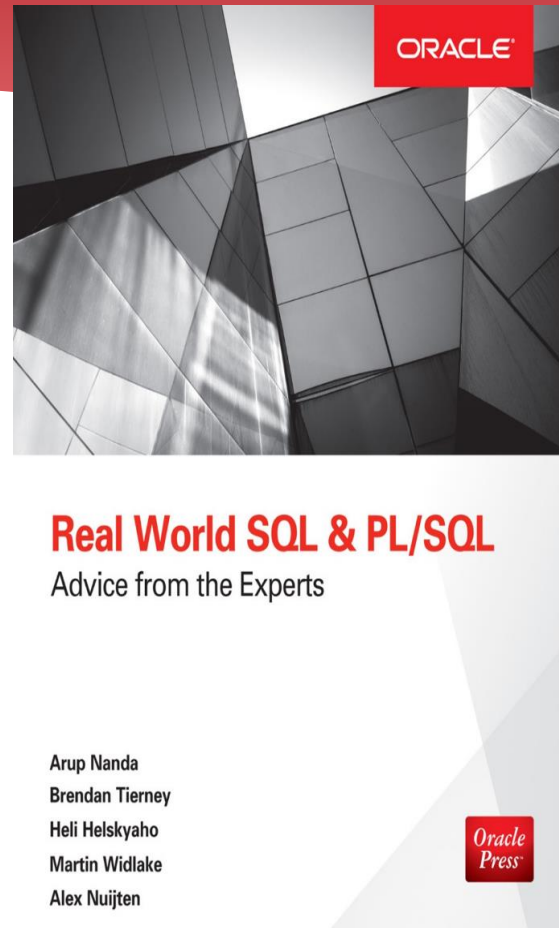
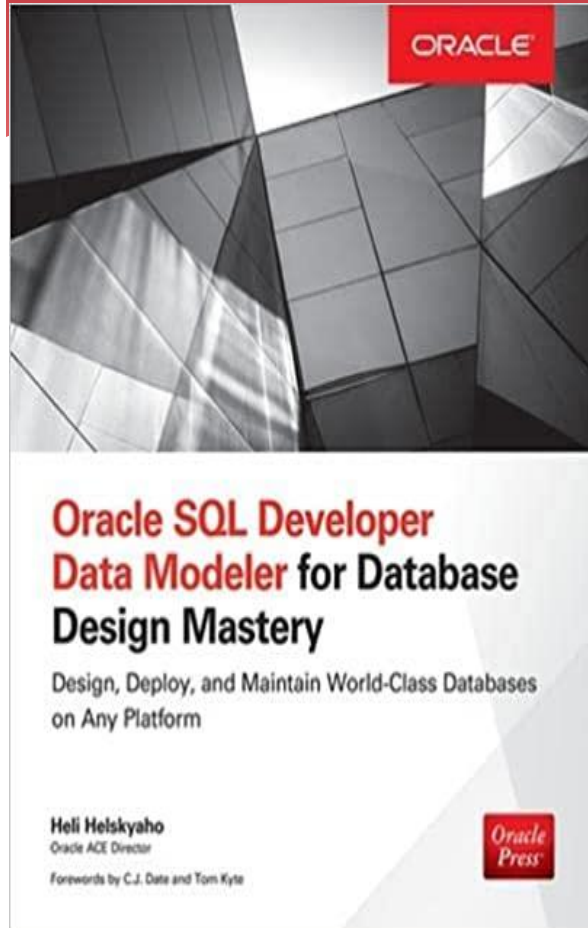
Heli



- * Graduated from University of Helsinki (Master of Science, computer science), currently a doctoral student at University of Helsinki
- * Worked with Oracle products since 1993, worked for IT since 1990
- * Data and Database!
- * CEO for Miracle Finland Oy
- * Oracle ACE Director
- * Public speaker and an author
- * Author of the book Oracle SQL Developer Data Modeler for Database Design Mastery (Oracle Press, 2015), co-author for Real World SQL and PL/SQL: Advice from the Experts (Oracle Press, 2016), Machine Learning for Oracle Database Professionals: Deploying Model-Driven Applications and Automation Pipelines (Apress, 2021), and Extending Oracle Application Express with Oracle Cloud Features: A Guide to Enhancing APEX Web Applications with Cloud-Native and Machine Learning Technologies (Apress, 2022)

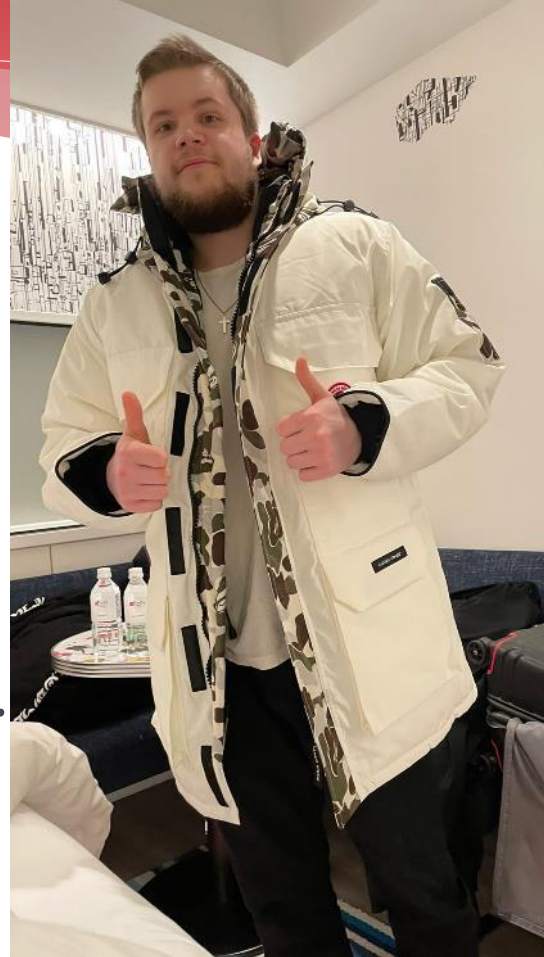


Books



Matias

- * Consultant
 - * Miracle Finland Oy
- * Who am I?
 - * On IT since birth
 - * Professionally a couple of years
 - * OCI, networks, IOT, ML, analytics,...
- * Hobbies
 - * Love learning cool stuff
 - * Playing with tech devices



Oracle ACE
Associate



Oracle offerings for machine learning (examples)

- * Connectors to the database from different programming languages
- * In-database machine learning
 - * OML4SQL
 - * OML4R
 - * OML4Py
- * Oracle Data Science
- * OCI AI Services
- * Oracle Analytics Cloud

Connectors to the database using different programming languages

Oracle DB and Python

* python-oracledb (python -m pip install oracledb --upgrade)

In-database machine learning

- * OML4SQL
- * OML4R
- * OML4Py

A closer look to OML4SQL

Data Dictionary Views for ODM

Table 2-1 Data Dictionary Views for Oracle Data Mining

View Name	Description
<code>ALL_MINING_MODELS</code>	Provides information about all accessible mining models
<code>ALL_MINING_MODEL_ATTRIBUTES</code>	Provides information about the attributes of all accessible mining models
<code>ALL_MINING_MODEL_PARTITIONS</code>	Provides information about the partitions of all accessible partitioned mining models
<code>ALL_MINING_MODEL_SETTINGS</code>	Provides information about the configuration settings for all accessible mining models
<code>ALL_MINING_MODEL_VIEWS</code>	Provides information about the model views for all accessible mining models
<code>ALL_MINING_MODEL_XFORMS</code>	Provides the user-specified transformations embedded in all accessible mining models.

Oracle PL/SQL Packages

- * DBMS_PREDICTIVE_ANALYTICS
 - * Routines for performing *predictive analytics*
- * DBMS_DATA_MINING_TRANSFORMING
 - * Routines for *transforming the data* for mining models
- * DBMS_DATA_MINING
 - * Routines for *creating and managing mining models*

DBMS_PREDICTIVE_ANALYTICS

- * routines that perform an automated data mining known as *predictive analytics*
- * no need to be aware of model building or scoring
 - * All mining activities are handled internally by the procedure.

DBMS_PREDICTIVE_ANALYTICS

- * EXPLAIN
 - * *ranks attributes* in order of influence in explaining the target column
- * PREDICT
 - * *predicts the value of a target column* based on values in the input data
- * PROFILE
 - * *generates rules* that describe the cases from the input data

EXPLAIN

BEGIN

```
DBMS_PREDICTIVE_ANALYTICS.EXPLAIN(  
    data_table_name      => 'beer_training_data',  
    explain_column_name  => 'overall',  
    result_table_name    => 'beer_explain');
```

END;

/

EXPLAIN

Welcome Page BEER_EXPLAIN - Structure HR 12.2 BEER_EXPLAIN				
Columns Data Model Constraints Grants Statistics Triggers Flashback Dependencies Details Partitions Indexes SQL				
Sort.. Filter: Actions...				
	ATTRIBUTE_NAME	ATTRIBUTE_SUBNAME	EXPLANATORY_VALUE	RANK
1	TASTE	(null)	0,1369959018036639046488620435087567350302	1
2	PALATE	(null)	0,1033038106866505612466700086317885929612	2
3	AROMA	(null)	0,08791526829406451506523536256475774282	3
4	APPEARANCE	(null)	0,0597994415746425094084278968669637963851	4
5	IDINDEX	(null)	0,0514743174015554311246566591668580953978	5
6	STYLE	(null)	0,0487182850810693894469596131763997020818	6
7	BEERID	(null)	0,0447651730492723089511095141894142199002	7
8	ABV	(null)	0,0282744648145592050251039059166762689924	8
9	BREWERID	(null)	0,0273217823948269630249887964806943058078	9
10	TEXT	(null)	0	10
11	PROFILENAME	(null)	0	10
12	NAME	(null)	0	10
13	GENDER	(null)	0	10
14	AGEINSECONDS	(null)	0	10
15	BIRTHDAYRAW	(null)	0	10
16	TIMEUNIX	(null)	0	10
17	TIMESTRUCT	(null)	0	10
18	BIRTHDAYUNIX	(null)	0	10

PREDICT

```
DECLARE
p_accuracy NUMBER(10,9);
BEGIN
    DBMS_PREDICTIVE_ANALYTICS.PREDICT(
        accuracy          => p_accuracy,
        data_table_name   => 'beer_train',
        case_id_column_name => 'idindex',
        target_column_name => 'overall',
        result_table_name  => 'Beer_predict');
    DBMS_OUTPUT.PUT_LINE('Accuracy: ' || p_accuracy);
END;
/
```

Accuracy: .24618951 (a measure of improved maximum average accuracy versus a naive model's maximum average accuracy)

PREDICT

IDINDEX	PREDICTION	PROBABILITY
0	2	.42799808967620911
1	3	.53057676003601528
2	3	.51728627054274079
3	4	.45808326842381863
4	4	.47447188708319082
5	4	.64845475978174982
6	4	.65424415909731026
7	3	.45898266110126107
9	4	.51363296269020753

PROFILE

```
BEGIN
    DBMS_PREDICTIVE_ANALYTICS.PROFILE (
        DATA_TABLE_NAME      => 'beer_train',
        TARGET_COLUMN_NAME    => 'overall',
        RESULT_TABLE_NAME     =>
'beer_profile_result');
END;
/
```

PROFILE

BEER_PROFILE_RESULT

PROFILE_ID	RECORD_COUNT	DESCRIPTION
1	50	[SYS.XMLTYPE]
2	100	[SYS.XMLTYPE]
3	34	[SYS.XMLTYPE]
4	462	[SYS.XMLTYPE]
5	566	[SYS.XMLTYPE]
6	279	[SYS.XMLTYPE]
7	2445	[SYS.XMLTYPE]
8	1624	[SYS.XMLTYPE]
9	1522	[SYS.XMLTYPE]
10	653	[SYS.XMLTYPE]
11	8024	[SYS.XMLTYPE]
12	3557	[SYS.XMLTYPE]
...	...	[SYS.XMLTYPE]

Preparing the data

- * This is usually the most difficult and time consuming part of machine learning...

Automatic Data Preparation (ADP)

- * If ADP is active (on), ADP automatically implements the *transformations required by the algorithm*.
- * The transformations are *embedded in the model and automatically executed whenever the model is applied*.

Transforming the data

- * Creating Nested Columns
 - * if you want to include transactional data etc.
- * Converting Column Data Types
 - * Age -> Child, Adult
- * Business and Domain-Sensitive Transformations
 - * Date of birth -> age
- * Text Transformation (a text column must be in a table, not a view)
- * ...
- * Binning
- * normalization

DBMS_DATA_MINING_TRANSFORM

A Record Transformation

- * Each transform_rec specifies the transformation instructions for an attribute.

```
TYPE transform_rec IS RECORD (  
    attribute_name      VARCHAR2 (30) ,  
    attribute_subname   VARCHAR2 (4000) ,  
    expression          EXPRESSION_REC ,  
    reverse_expression  EXPRESSION_REC ,  
    attribute_spec      VARCHAR2 (4000) ) ;
```

A Transformation List

- * is defined as a table/*collection of transformation records* (transform_rec)
- * A list can be created using:
 - * The SET_TRANSFORM procedure in DBMS_DATA_MINING_TRANSFORM
 - * The STACK interface in DBMS_DATA_MINING_TRANSFORM
 - * The GET_MODEL_TRANSFORMATIONS and GET_TRANSFORM_LIST functions in DBMS_DATA_MINING

User-specified Transformation List

- * can be *embedded a Transformation List in the model* and reapply whenever the model is applied
- * you do not need to specify them for the test or scoring data sets because the transformation instructions are embedded in the model

Embedding Transformations in a Model

* You can create a transformation list and pass it to DBMS_DATA_MINING.CREATE_MODEL:

```
PROCEDURE create_model(  
    model_name           IN VARCHAR2,  
    mining_function      IN VARCHAR2,  
    data_table_name      IN VARCHAR2,  
    case_id_column_name  IN VARCHAR2,  
    target_column_name   IN VARCHAR2 DEFAULT NULL,  
    settings_table_name  IN VARCHAR2 DEFAULT NULL,  
    data_schema_name     IN VARCHAR2 DEFAULT NULL,  
    settings_schema_name IN VARCHAR2 DEFAULT NULL,  
    xform_list           IN TRANSFORM_LIST DEFAULT NULL);
```

ADP and Transformation List

- * If you enable ADP and you specify a transformation list
 - * the transformation list is *embedded* with the automatic, system-generated transformations
 - * is executed *before* the automatic transformations

Creating a Model

Creating a Model

- * Choose the mining function
- * Choose the algorithm
- * Create and populate the settings table

Choose the Mining Function

- * Supervised Learning
 - * Regression
 - * Classification
 - * Feature Selection
- * Unsupervised Learning
 - * Anomaly Detection
 - * Clustering
 - * Association
 - * Feature Extraction

Choosing the Algorithm

- * Decision Tree (classification)
- * Naive Bayes (classification)
- * Generalized Linear Models (regression and classification)
- * Support Vector Machines (classification, regression, and anomaly detection)
- * k-Means (clustering)
- * O-Cluster (clustering)
- * Minimum Description Length (for calculating attribute importance)
- * Apriori (for calculating association rules)
- * Non-Negative Matrix Factorization, NMF (feature extraction)
- * ... Each version brings more algorithms to choose from...

Choosing the Algorithm

- * Decision Tree (classification)
- * Naive Bayes (classification)
- * Generalized Linear Models (regression and classification)
- * Support Vector Machines (classification, regression, and anomaly detection)
- * k-Means (clustering)
- * O-Cluster (clustering)
- * Minimum Description Length (for calculating attribute importance)
- * Apriori (for calculating association rules)
- * Non-Negative Matrix Factorization, NMF (feature extraction)
- * ... Each version brings more algorithms to choose from...

A Settings table

```
CREATE TABLE Beer_settings_DT (  
  setting_name  VARCHAR2(30),  
  setting_value VARCHAR2(4000));  
  
BEGIN  
  INSERT INTO Beer_settings_DT VALUES  
    (dbms_data_mining.algo_name, dbms_data_mining.algo_decision_tree);  
  ...  
END;  
/
```

Other possible settings

- * Cost table and matrix (Decision Tree model)
- * Prior Probabilities (Naive Bayes)
- * Class Weights (Logistic Regression or Support Vector Machine)
- * ...

Create a new model

```
BEGIN
  DBMS_DATA_MINING.CREATE_MODEL(
    model_name          => 'Beer_DT',
    mining_function      => dbms_data_mining.classification,
    data_table_name     => 'Beer_training_data',
    case_id_column_name => 'IDIndex',
    target_column_name  => 'Overall',
    settings_table_name => 'Beer_settings_DT');
END;
/
```


Model Signature

- * The set of data attributes that are used to build a model

Model Signature

```
SELECT attribute_name, attribute_type  
FROM TABLE(DBMS_DATA_MINING.GET_MODEL_SIGNATURE('BEER_DT'))  
ORDER BY attribute_name;
```

	ATTRIBUTE_NAME	ATTRIBUTE_TYPE
1	ABV	NUMBER
2	BREWERID	NUMBER
3	STYLE	VARCHAR2

Partitioned Model

- * enables to build and manage models tailored to independent slices of data
- * a partitioning key (comma-separated list of one or more columns) is required
- * the partitioning key is set through settings table
- * the partition columns must be part of the USING clause when scoring

Testing and Evaluating the model

- * Test: the model with new data (known input, known output)
- * Evaluation: depends on the chosen metrics

Apply the Model

BEGIN

```
DBMS_DATA_MINING.APPLY (  
  model_name      => 'beer_DT',  
  data_table_name => 'beer_test',  
  case_id_column_name => 'idindex',  
  result_table_name => 'beer_result_table_DT');
```

END;

/

What to measure?

- * Number of positives, number of negatives, number of true positives, number of false positives, number of true negatives, number of false negatives
- * Portion of positives, portion of negatives
- * Class ratio
- * Accuracy, Error rate
- * ROC curve, coverage curve,
- * ...
- * It all depends on how we define the performance for the answer to our question (experiment): *experimental objective*

BEER_RESULT_TABLE_DT

Indexes Model Constraints Grants Statistics UI Defaults Triggers Dependencies SQL REST

10 Rows

IDINDEX	PREDICTION	PROBABILITY	COST
13803	4	.61222339304531082	.38777660695468918
13803	5	.25594192717480391	.74405807282519609
13803	3	.11017445264020606	.88982554735979391
13803	2	.01803067556492214	.98196932443507789
13803	1	.0036295515747570544	.99637044842524292
13803	0	0	1
13960	4	.61222339304531082	.38777660695468918
13960	5	.25594192717480391	.74405807282519609
13960	3	.11017445264020606	.88982554735979391

Evaluating

- * COMPUTE_CONFUSION_MATRIX Procedure
- * COMPUTE_LIFT Procedure
- * COMPUTE_ROC Procedure
- * RANK_APPLY Procedure

Deployment

- * Deployment is implementing the models in the target environment
- * Moving a model from the database where it was built to the database where it will be used (export/import)
- * Or creating an endpoint

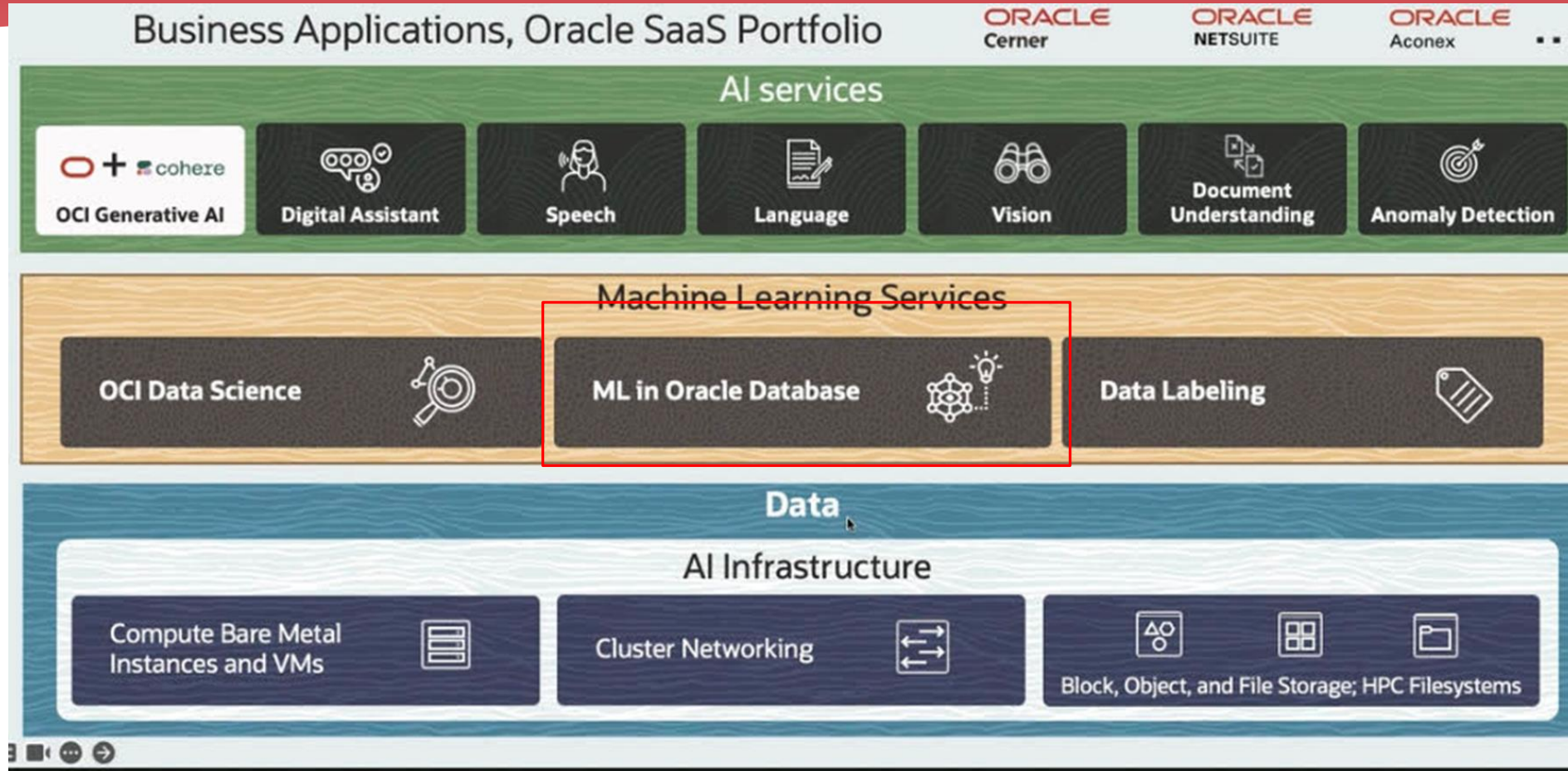
Real-time scoring

What is the probability for beer 43548 to get overall 5?

```
SELECT PREDICTION_PROBABILITY(Beer_DT, 5 USING *) beer_overall_prob  
FROM beer_test  
WHERE idindex = 43658;  
1.5087463556851313E-001
```

```
SELECT PREDICTION_PROBABILITY(Beer_DT, 5 USING STYLE)  
beer_overall_prob  
FROM beer_test  
WHERE idindex = 43658;  
2.4140018157974377E-001
```

Machine Learning in OCI



ORACLE Cloud

Search resources, services, documentation, and Marketplace

Germany Central (Frankfurt)

Overview > Autonomous Database > Autonomous Database details

ADW

AVAILABLE

Announcements

- New Feature: Built-in encryption / decryption support for exporting and loading data in Autonomous Database. [Learn more.](#)
- Learn about other [new features in Autonomous Database here.](#)

OMLDB

Primary

Database actions

Database

Autonomous Database information

Edit tool configuration

Tools

ORACLE Cloud

Search resources, services, documentation, and Marketplace

Germany Central (Frankfurt)

Database actions	<p>Oracle Autonomous Database. Learn more.</p> <p>Access URL:</p> <p>https://l /ords/sql-</p> <p>developer Copy</p>	Enabled
Graph Studio	<p>Graph Studio automates the creation of knowledge (RDF) and property graphs and includes interactive tooling for query, analysis, and visualization of these graphs in the Autonomous Database. You must log in as a graph-enabled user to access Graph Studio. Create this user in database actions. Learn more.</p> <p>Access URL:</p> <p>https://l /graphstudio/</p> <p>Copy</p>	Enabled
Oracle Machine Learning user interface	<p>The Oracle Machine Learning user interface provides immediate access to the Oracle Machine Learning components and functionality on Autonomous Database, including OML Notebooks, OML AutoML UI, OML Models, and template example notebooks. Learn more.</p> <p>Access URL:</p> <p>https://l /oml/ Copy</p>	Enabled
Data transforms	<p>Oracle data transforms allows you to design graphical data transformations in the form of data flows and workflows. The data flows define how the data is moved and transformed between different systems, while the workflows define the sequence in which the data flows are executed. Learn more.</p> <p>Access URL:</p> <p>https://l /nodi/ Copy</p>	Enabled

Terms of Use and Privacy

Cookie Preferences

Copyright © 2024. Oracle and/or its affiliates. All rights reserved.



SIGN IN

Database name:

DE 3

Sign in with your Oracle Machine Learning Database User credentials

USERNAME

PASSWORD

Sign In

Important changes

- Scratchpad notebooks will now be opened using the OML Notebooks EA interface and appear in your Notebooks EA listing.
- Notebooks generated for a model from an AutoML experiment or created from templates (Example, Shared or Personal) will appear in your Notebooks EA listing and use the OML Notebooks EA interface.

> How Do I?

Quick Actions



AutoML

Create and run
AutoML Experiments



Models

Manage and Deploy
Machine Learning
Models



Data Monitors

Monitor Data Drift



Model Monitors

Monitor Model Drift



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 Notebooks

[See More...](#)

Recent Experiments

Recent Activities



Classification Prediction Model



Predicting Target Customers using Classification

In this notebook, we predict customers most likely to be positive responders to an Affinity Card loyalty program. High Affinity Card responders (target value = 1) are defined as those customers who when given a loyalty or affinity card hyper-respond i.e. they increase their purchasing higher than the Affinity Card program's offered discount percentage. This notebook builds and applies classification models (decision tree) using the SH schema data. All processing occurs inside Oracle Autonomous Database.

Updated August 2019 By Charlie Berger, Dhvani Sheth, Siddesh C Prabhu Dev Ujjini, Mark Hornick, Marcos Arancibia
Copyright (c) 2019 Oracle and/or its affiliates. All rights reserved.

Took 2 sec. Last updated by USER07 at August 16 2019, 1:12:03 AM.

FINISHED



Took 0 sec. Last updated by USER07 at August 14 2019, 12:25:01 AM.

For more information, check the Oracle ADWC Documentation <https://docs.oracle.com/en/cloud/paas/autonomous-data-warehouse-cloud/index.html>

Oracle Machine Learning folder on Oracle on Github <https://github.com/oracle/oracle-db-examples/tree/master/machine-learning>

Oracle Machine Learning web page: <https://www.oracle.com/database/technologies/datawarehouse-bigdata/machine-learning.html>

Introducing Oracle Machine Learning blog post: <https://blogs.oracle.com/datamining/introducing-oracle-machine-learning-sql-notebooks-for-the-oracle-autonomous-data-warehouse-cloud>

Took 0 sec. Last updated by USER07 at August 14 2019, 12:25:02 AM. (outdated)

FINISHED

Display the SH.SUPPLEMENTARY_DEMOGRAPHICS data

```
%sql
SELECT * FROM SH.SUPPLEMENTARY_DEMOGRAPHICS;
```



CUST_ID	EDUCATION	OCCUPATION	HOUSEHOLD_SIZE	YRS_RESIDENCE	AFFINITY_CARD	BULK_PACK_DISKETTES	FLAT_PANEL_MONITOR	HOME_THEATER_PACKAGE	BOOKKEEPING_APPLICATION
102547	10th	Other	1	0	0	1	1	0	0

<https://docs.oracle.com/en/cloud/paas/autonomous-data-warehouse-cloud/index.html>

Machine Learning Process, Example

1. Defining the Task, understanding the Task
2. Collecting the data, understanding the data
3. Attributes (Features, Columns)
4. Preparing the data/Transforming the data
5. Creating models
6. Evaluating models
7. Scoring and Deployment
8. (monitoring the data and the model)

Machine Learning words

- * Exploratory data analysis
- * Data visualization
- * Feature selection
- * Feature engineering
- * Algorithm selection
- * Feature encoding
- * Hyperparameter tuning
- * Model evaluation
- * Model interpretation and explainability...

AutoML

Important changes

- Scratchpad notebooks will now be opened using the OML Notebooks EA interface and appear in your Notebooks EA listing.
- Notebooks generated for a model from an AutoML experiment or created from templates (Example, Shared or Personal) will appear in your Notebooks EA listing and use the OML Notebooks EA interface.

> How Do I?

Quick Actions



AutoML

Create and run
AutoML Experiments



Models

Manage and Deploy
Machine Learning
Models



Data Monitors

Monitor Data Drift



Model Monitors

Monitor Model Drift



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 Notebooks

[See More...](#)

Recent Experiments

Recent Activities





AutoML Experiments

Filter

<input type="checkbox"/>	Name ^	Comment ◇	Created On ◇	Created By ◇	Updated By ◇	Status ◇
<input type="checkbox"/>	AffinityCard1	The First experiment on affinity card	11/4/23, 12:14 PM			Completed
<input type="checkbox"/>	Bicycle prediction v1	Bicycle prediction v1	7/25/23, 2:17 PM			Completed
<input type="checkbox"/>	Bicycle Prediction v2	Bicycle Prediction v2	7/25/23, 4:00 PM			Completed
<input type="checkbox"/>	CustomerModel		11/28/23, 9:55 AM			Failed

Create Experiment

▶ Start ▾ ⬆ Save Cancel

Name *

Comments

Data Source *



Prediction Type *

Predict *

Case ID

▶ Additional Settings

▴ Features

↻ Refresh

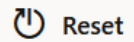
Search...

<input type="checkbox"/> Name	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
-------------------------------	------	---------------	-----------------	-----	-----	------	---------

No data to display.

(For Classification)

Additional Settings



Reset

Maximum Top Models

5



Maximum Run Duration (Hours)

8



Database Service Level

Low



Model Metric

Balanced Accuracy



Model Metric

Balanced Accuracy



Accuracy

Balanced Accuracy

ROC AUC

F1

Precision

Recall

Algorithms



Name ^



Decision Tree



Generalized Linear Model



Generalized Linear Model (Ridge Regression)



Naive Bayes



Neural Network



Random Forest



Support Vector Machine (Gaussian)



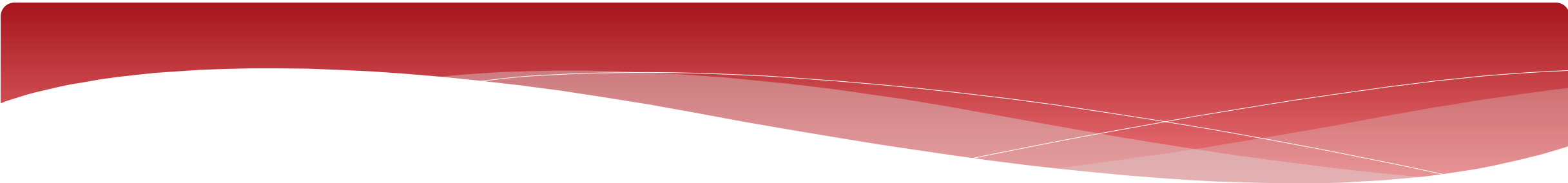
Support Vector Machine (Linear)

Features

Refresh

Search...

<input type="checkbox"/> Name	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
<input checked="" type="checkbox"/> TARGET	NUMBER	0	2	0	1	0.63	0.68
<input checked="" type="checkbox"/> area error	NUMBER	0	518	6.802	542.2	43.22	56.54
<input checked="" type="checkbox"/> compactness error	NUMBER	0	535	0.0023	0.1354	0.03	0.02
<input checked="" type="checkbox"/> concave points error	NUMBER	0	503	0	0.0528	0.01	0.01
<input checked="" type="checkbox"/> concavity error	NUMBER	0	543	0	0.396	0.03	0.04
<input checked="" type="checkbox"/> fractal dimension error	NUMBER	0	550	0.0009	0.0298	0	0.01
<input checked="" type="checkbox"/> mean area	NUMBER	0	544	143.5	2501	680.01	393.06
<input checked="" type="checkbox"/> mean compactness	NUMBER	0	523	0.0194	0.3454	0.11	0.06



Create Experiment

Name *

MyFirstAutoML

Start

Save

Cancel


Faster Results

Better Accuracy

MyFirstAutoML

✓ Completed

▶ Start ▼

▶ Experiment Settings  Edit

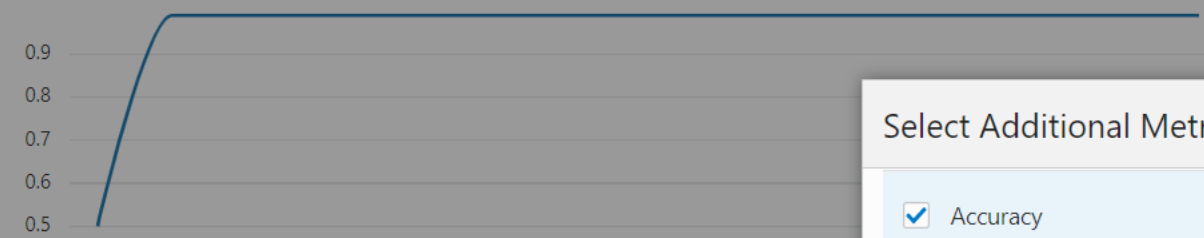
Balanced Accuracy



Leader Board

Deploy	Rename	Create Notebook	Metrics	
Algorithm			Model Name	Balanced Accuracy
Support Vector Machine (Gaussian)			svmg_fcd3cb7701	0.9896
Neural Network			nn_6d09b27a4a	0.9896
Generalized Linear Model (Ridge Regress...			glm_r_71fd0d3403	0.9792
Support Vector Machine (Linear)			svml_801e8b40f9	0.9792
Generalized Linear Model			glm_86fbdeec63	0.9715

Balanced Accuracy



Leader Board

Deploy

Rename

Create Notebook

Metrics

Algorithm

Model Name

Support Vector Machine (Gaussian)

svmg_fcd3cb7701

Neural Network

nn_6d09b27a4a

Generalized Linear Model (Ridge Regress...

glm_r_71fd0d3403

Support Vector Machine (Linear)

svml_801e8b40f9

Generalized Linear Model

glm_86fbdeec63

☒ Precision

☒ Recall

☐ ROC AUC

☐ F1 Micro

☐ F1 Macro

Recall

Accuracy

0.9792

0.9912

0.9792

0.9912

0.9583

0.9823

0.9583

0.9823

0.9583

0.9735

Select Additional Metrics

☒ Accuracy

☐ F1

☒ Precision

☒ Recall

☐ ROC AUC

☐ F1 Micro

☐ F1 Macro

MyFirstAutoML

✓ Completed

▶ Start ▼

▶ Experiment Settings  Edit

Balanced Accuracy



Leader Board

Deploy	Rename	Create Notebook	Metrics				
Algorithm	Model Name		Balanced Accuracy ▼	Precision	Recall	Accuracy	F1
Support Vector Machine (Gaussian)	svmg_fcd3cb7701		0.9896	1.0000	0.9792	0.9912	0.9895
Neural Network	nn_6d09b27a4a		0.9896	1.0000	0.9792	0.9912	0.9895

Rename Model - svmg_fcd3cb7701

New Model Name *

SVM Model

Enter an alphanumeric Name (max 123 characters)

OK


Cancel



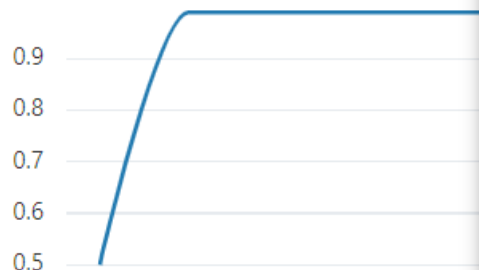
The selected model has been renamed

<- Experiments

MyFirstAutoML

► Experiment Settings  Edit

Balanced Accuracy



Leader Board

Deploy Rename Create Notebook

Algorithm

Support Vector Machine (Gaussian)

Neural Network

Generalized Linear Model (Ridge Regression)

Model Detail - SVM_Model



Prediction Impacts

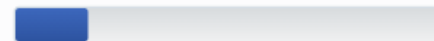
Confusion Matrix

Name

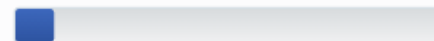


Prediction Impact

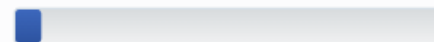
worst texture



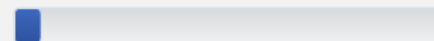
compactness error



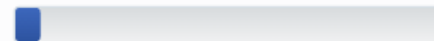
worst area



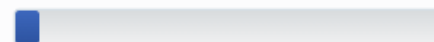
mean concavity



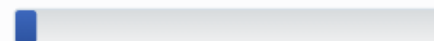
worst radius



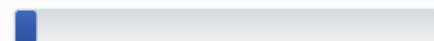
worst concavity



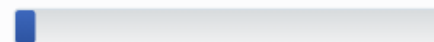
fractal dimension error



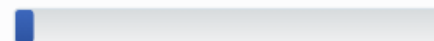
worst concave points



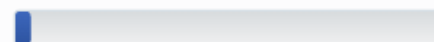
worst perimeter



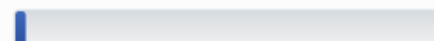
worst smoothness



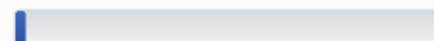
mean smoothness



concave points error



texture error



The selected model has been renamed

<- Experiments

MyFirstAutoML

► Experiment Settings  Edit

Balanced Accuracy

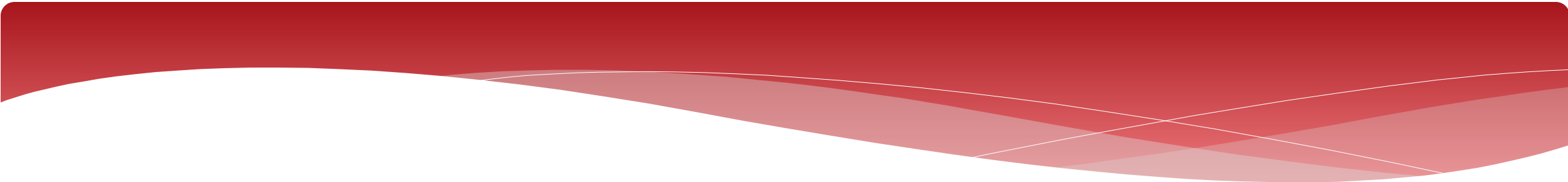


Model Detail - SVM_Model

Prediction Impacts

Confusion Matrix

	Predicted: 0	Predicted: 1
Actual: 0	47	1
Actual: 1	0	65



Leader Board

Deploy	Rename	Create Notebook	Metrics			
Algorithm	Model Name	Balanced Accuracy ▼	Precision	Recall	Accuracy	F1
Support Vector Machine (Gaussian)	SVM_Model	0.9896	1.0000	0.9792	0.9912	0.9895
Neural Network	nn_6d09b27a4a	0.9896	1.0000	0.9792	0.9912	0.9895
Generalized Linear Model (Ridge Regress...	glmr_71fd0d3403	0.9792	1.0000	0.9583	0.9823	0.9787
Support Vector Machine (Linear)	svml_801e8b40f9	0.9792	1.0000	0.9583	0.9823	0.9787
Generalized Linear Model	glm_86fbdeec63	0.9715	0.9787	0.9583	0.9735	0.9684

Deploy Model - SVM_Model



Name *

SVM Model

URI *

Version *

Namespace

☐ Shared




OK


Cancel

Create a Notebook


Leader Board


Deploy	Rename	Create Notebook	Metrics				
Algorithm	Model Name		Balanced Accuracy ▼	Precision	Recall	Accuracy	F1
Support Vector Machine (Gaussian)	SVM_Model		0.9896	1.0000	0.9792	0.9912	0.9895
Neural Network	nn_6d09b27a4a		0.9896	1.0000	0.9792	0.9912	0.9895
Generalized Linear Model (Ridge Regress...	glmr_71fd0d3403		0.9792	1.0000	0.9583	0.9823	0.9787
Support Vector Machine (Linear)	svml_801e8b40f9		0.9792	1.0000	0.9583	0.9823	0.9787
Generalized Linear Model	glm_86fbdeec63		0.9715	0.9787	0.9583	0.9735	0.9684


 Home


 Project




 Notebooks


 Notebooks EA


 AutoML Experiments


 Monitoring




 Data

 Models

 Models

 Templates



 Jobs

MyFirstAutoMLNotebook



1 keyboard settings default ▼

Oracle Machine Learning AutoML UI - Experiment - Generated Notebook

READY ▶ 🔍 📖 ⚙️

Get proxy object for selected data

READY ▶ 🔍 📖 ⚙️

```
%python

import oml

columns = '"area error"', '"compactness error"', '"concave points error"', '"concavity error"', '"fractal dimension error"', '"mean area"', '"mean compactness"', '"mean concave points"', '"mean concavity"',
          '"mean fractal dimension"', '"mean perimeter"', '"mean radius"', '"mean smoothness"', '"mean symmetry"', '"mean texture"', '"perimeter error"', '"radius error"', '"smoothness error"', '"symmetry error"',
          '"texture error"', '"worst area"', '"worst compactness"', '"worst concave points"', '"worst concavity"', '"worst fractal dimension"', '"worst perimeter"', '"worst radius"', '"worst smoothness"',
          '"worst symmetry"', '"worst texture"', '"TARGET"'
schema="HELI ML"
table="BREASTCANCER"

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

build_data = oml.sync(query=query)
x_train = build_data
```

Prepare training data

READY ▶ 🔍 📖 ⚙️

```
%python

import oml

X_train = build_data[:,['area error', 'compactness error', 'concave points error', 'concavity error', 'fractal dimension error', 'mean area', 'mean compactness', 'mean concave points', 'mean concavity', 'mean fractal dimension',
                        'mean perimeter', 'mean radius', 'mean smoothness', 'mean symmetry', 'mean texture', 'perimeter error', 'radius error', 'smoothness error', 'symmetry error', 'texture error', 'worst area',
                        'worst compactness', 'worst concave points', 'worst concavity', 'worst fractal dimension', 'worst perimeter', 'worst radius', 'worst smoothness', 'worst symmetry', 'worst texture']]
v_train = build_data[:, 'TARGET']
```

Important changes

- Scratchpad notebooks will now be opened using the OML Notebooks EA interface and appear in your Notebooks EA listing.
- Notebooks generated for a model from an AutoML experiment or created from templates (Example, Shared or Personal) will appear in your Notebooks EA listing and use the OML Notebooks EA interface.

> How Do I?

Quick Actions



AutoML

Create and run
AutoML Experiments



Models

Manage and Deploy
Machine Learning
Models



Data Monitors

Monitor Data Drift



Model Monitors

Monitor Model Drift



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 Notebooks

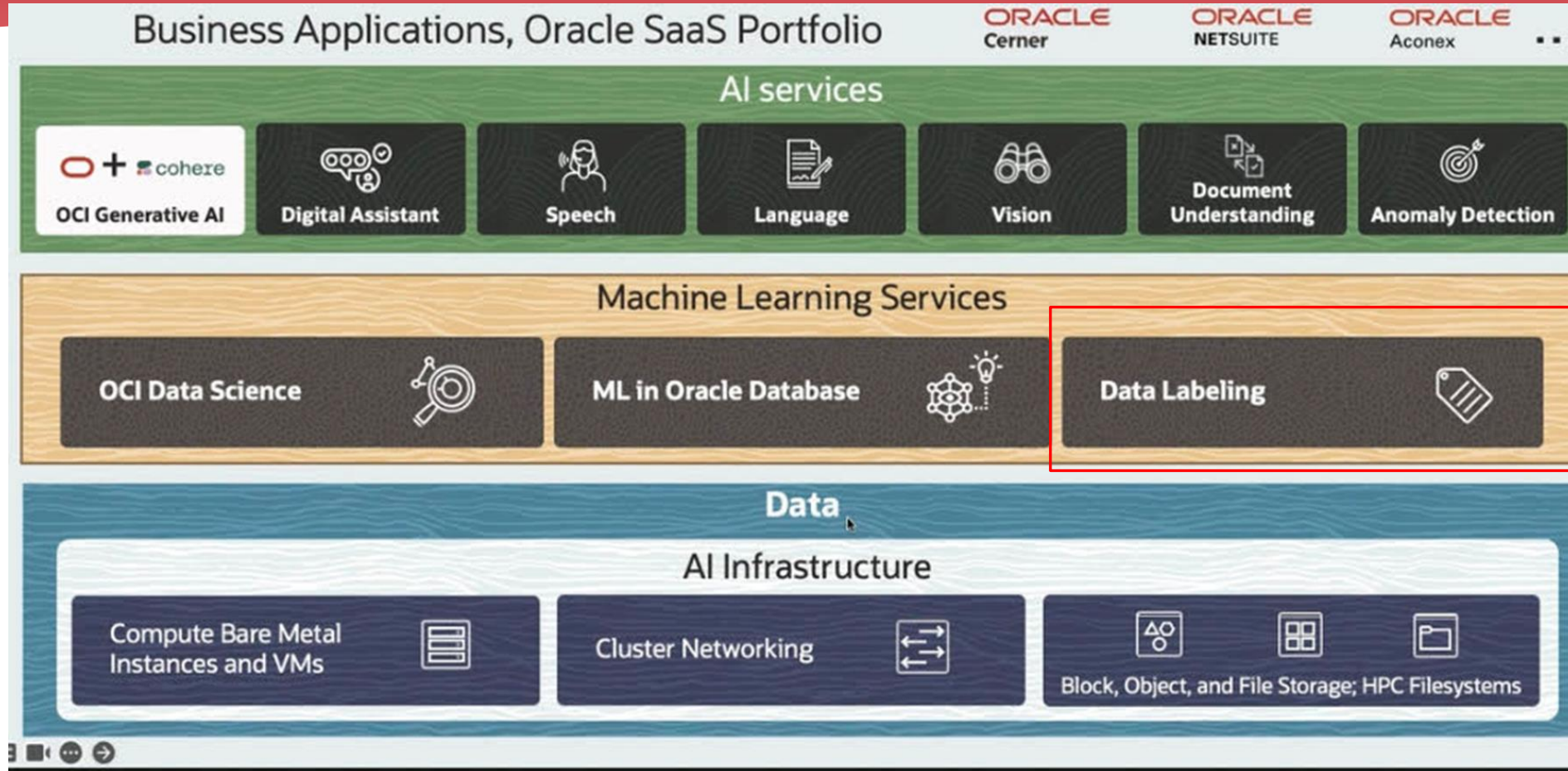
[See More...](#)

Recent Experiments

Recent Activities



Machine Learning in OCI



Data Labeling

The screenshot displays the Oracle Cloud console interface. At the top, the header includes the Oracle Cloud logo, a search bar, and the region 'Germany Central (Frankfurt)'. The left sidebar contains a navigation menu with categories like Home, Compute, Storage, Networking, Oracle Database, Databases, Analytics & AI (highlighted), Developer Services, Identity & Security, Observability & Management, Hybrid, Migration & Disaster Recovery, Billing & Cost Management, and Governance & Administration. The main content area is titled 'Analytics & AI' and is organized into columns. The 'Machine Learning' column has 'Data Labeling' highlighted with a red rectangle. Other columns include 'Analytics', 'Data Lake', 'Messaging', 'AI Services', 'Media Services', and 'Related services'. The footer contains links for 'Terms of Use and Privacy' and 'Cookie Preferences', and a copyright notice for 2023.

ORACLE Cloud Search resources, services, documentation, and Marketplace Germany Central (Frankfurt)

Q Search

Analytics & AI

Analytics
Analytics Cloud
Fusion Analytics Warehouse

Data Lake
Big Data Service
Data Catalog
Data Integration
Data Flow

Messaging
Streaming
Service Connector Hub

Machine Learning
Data Science
Data Labeling

AI Services
Language
Speech
Vision
Document Understanding
Anomaly Detection
Digital Assistant

Media Services
Media Flow
Media Flow Jobs
Media Streams

Related services
Autonomous Database
MySQL
Oracle NoSQL Database
Object Storage & Archive Storage
Application Performance Monitoring
Logging Analytics
Operations Insights
Logging

Help
Analytics Cloud Documentation
Big Data Documentation
Data Science Documentation
Digital Assistant Documentation
REST APIs

Terms of Use and Privacy Cookie Preferences

Copyright © 2023, Oracle and/or its affiliates. All rights reserved.

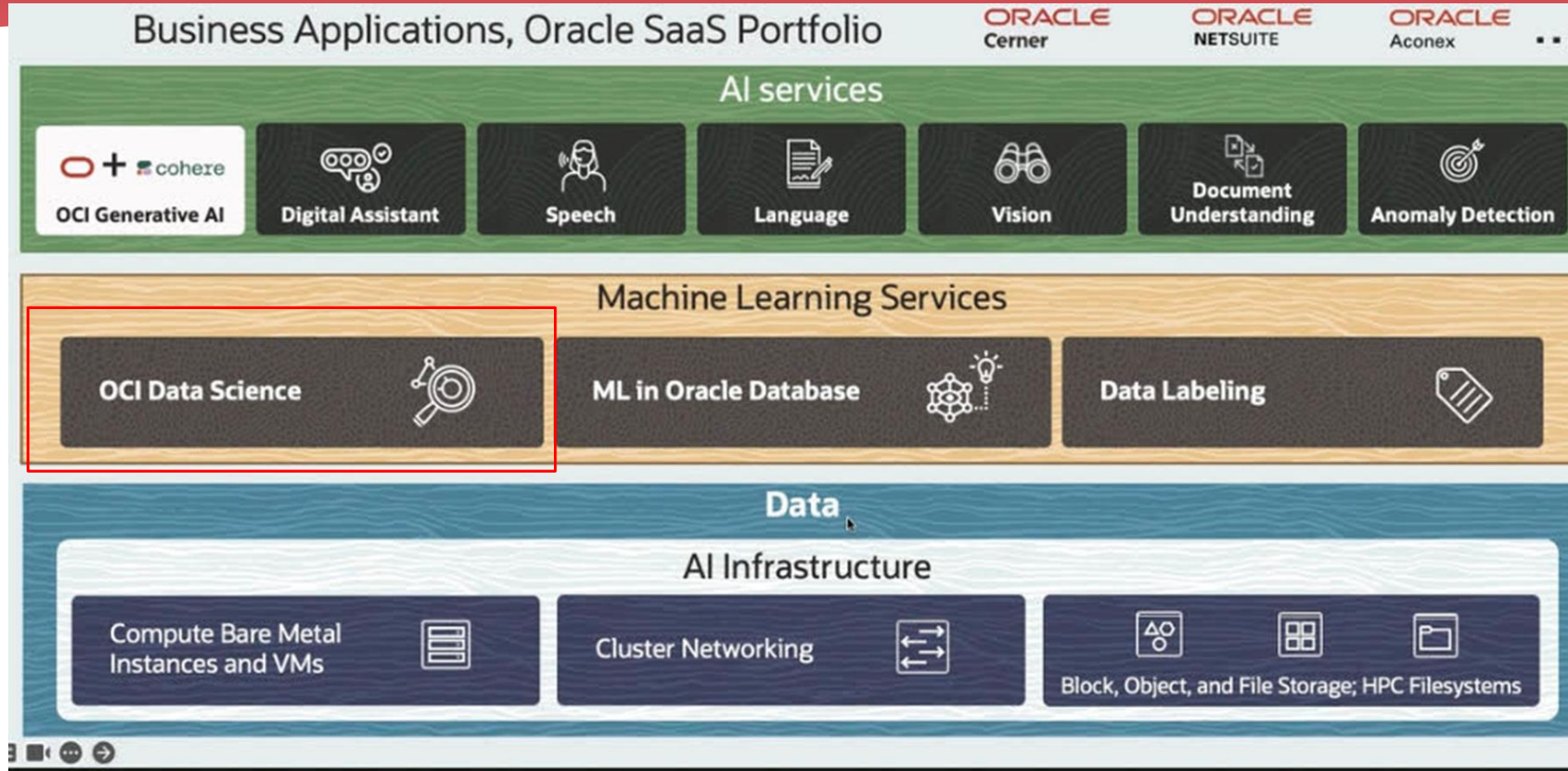
Labeled data

- * Needed for Supervised learning
- * Labeling is identifying properties (labels) for images/text data and annotating (labeling) them with those properties
- * Single/multiple labels
- * The data labeled can be used with AI Service Custom models

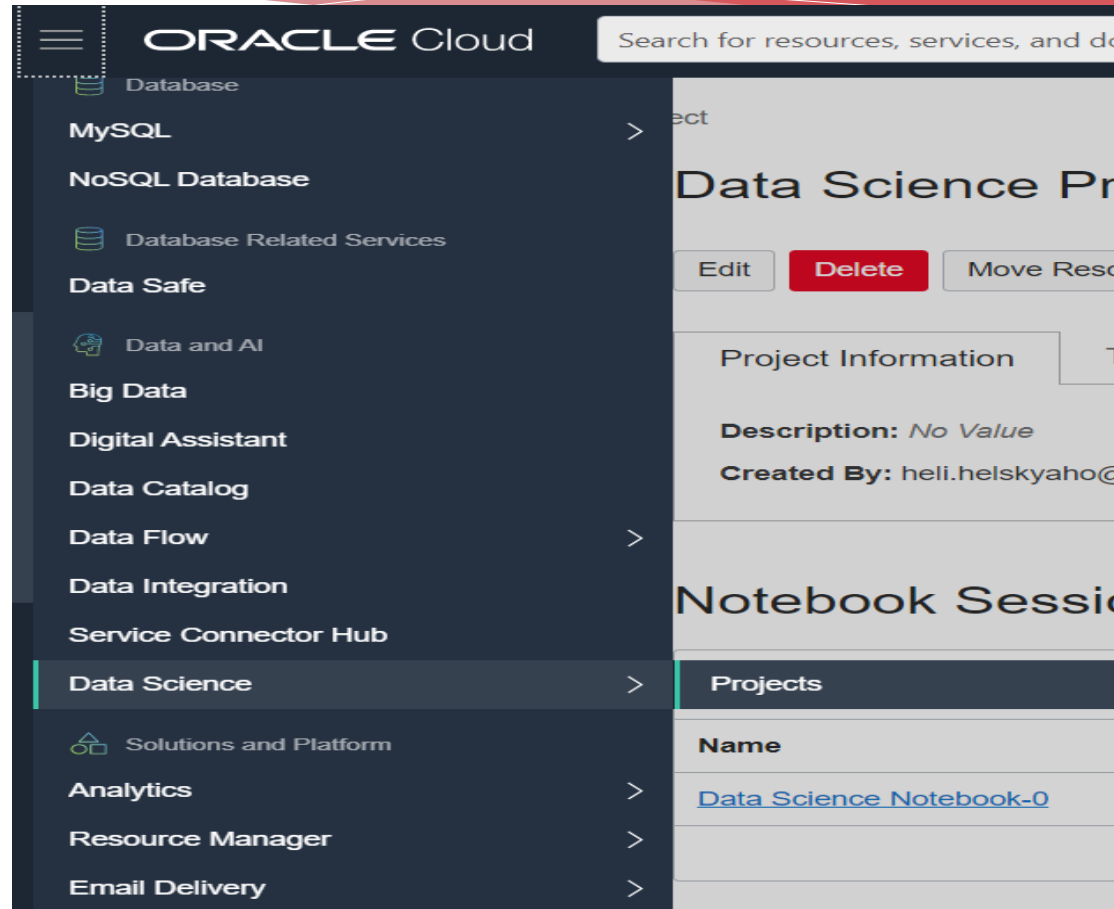
OCI Data Labeling Service

- * Images
 - * Classification, object detection,...
- * Text (evaluations, feedback,...)
 - * Classification, highlighting words or span of words within a text (can be used to train a custom NLP model)
- * Documents (PO, Invoice, Receipt,...)
 - * Classification


Oracle Machine Learning in OCI










OCI Data Science



Project

 ORACLE Cloud

Search for resources, services, and documentation

Netherlands Northwest (Amsterdam)       

Data Science

Projects

List Scope

COMPARTMENT

hellhelskyaho (root)

Filters

STATE

Any state

Tag Filters

[add](#) | [clear](#)


no tag filters applied

Create Project








Name	Status	Description	Created By	Created On
Data Science Project	● Active		hell.helskyaho@miracleoy.fi	Tue, Sep 15, 2020, 13:19:08 UTC

Showing 1 Project < Page 1 >


Notebook session



Search for resources, services, and documentation

Netherlands Northwest (Amsterdam)       

Data Science » [Projects](#) » Data Science Project



ACTIVE

Data Science Project

EditDeleteMove ResourceAdd Tags

Project InformationTags

Description: *No Value*OCID: ...vy3pz7xq [Show](#) [Copy](#)

Created By: heli.helskyaho@miracleoy.fiCreated On: Tue, Sep 15, 2020, 13:19:08 UTC

Resources

Notebook Sessions

Models

List Scope

COMPARTMENT

helihelskvaho (root)

[Terms of Use and Privacy](#) [Cookie Preferences](#)

Notebook Sessions *in helihelskyaho (root) Compartment*

Create Notebook Session

Name	Status	Compute Instance Shape	Created By	Created On
Data Science Notebook-0	● Active	VM.Standard2.2	heli.helskyaho@miracleoy.fi	Tue, Sep 15, 2020, 13:19:10 UTC

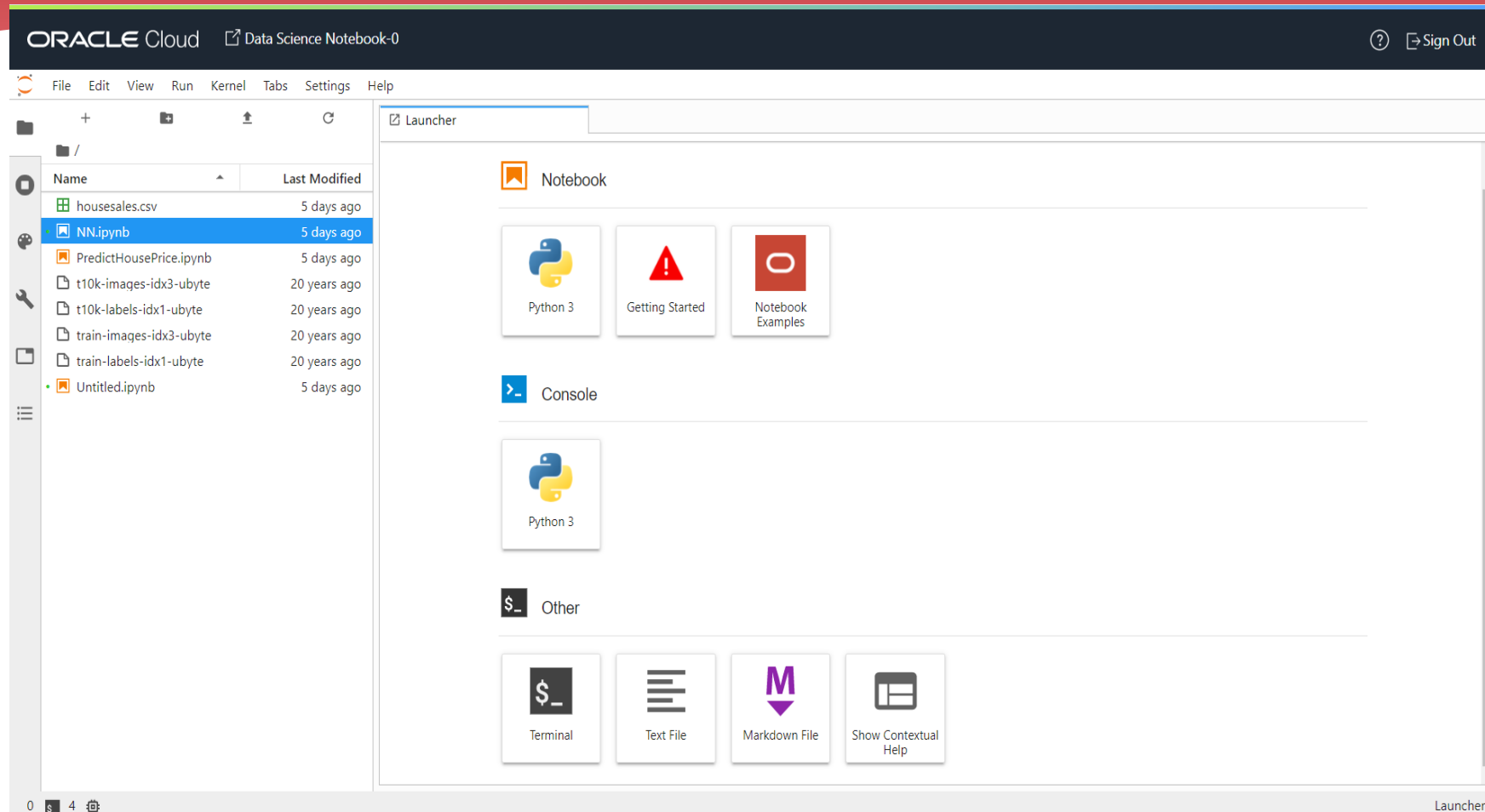
Showing 1 Notebook Session < Page 1 >

Copyright © 2020. Oracle and/or its affiliates. All rights reserved.

MIRACLE
Miracle Finland Oy

Copyright © Miracle Finland Oy

Notebook session



Notebook

The screenshot displays the Oracle Cloud Data Science Notebook interface. The top bar shows 'ORACLE Cloud' and 'Data Science Notebook-0' with a 'Sign Out' button. Below this is a menu bar with 'File', 'Edit', 'View', 'Run', 'Kernel', 'Tabs', 'Settings', and 'Help'. The left sidebar contains a file explorer with a list of files and folders:

Name	Last Modified
housesales.csv	5 days ago
NN.ipynb	5 days ago
PredictHousePrice.ipynb	5 days ago
t10k-images-idx3-ubyte	20 years ago
t10k-labels-idx1-ubyte	20 years ago
train-images-idx3-ubyte	20 years ago
train-labels-idx1-ubyte	20 years ago
Untitled.ipynb	5 days ago

The main area is a code editor for 'PredictHousePrice.ipynb'. It shows a Python 3 kernel and a code cell with the following code:

```
import numpy as np
from scipy.stats import norm
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn import linear_model
from sklearn.metrics import mean_squared_error
from scipy import stats
import warnings
warnings.filterwarnings('ignore')
%matplotlib inline
```

Below the code cell are five output cells showing the execution results:

```
[2]: alldata = pd.read_csv('./housesales.csv')
```

```
[3]: alldata.shape
```

```
[3]: (1460, 81)
```

```
[4]: alldata['SalePrice'][:10]
```

```
[4]: 0    208500
     1    181500
     2    223500
     3    140000
     4    250000
     5    143000
     6    307000
     7    200000
     8    129900
     9    118000
     Name: SalePrice, dtype: int64
```

```
[5]: alldata['SalePrice'].isnull().sum()
```

The bottom status bar indicates 'Mode: Command', 'Ln 1, Col 1', and 'PredictHousePrice.ipynb'.

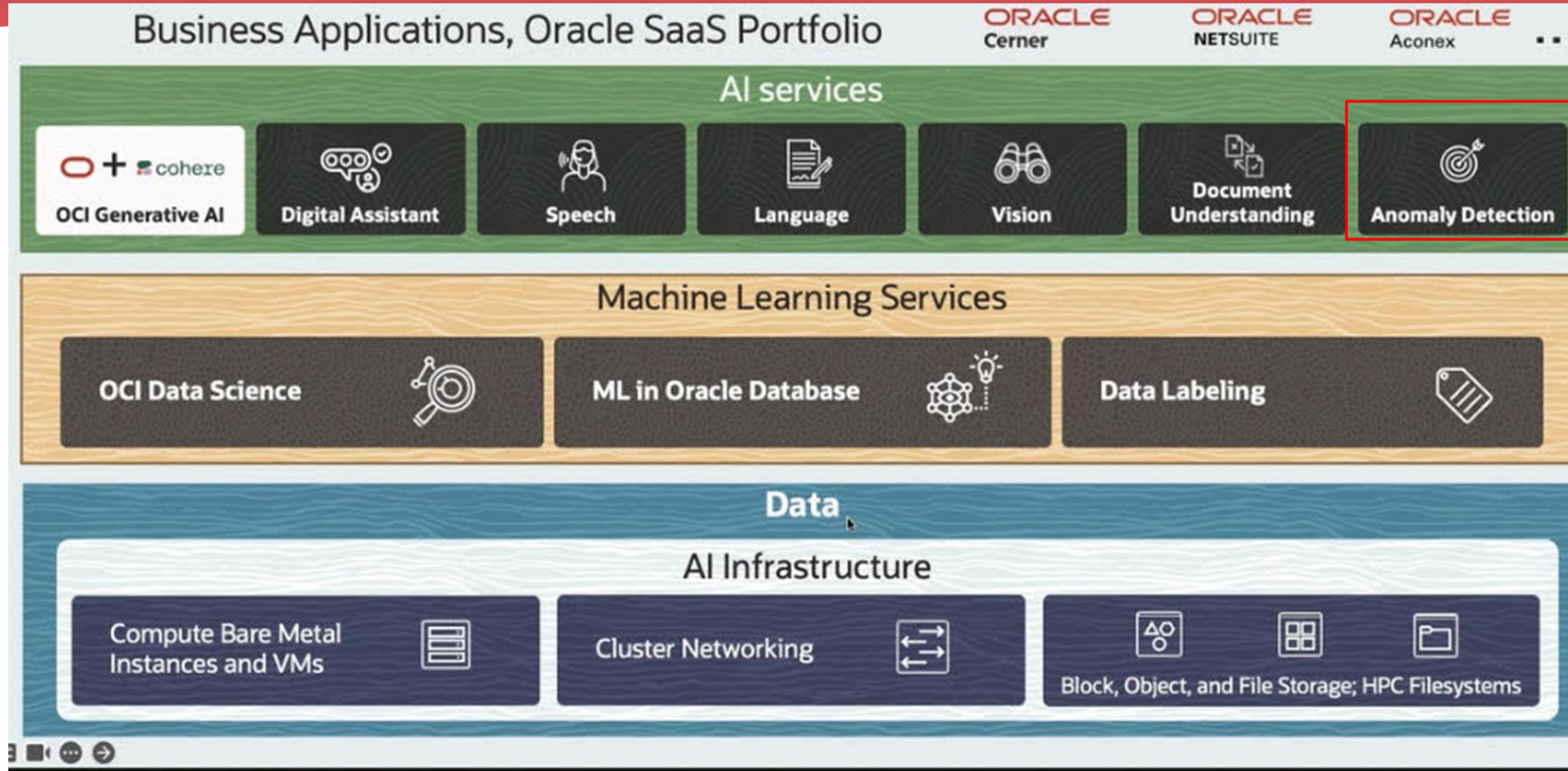
Oracle Accelerated Data Science (ADS) SDK

- * Python library included into the Oracle Cloud Infrastructure Data Science service
- * Can be pip installed also in other environments
- * offers a friendly user interface with objects and methods that describe the steps involved in the lifecycle of machine learning models

ADS, Oracle AutoML engine

- * Exploratory data analysis
- * Automatic data visualization
- * Feature selection
- * Feature engineering
- * Algorithm selection
- * Feature encoding
- * Hyperparameter tuning
- * Model evaluation
- * Model interpretation and explainability...

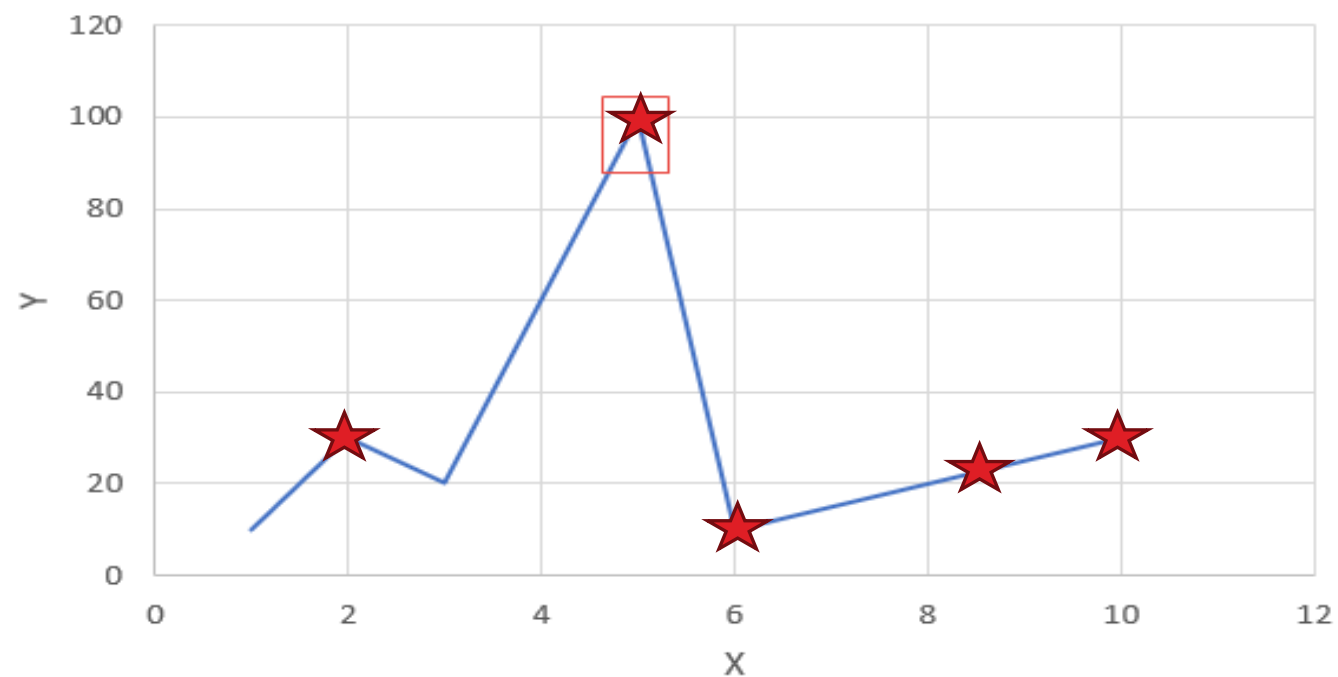
Machine Learning in OCI



Anomaly Detection

- * Identificating rare events/observations, very different from majority of the data

Anomaly Detection



Anomaly Detection, Create and Train Model

Create and Train Model

- 1 [Select Data](#)
- 2 **Train Model**
- 3 [Review](#)

A model is trained until the accuracy options are met, and then it is saved with a unique model OCID.

Select Model Compartment ⓘ

heli

demospace (root)/heli

Name *Optional*

AnomalyDetectionModel

Description *Optional*

Target False Alarm Probability(FAP) ⓘ

0,05

FAP stands for False Alarm Probability, which is basically the likelihood (percentage) of a timestamp is flagged as anomaly in the clean (anomaly-free) training data. It is calculated at every signal level and then averaged across all signals as the final achieved FAP by our model.

Training Fraction Ratio ⓘ

0,7

Training Fraction Ratio specifies the ratio of the whole training data used for our algorithm to learn the pattern and train the model. The rest (1-ratio) of training data will be used for our algorithm to evaluate and report model performance (e.g.. FAP).

[Previous](#)

[Next](#)

[Cancel](#)

Anomaly Detection, Detect Anomaly

Anomalies

Detect the anomalies for the data contained in the request using the stored model

Detect Anomalies

Download JSON

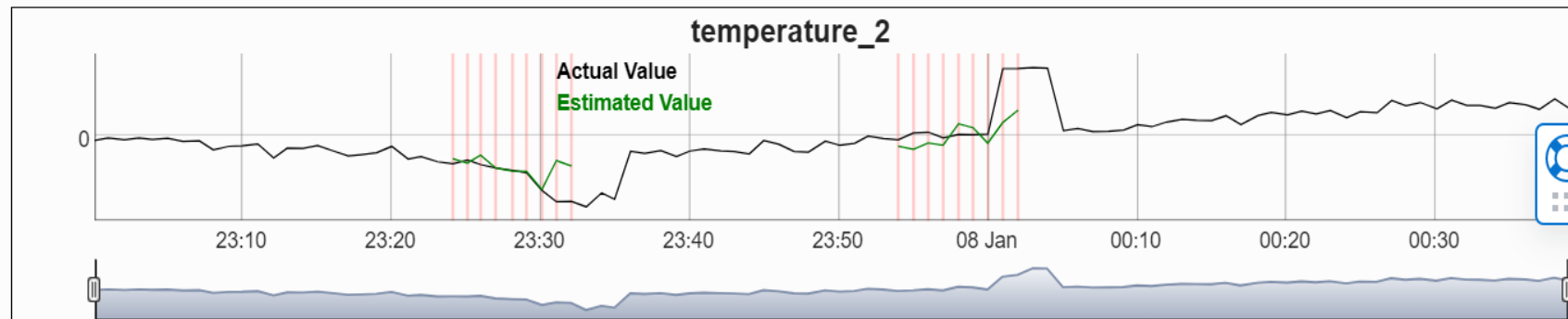
[Reset](#)



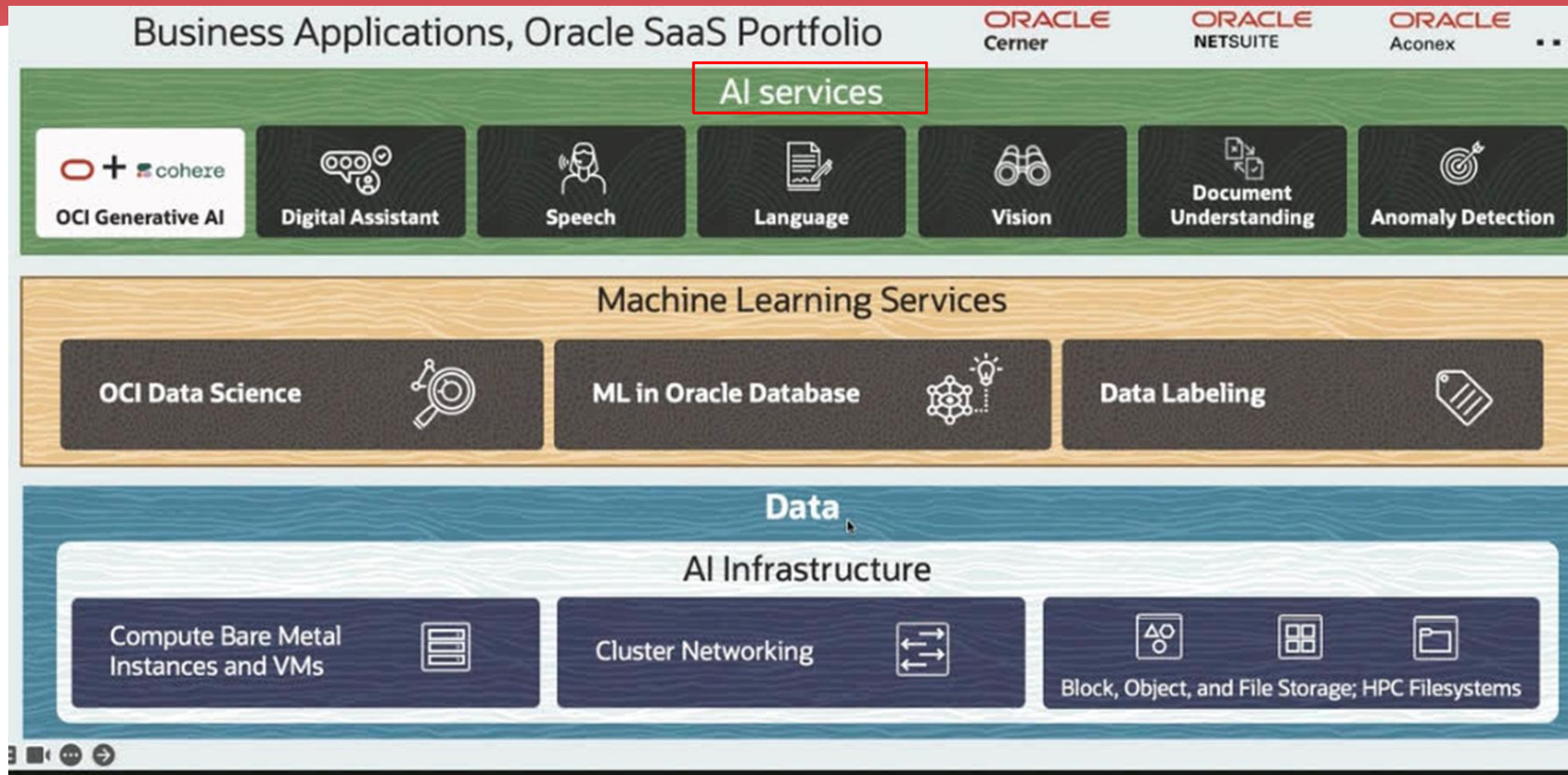
- Black line indicates the actual input value of a signal, Green line indicates the predicted value by the machine learning model, and Red line indicates anomaly being detected at that timestamp.
- The Anomaly Score Per Signal shows the significance of anomaly at individual signal level for a given timestamp. Note that not all the signals will flag anomalies at the same time.
- The Aggregated Anomaly Score indicates the significance of anomaly for a given timestamp by considering the anomaly from all signals together.


Select column labels(with anomalies) for visualization

Select column labels(with anomalies) for visualization



Machine Learning in OCI



 Search

- Home
- Compute
- Storage
- Networking
- Oracle Database
- Databases
- Analytics & AI**
- Developer Services
- Identity & Security
- Observability & Management
- Hybrid
- Migration & Disaster Recovery
- Billing & Cost Management
- Governance & Administration

Analytics & AI

Analytics

Analytics Cloud
Fusion Analytics Warehouse

Data Lake

Big Data Service
Data Catalog
Data Integration
Data Flow

Messaging

Streaming
Service Connector Hub

Machine Learning

Data Science
Data Labeling

AI Services

Language
Speech
Vision
Document Understanding

Anomaly Detection
Digital Assistant

Media Services

Media Flow
Media Flow Jobs
Media Streams

Related services

Autonomous Database
MySQL
Oracle NoSQL Database
Object Storage & Archive Storage
Application Performance Monitoring
Logging Analytics
Operations Insights
Logging

Help

Analytics Cloud Documentation
Big Data Documentation
Data Science Documentation
Digital Assistant Documentation
REST APIs

AI Services

Generative AI

Oracle Analytics Cloud, OAC

Oracle Analytics Cloud (OAC) is...

- * Scalable and secure public cloud service in OCI
- * For example:
 - * Data connectivity
 - * Data preparation, Data flow
 - * Data visualization
 - * Data collaboration

Q Search

[Home](#)[Compute](#)[Storage](#)[Networking](#)[Oracle Database](#)[Databases](#)[Analytics & AI](#)[Developer Services](#)[Identity & Security](#)[Observability & Management](#)[Hybrid](#)[Migration & Disaster Recovery](#)[Billing & Cost Management](#)[Governance & Administration](#)

Analytics & AI

Analytics

[Analytics Cloud](#)[Fusion Analytics Warehouse](#)

Data Lake

[Big Data Service](#)[Data Catalog](#)[Data Integration](#)[Data Flow](#)

Messaging

[Streaming](#)[Service Connector Hub](#)

Machine Learning

[Data Science](#)[Data Labeling](#)

AI Services

[Language](#)[Speech](#)[Vision](#)[Document Understanding](#)[Anomaly Detection](#)[Digital Assistant](#)

Media Services

[Media Flow](#)[Media Flow Jobs](#)[Media Streams](#)

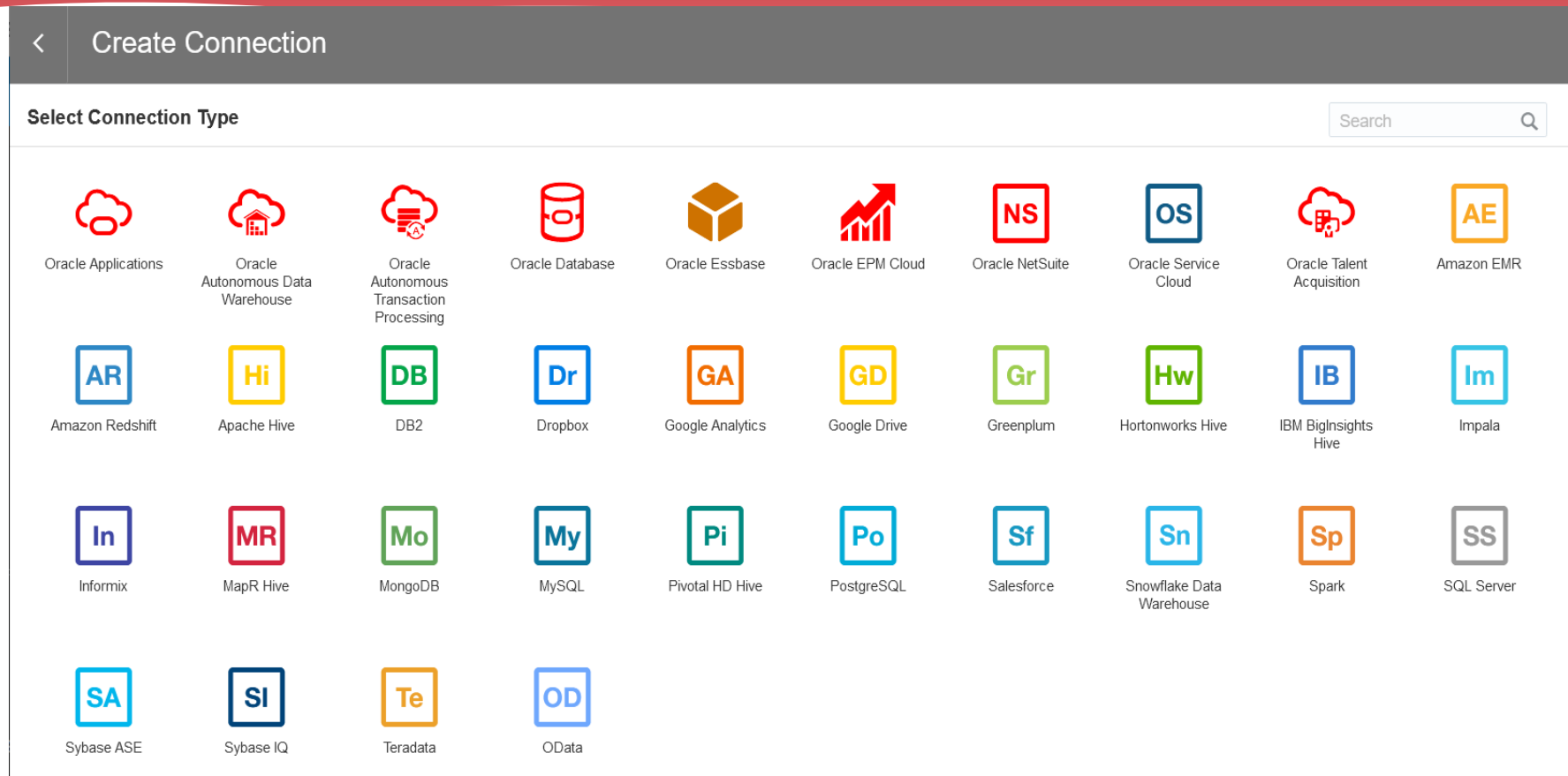
Related services

[Autonomous Database](#)[MySQL](#)[Oracle NoSQL Database](#)[Object Storage & Archive Storage](#)[Application Performance Monitoring](#)[Logging Analytics](#)[Operations Insights](#)[Logging](#)

Help

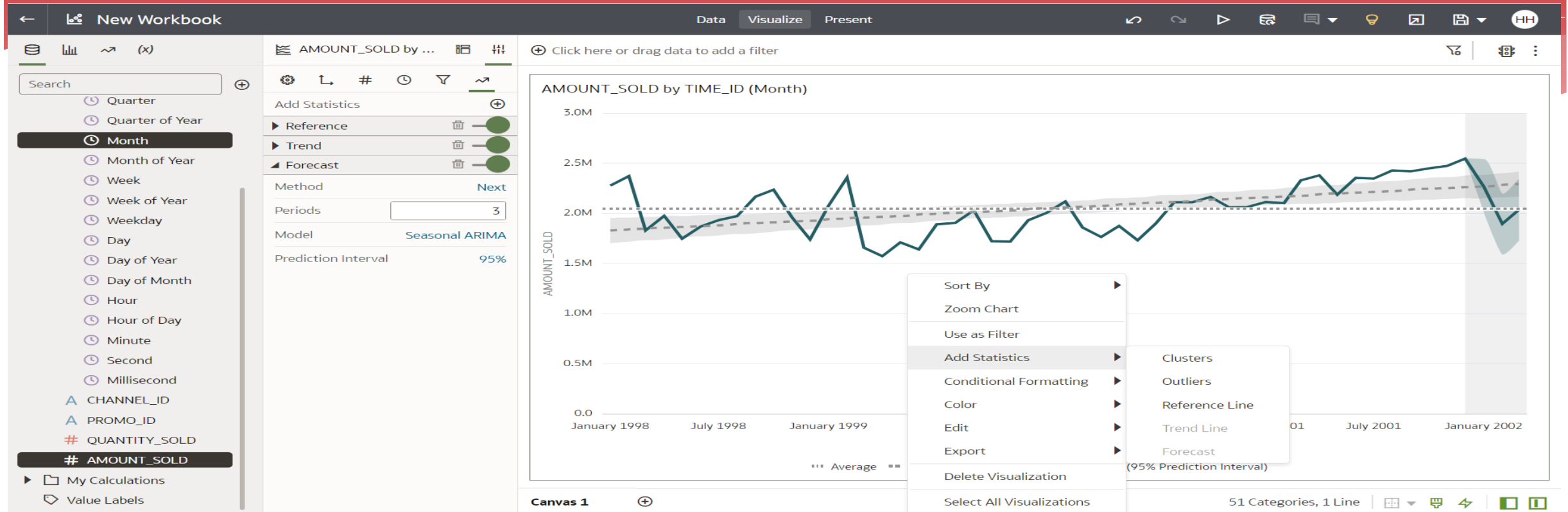
[Analytics Cloud Documentation](#)[Big Data Documentation](#)[Data Science Documentation](#)[Digital Assistant Documentation](#)[REST APIs](#)

Plenty of data source candidates

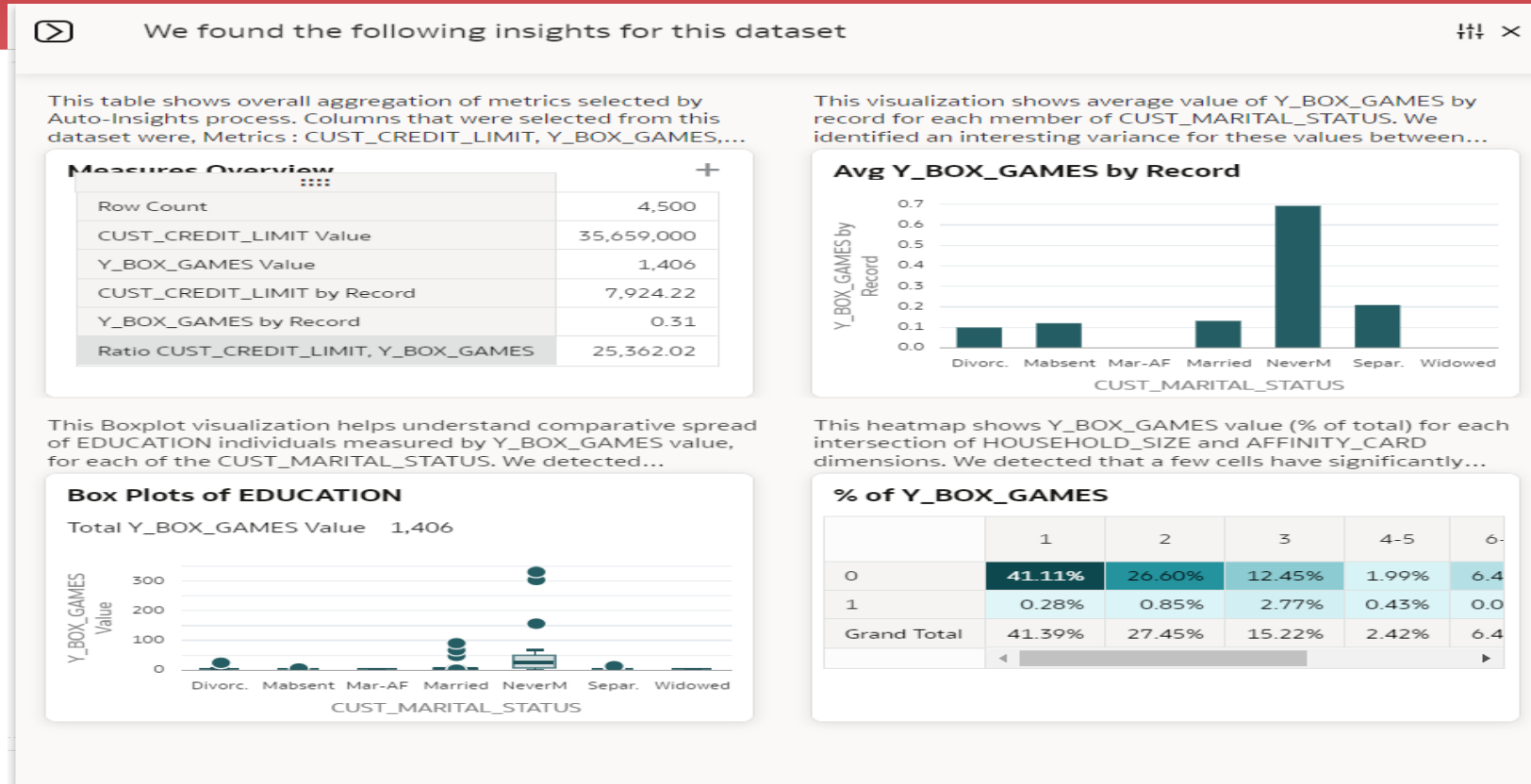


Machine Learning?

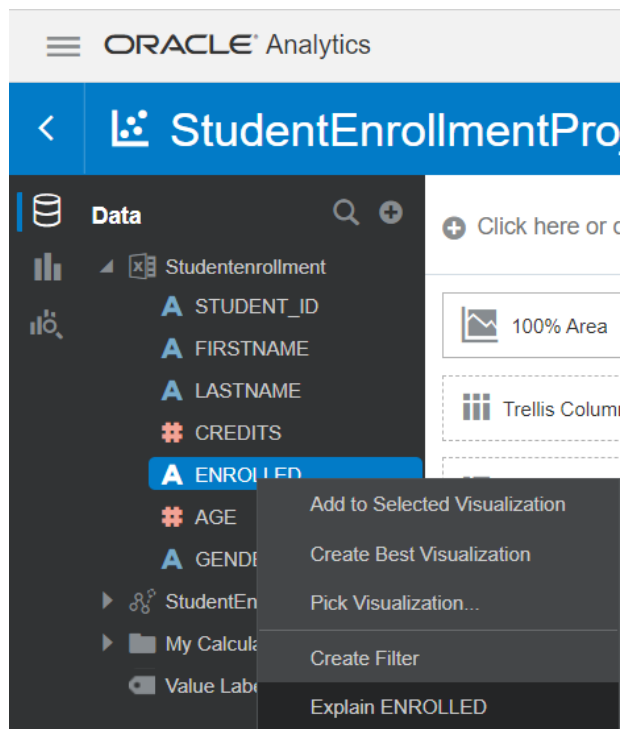
ML build in OAC



Insights



Explain



Explain

Explain ENROLLED

Add Selected X

Basic Facts about ENROLLED

What are the values of ENROLLED and how do they relate to each other?

Key Drivers of ENROLLED

What elements in this data best explain the values of ENROLLED?

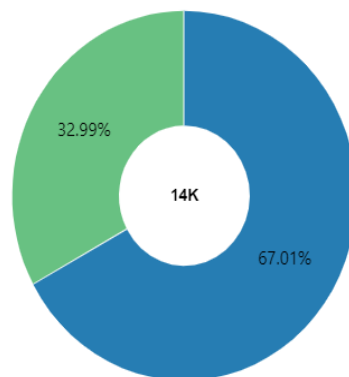
Segments that Explain ENROLLED

What hidden groups in the data can predict outcomes for ENROLLED?

Anomalies of ENROLLED

What groups in the data exhibit unexpected results for ENROLLED?

Basic facts about ENROLLED



ENROLLED ■ 0 ■ 1



Select for Canvas

ENROLLED is a Numeric Attribute with 2 unique values across 13848 rows (100%). The most common ENROLLED is 0 (67%) and the least common is 1 (33%).

Creating Models in OAC, Classification

AutoML

CreatingAnAutoMLModel

Search

Add Data

Join

Union Rows

Filter

Aggregate

Save Dataset

Create Essbase Cube

Add Columns

Select Columns

Rename Columns

Transform Column

Merge Columns

Split Columns

Bin

Group

Branch

Cumulative Value

Time Series Forecast

Analyze Sentiment

Train Numeric Prediction

Train Multi-Classifier

Train Clustering

Train Binary Classifier

Customer...

AutoML

Save Model

Save Model

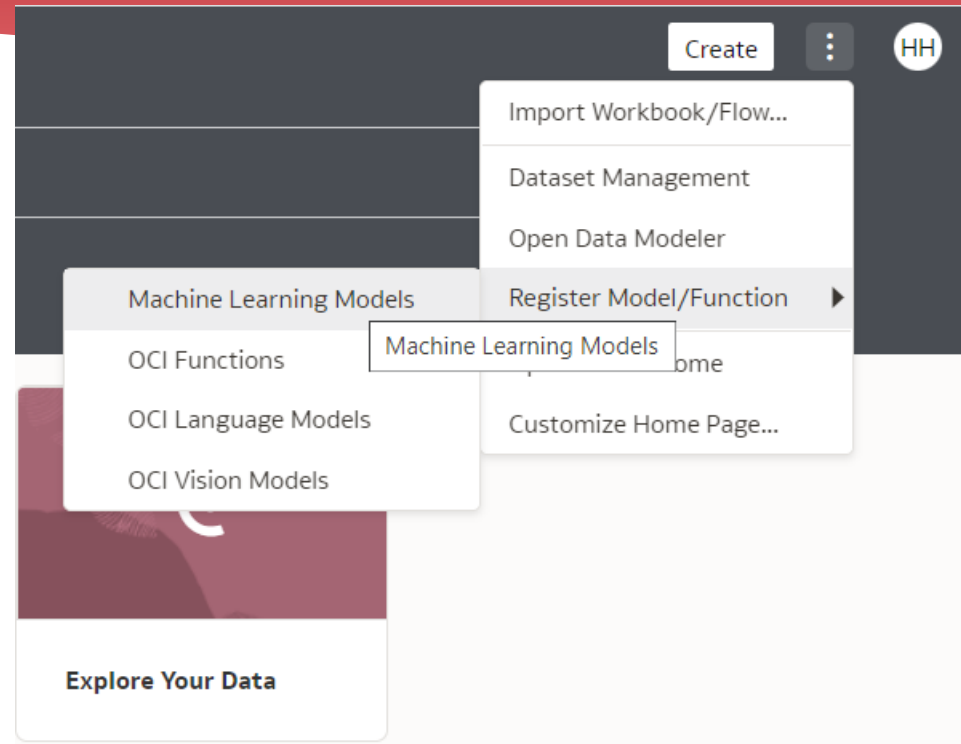
Model name

AutoMLModel

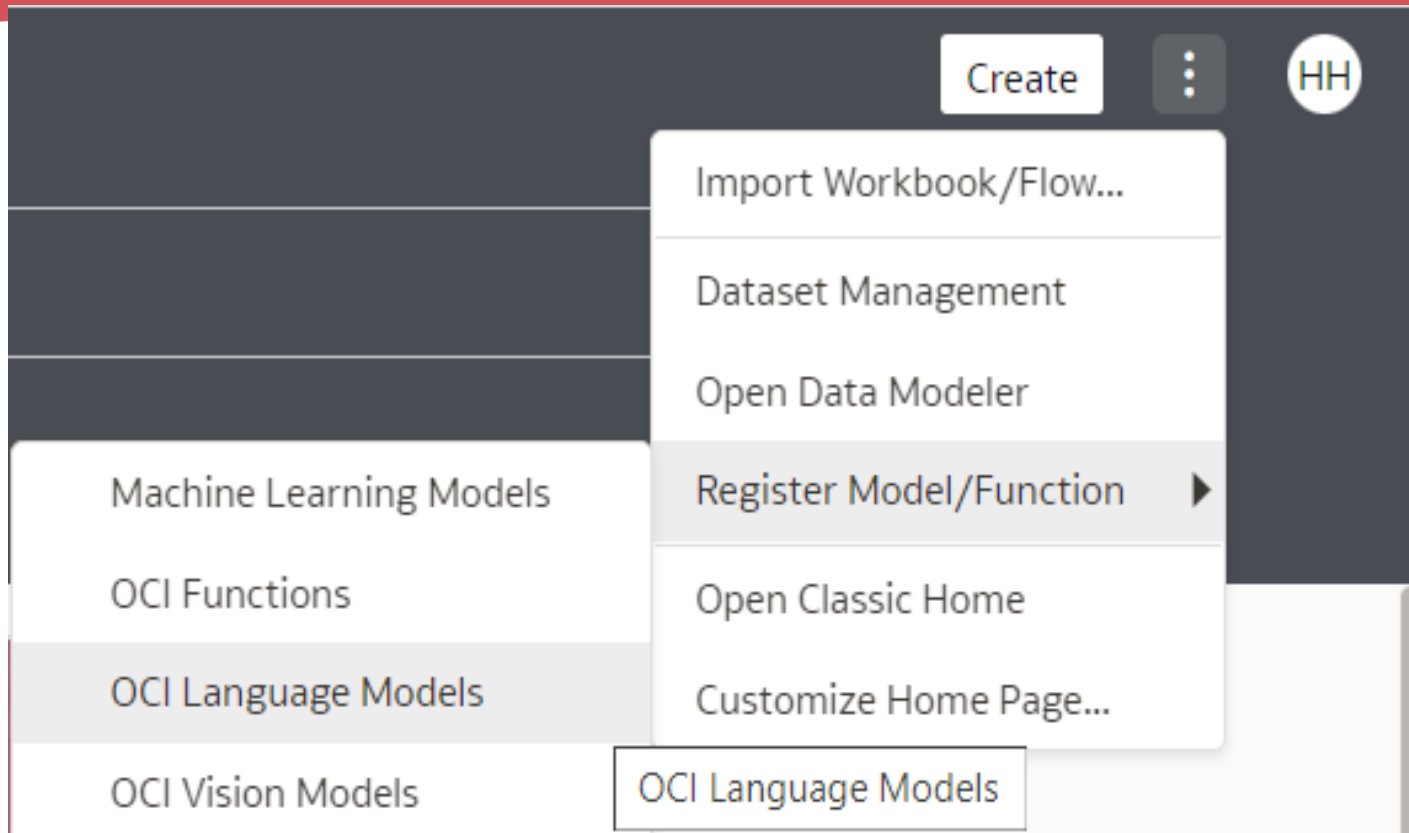
Model description

99 CUST_ID	ab CUST_GENDER	ab CUST_MARIT...	99 CUST_YEAR_...	ab CUST_INCOME_...	99 CUST_CREDIT...	ab EDUCATION	ab AFFINITY_CARD	ab HOUSEHOLD...	ab OCCUPATION	99 YRS_RESID
100100	F	Married	1959	J: 190,000 - 249,999	10000	Masters	1	4-5	Prof.	4
100200	M	NeverM	1983	L: 300,000 and above	9000	< Bach.	0	1	Other	2
100300	M	Married	1961	G: 130,000 - 149,999	10000	Bach.	1	3	Prof.	4

Use OML models from the database



AI Services



Using a model in OAC

Conclusions

- * ML is here now and it is the future
- * Oracle offers several tools for ML
- * Database is the best place for machine learning
 - * the data is there
 - * It's been designed for data processing
- * It is all about good quality data

Conclusions

- * SQL, PL/SQL, R, Python, ...
- * AutoML
- * Data Science Service
- * AI Services
- * Analytics Cloud
- * No excuses! Start learning, now!

THANK YOU!

QUESTIONS?

Email: heli@miracleoy.fi

Twitter: [@HeliFromFinland](https://twitter.com/HeliFromFinland)

Blog: Helifromfinland.com