Oracle Tools for Machine Learning

Heli Helskyaho, Matias Helskyaho
@helifromfinland

MIRACLE
Miracle Finland Oy

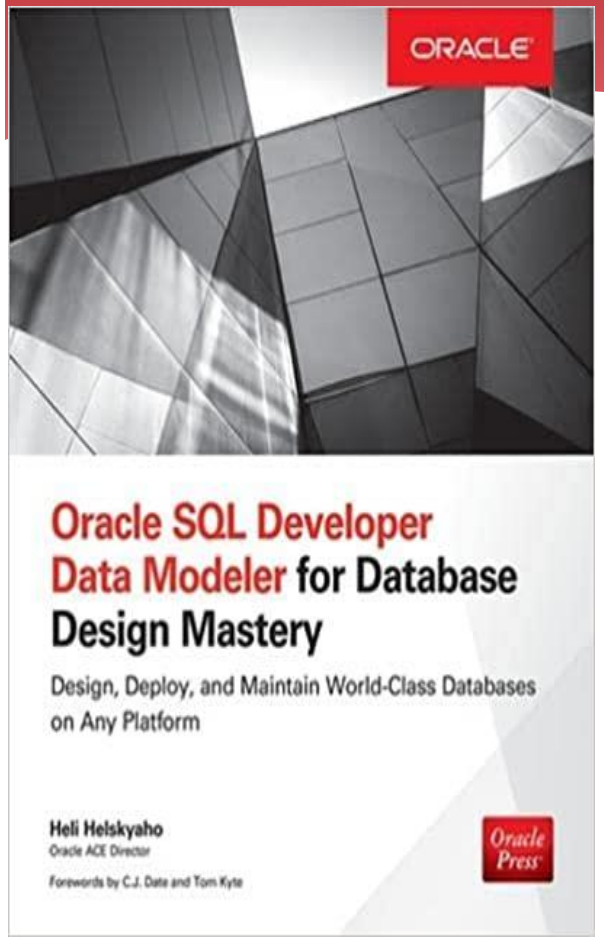
Heli

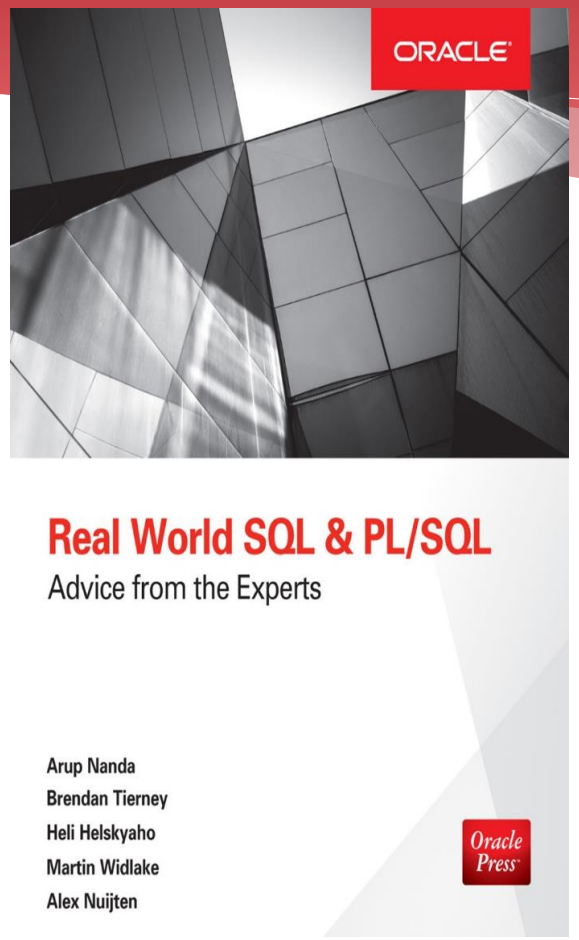
- Graduated from University of Helsinki (Master of Science, computer science), currently a doctoral student at University of Helsinki
- Worked with Oracle products since 1993, worked for IT since 1990
- Data and Database!
- CEO for Miracle Finland Oy
- Oracle ACE Director
- Public speaker and an author
- 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)

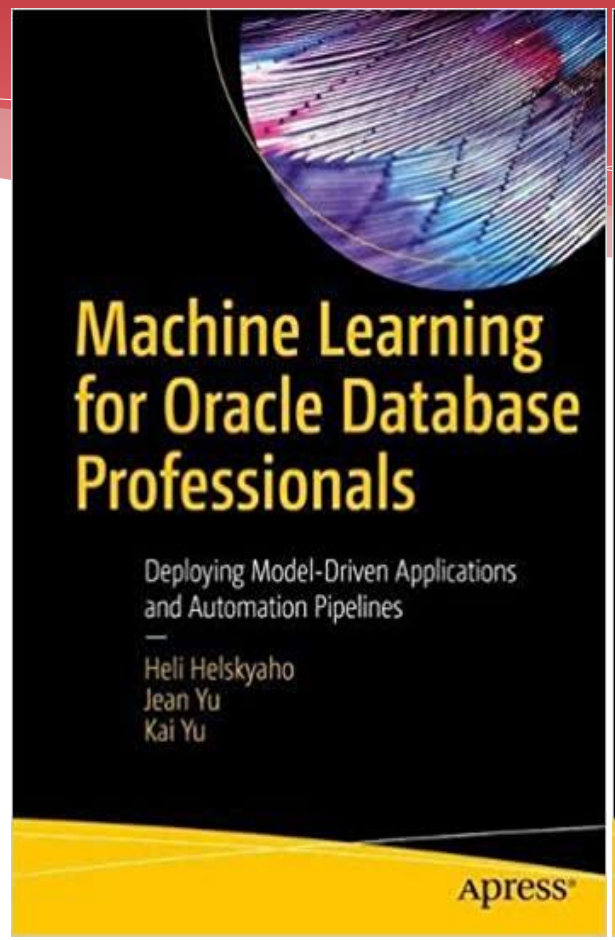
Oracle ACE Director

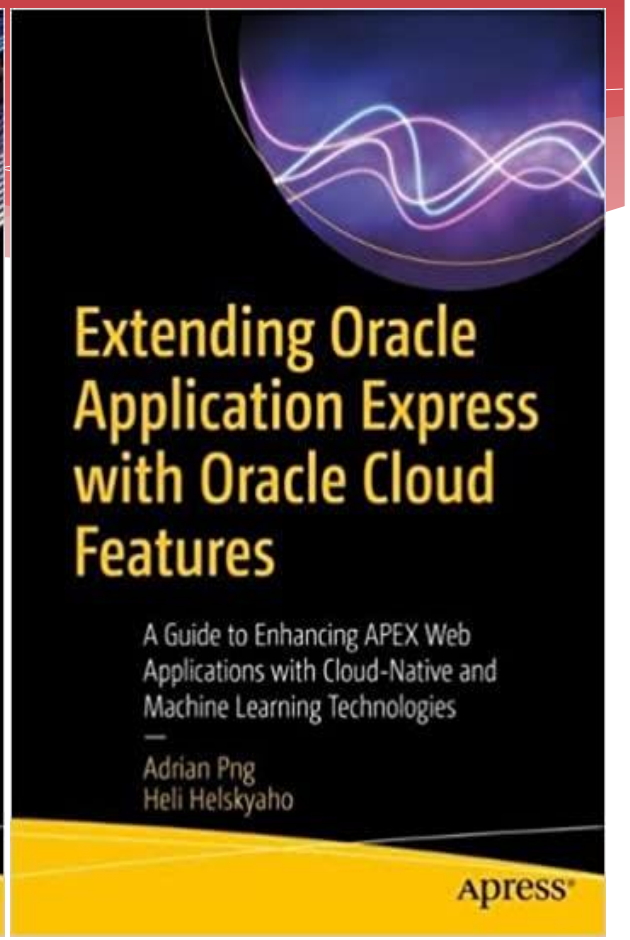
MIRACLE
Miracle Finland Oy

Books









MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

Connectors to the database using different programming languages

Miracle Finland Oy

Oracle DB and Python

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

MIRACLE
Miracle Finland Oy

In-database machine learning

* OML4SQL
* OML4R
* OML4Py

Miracle Finland Oy

A closer look to OML4SQL

Miracle Finland Oy

Data Dictionary Views for ODM

Table 2-1 Data Dictionary Views for Oracle Data Mining

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

MIRACLE
Miracle Finland Oy

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*

MIRACLE
Miracle Finland Oy

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.

MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

EXPLAIN

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

MIRACLE
Miracle Finland Oy

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)	0,136995901803663904648862043508756735030 2	1
2	PALATE	(null)	0,103303810686650561246670008631788592961 2	2
3	AROMA	(null)	0,087915268294064515065235362564757742 82	3
4	APPEARANCE	(null)	0,059799441574642509408427896866963796385 1	4
5	IDINDEX	(null)	0,051474317401555431124656659166858095397 8	5
6	STYLE	(null)	0,048718285081069389446959613176399702081 8	6
7	BEERID	(null)	0,044765173049272308951109514189414219900 2	7
8	ABV	(null)	0,028274464814559205025103905916676268992 4	8
9	BREWERID	(null)	0,027321782394826963024988796480694305807 8	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

MIRACLE
Miracle Finland Oy

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)

MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

PROFILE

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

MIRACLE
Miracle Finland Oy

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]

MIRACLE
Miracle Finland Oy

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.*

MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

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));
```

MIRACLE
Miracle Finland Oy

A Transformation List

* is defined as a table/*collection of transformation records* (transform_rec)
* A list can be created using:
  * The SET_TRANFORM 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

MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

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);
```

MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

Choose the Mining Function

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

MIRACLE
Miracle Finland Oy

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…

MIRACLE
Miracle Finland Oy

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…

MIRACLE
Miracle Finland Oy

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;
/
```

MIRACLE
Miracle Finland Oy

Other possible settings

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

MIRACLE
Miracle Finland Oy

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;
/
```

MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

Testing and Evaluating the model

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

MIRACLE
Miracle Finland Oy

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;
/
```

MIRACLE
Miracle Finland Oy

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*

MIRACLE
Miracle Finland Oy

BEER_RESULT_TABLE_DT

Indexes | Model | Constraints | Grants | Statistics | UI Defaults | Triggers | Dependencies | SQL | REST

t Rows | Insert Row

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

MIRACLE
Miracle Finland Oy

Evaluating

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

MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

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;
```

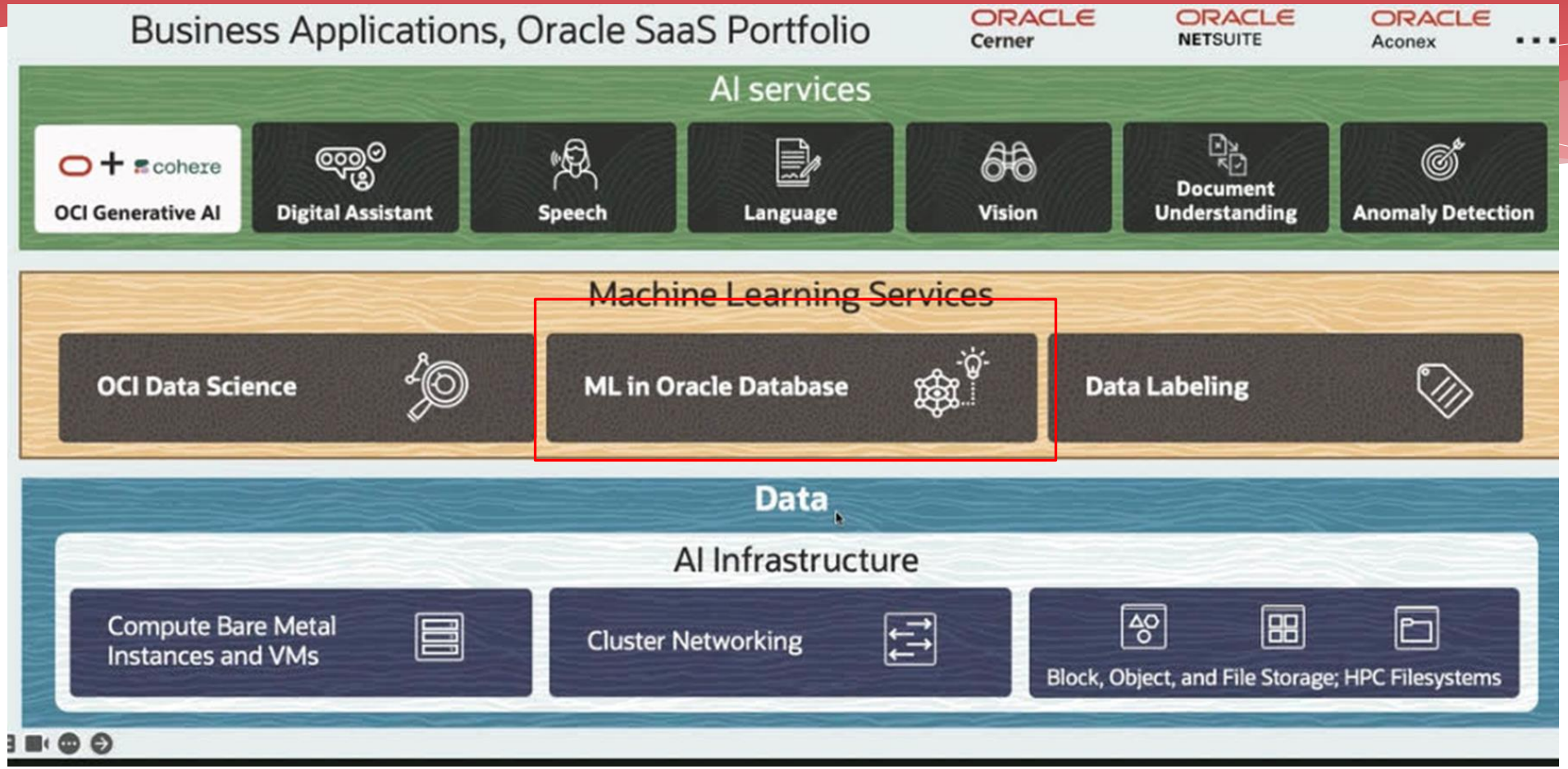
1.50874635568513l3E-001

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

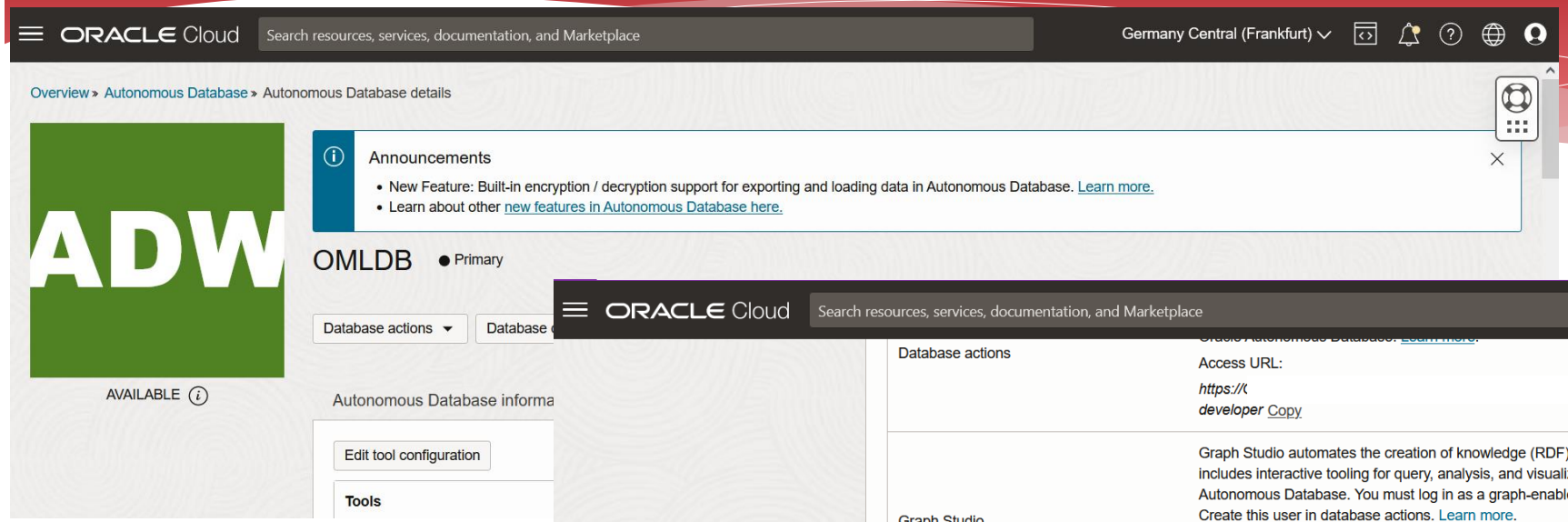
2.41400181579743777E-001

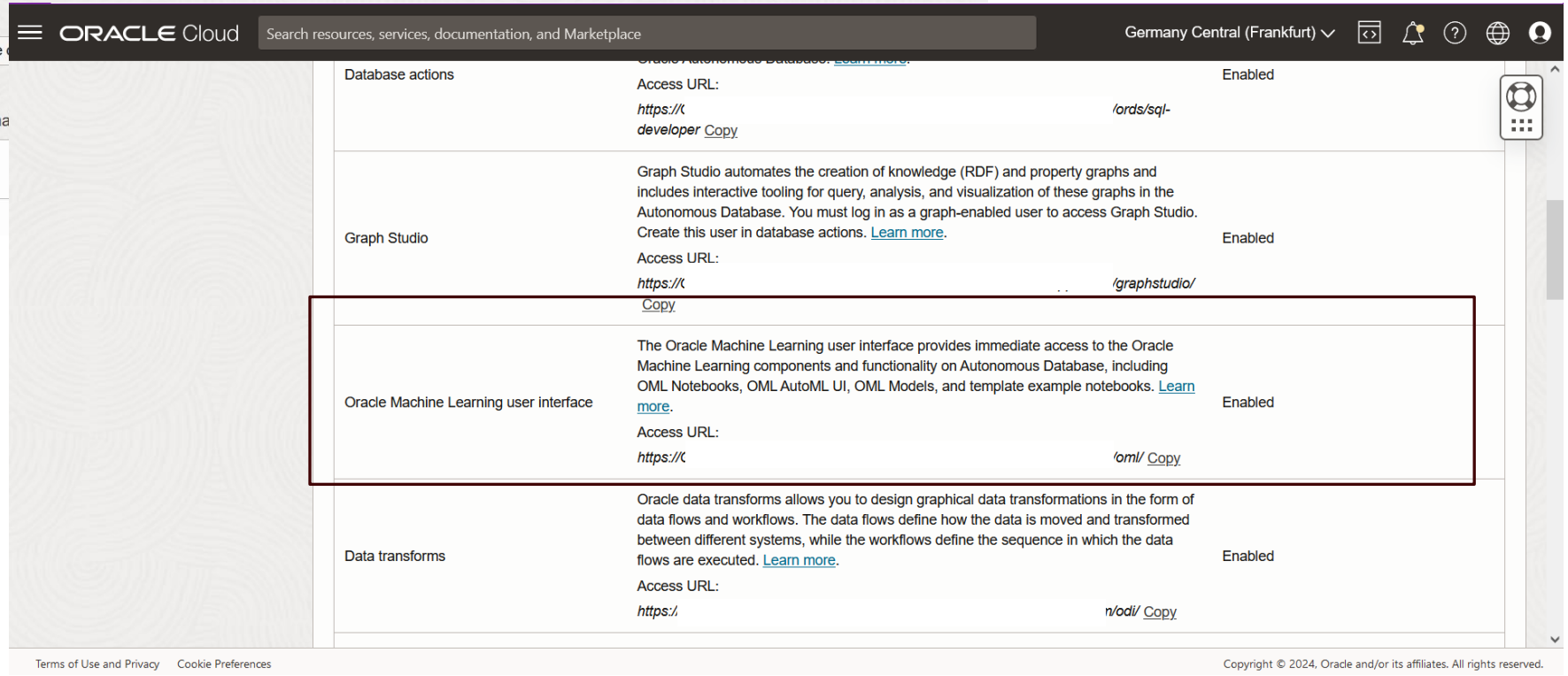
MIRACLE
Miracle Finland Oy

Machine Learning in OCI

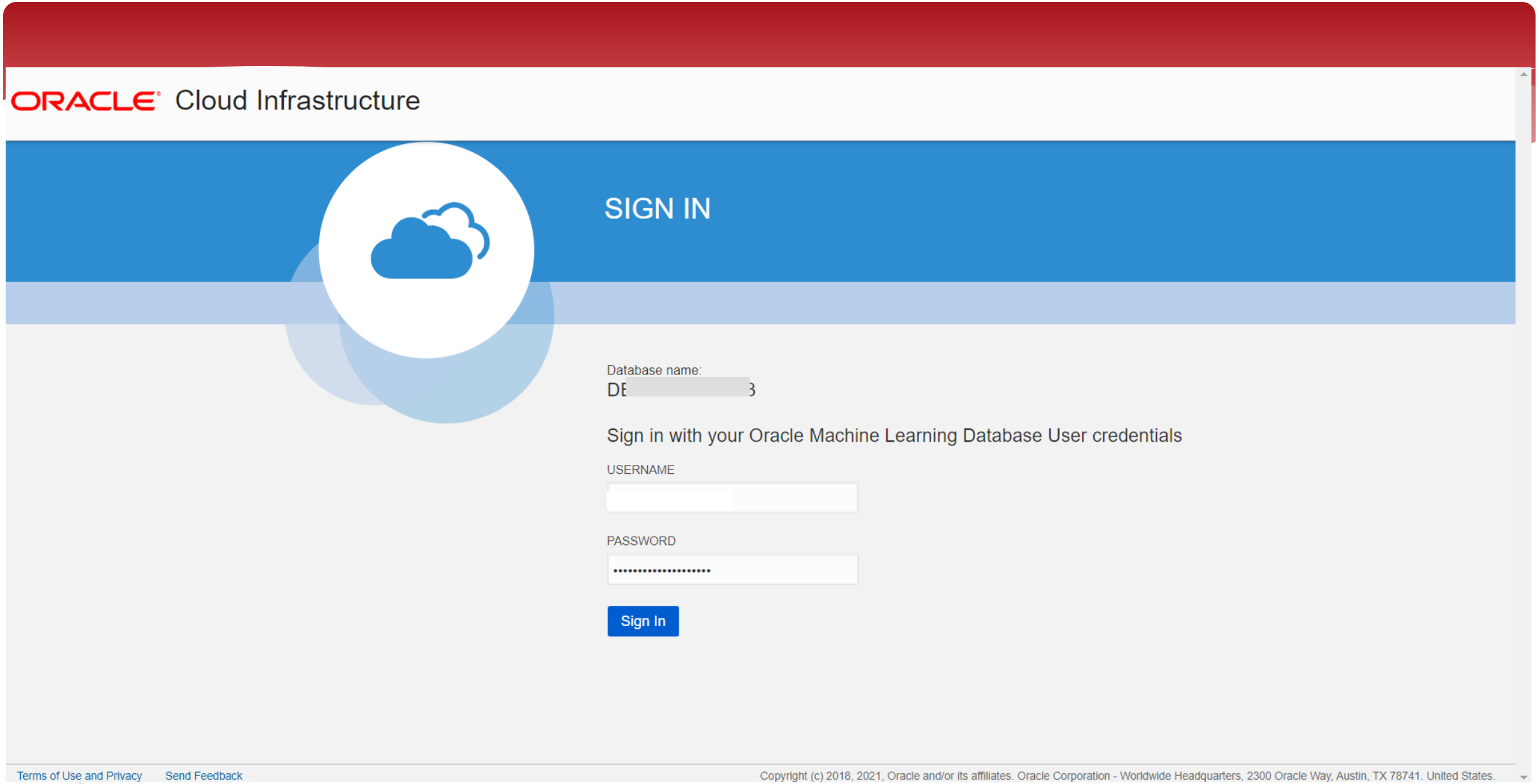


MIRACLE
Miracle Finland Oy

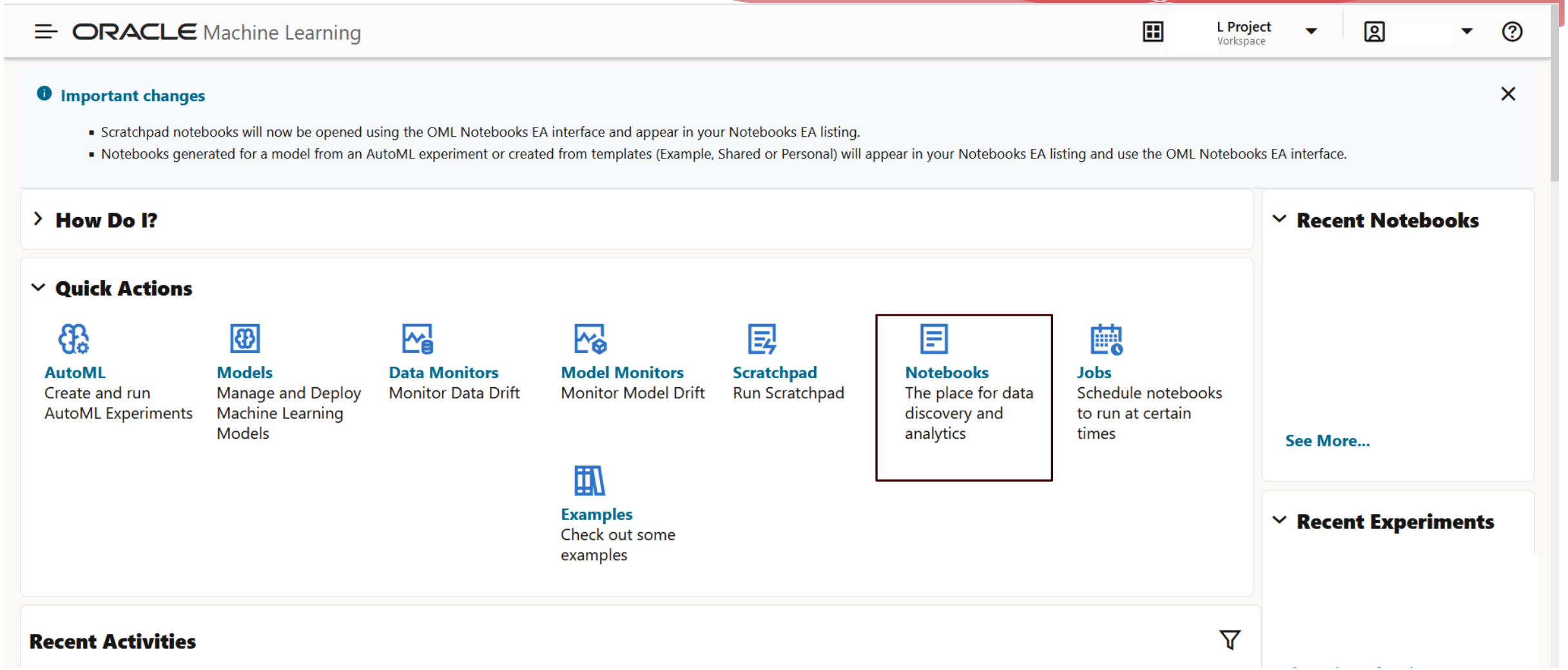




MIRACLE
Miracle Finland Oy



MIRACLE
Miracle Finland Oy



MIRACLE
Miracle Finland Oy

ORACLE Machine Learning

HELIML Project [HELIML Workspa...

HELIML

Connected

Classification Prediction Model

default

FINISHED

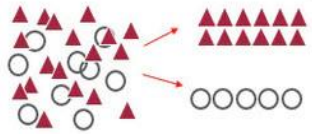
Predicting Target Customers using Classification

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 loyaly or affinity card hyper-respond i.e. they increae 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 Ujjni, Mark Hornick, Marcos Arancibia

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.

FINISHED

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;
```

FINISHED

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

MIRACLE
Miracle Finland Oy

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)

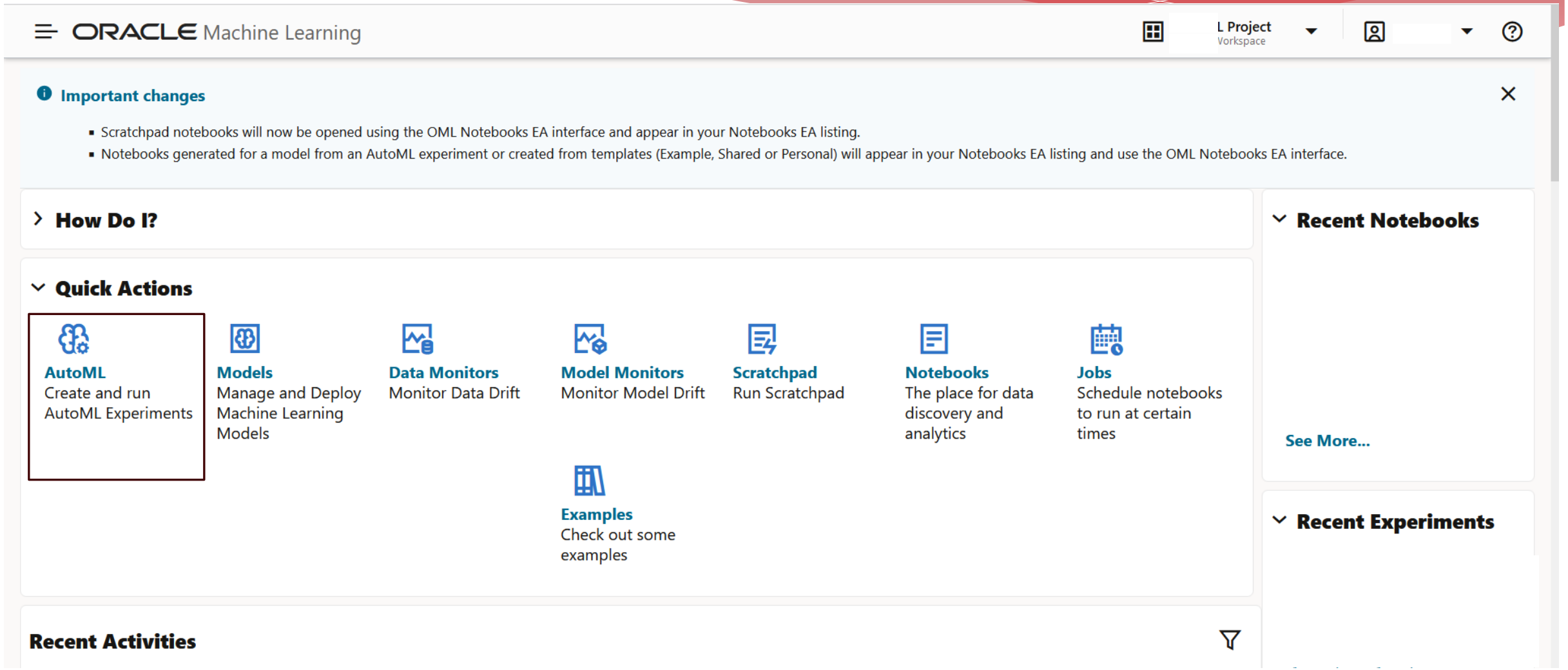
MIRACLE
Miracle Finland Oy

Machine Learning words

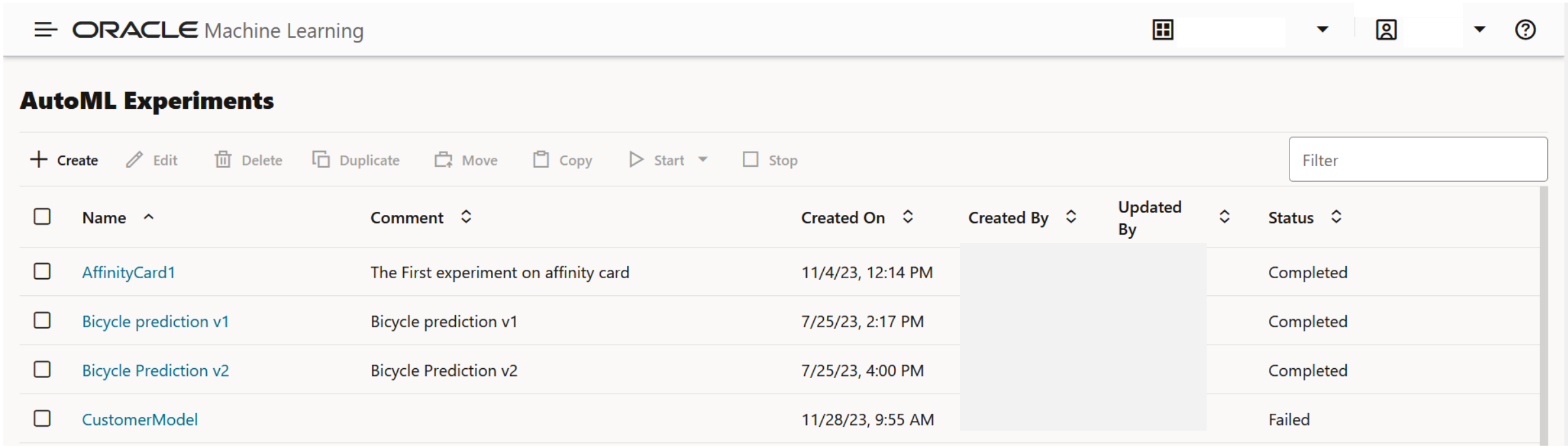
* Exploratory data analysis
* Data visualization
* Feature selection
* Feature engineering
* Algorithm selection
* Feature encoding
* Hyperparameter tuning
* Model evaluation
* Model interpretation and explainablility…

MIRACLE
Miracle Finland Oy

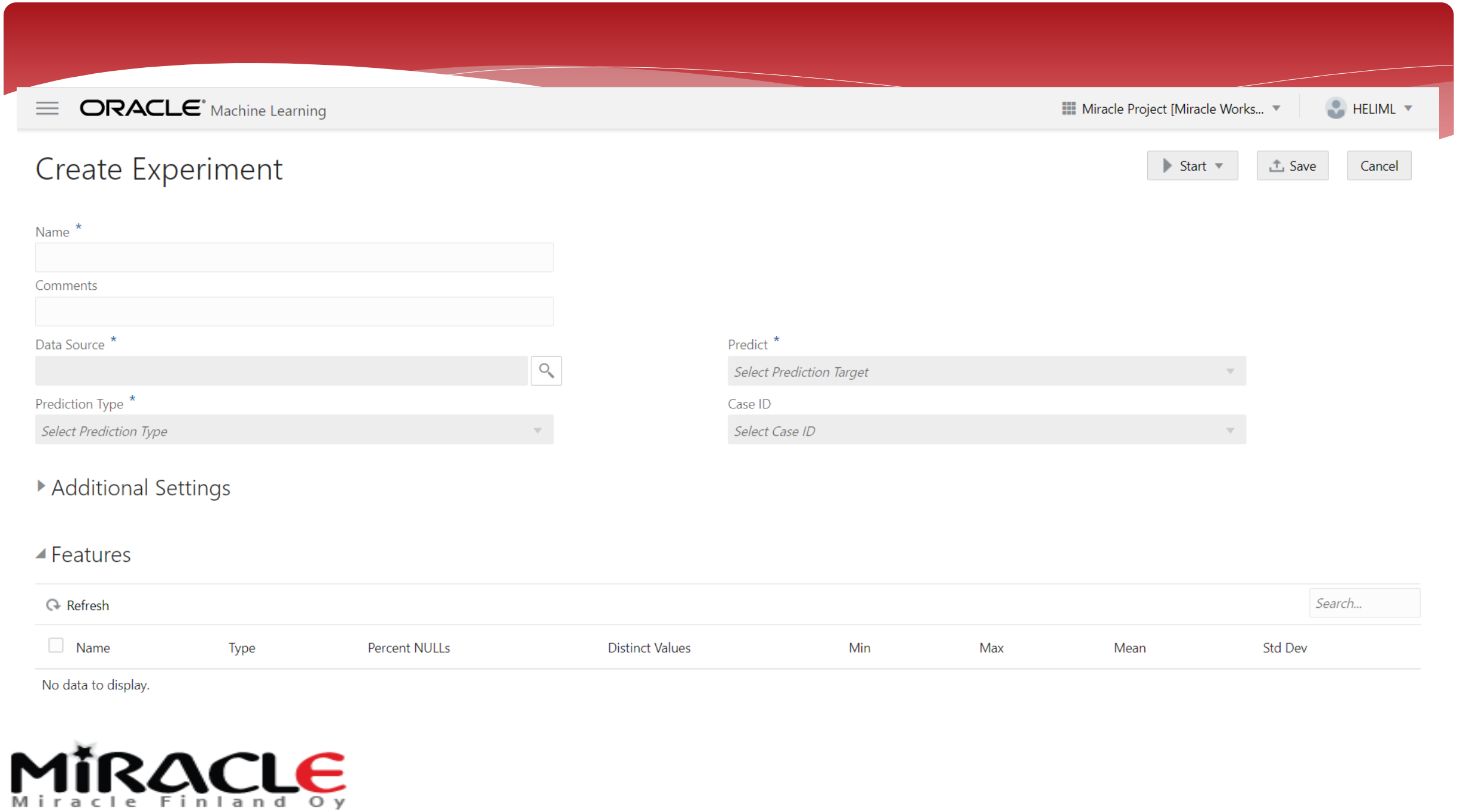
AutoML



MIRACLE
Miracle Finland Oy

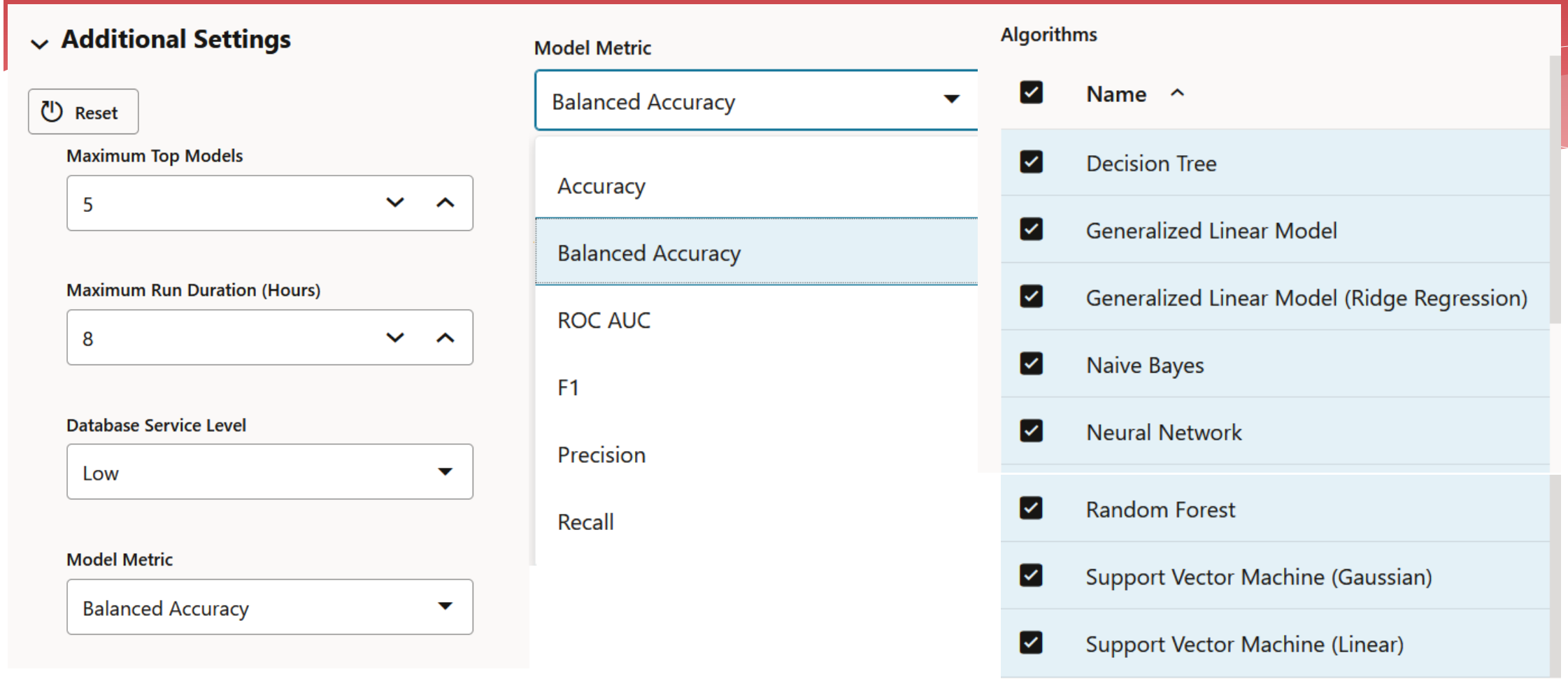


MIRACLE
Miracle Finland Oy



MIRACLE
Miracle Finland Oy

(For Classification)



MIRACLE
Miracle Finland Oy

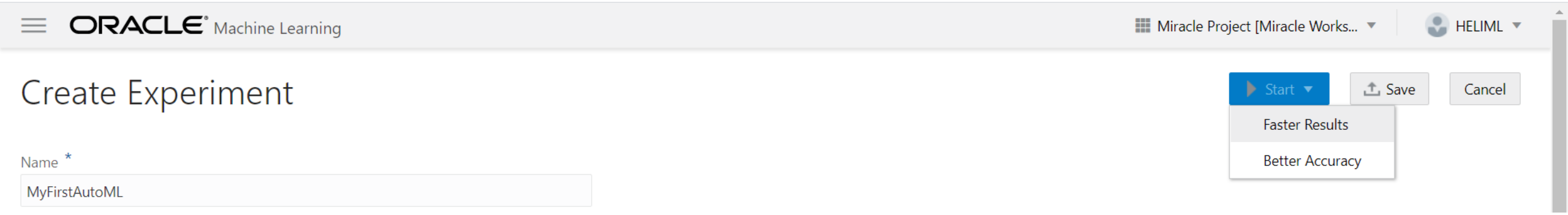
Features

Refresh

Search...

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

MIRACLE
Miracle Finland Oy



MIRACLE
Miracle Finland Oy

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...	glmr_71fd0d3403	0.9792
Support Vector Machine (Linear)	svml_801e8b40f9	0.9792
Generalized Linear Model	glm_86fbdeec63	0.9715

MIRACLE
Miracle Finland Oy

Balanced Accuracy

Leader Board

Deploy Rename Create Notebook Metrics

Algorithm	Model Name	Recall	Accuracy
Support Vector Machine (Gaussian)	svmg_fcd3cb7701	0.9792	0.9912
Neural Network	nn_6d09b27a4a	0.9792	0.9912
Generalized Linear Model (Ridge Regress...	glmr_71fd0d3403	0.9583	0.9823
Support Vector Machine (Linear)	svml_801e8b40f9	0.9583	0.9823
Generalized Linear Model	glm_86fbdeec63	0.9583	0.9735

Select Additional Metrics

- [x] Accuracy
- [ ] F1
- [x] Precision
- [x] Recall
- [ ] ROC AUC
- [ ] F1 Micro
- [ ] F1 Macro

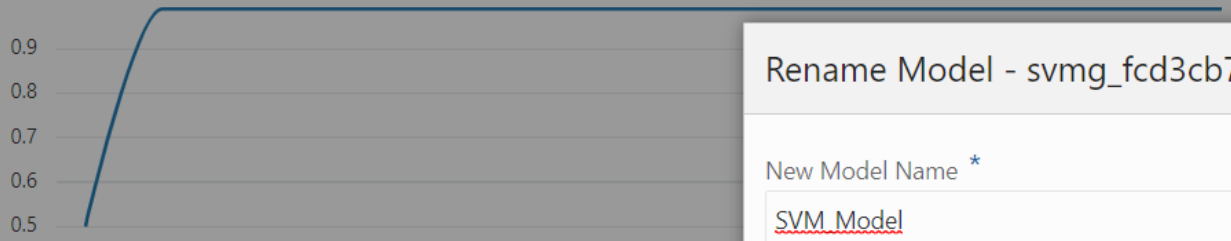
MIRACLE
Miracle Finland Oy

MyFirstAutoML

Completed | Start

Experiment Settings | Edit

Balanced Accuracy



Rename Model - svmg_fcd3cb7701

New Model Name *

SVM_Model

Enter an alphanumeric Name (max 123 characters)

OK | Cancel

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

MIRACLE
Miracle Finland Oy

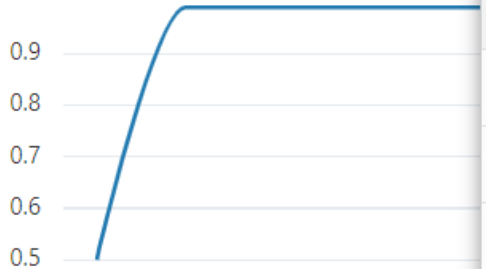
The selected model has been rename

<- Experiments

MyFirstAutoML

Experiment Settings Edit

Balanced Accuracy



Leader Board

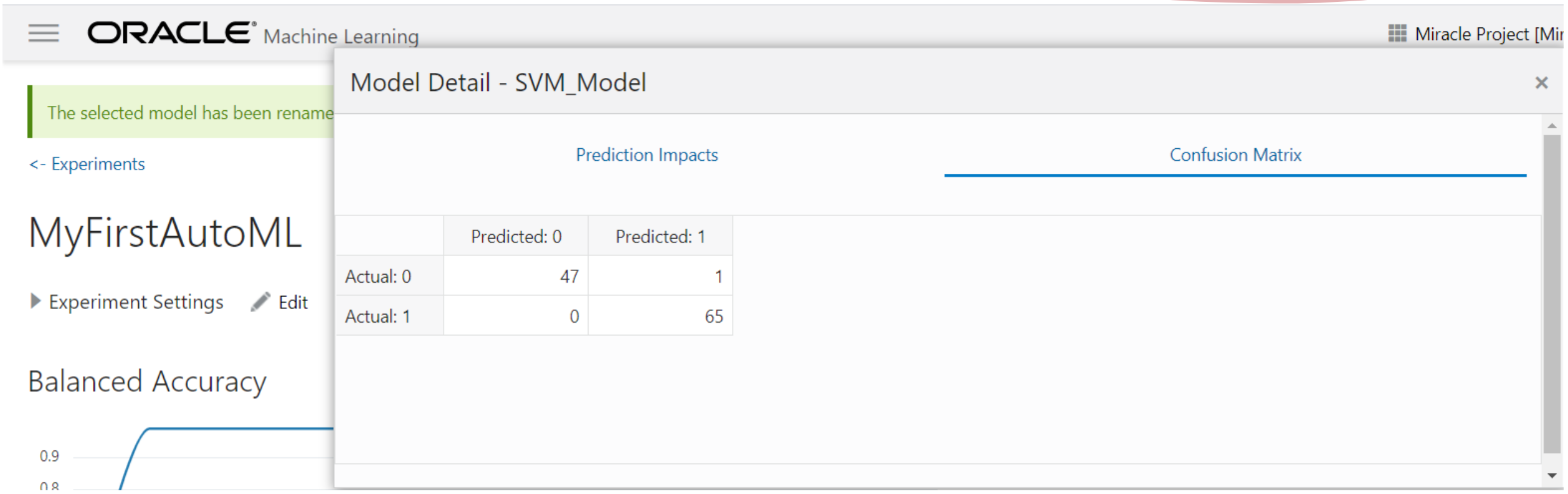
Deploy Rename Create Noteb

Algorithm
Support Vector Machine (Gaussian)
Neural Network

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	



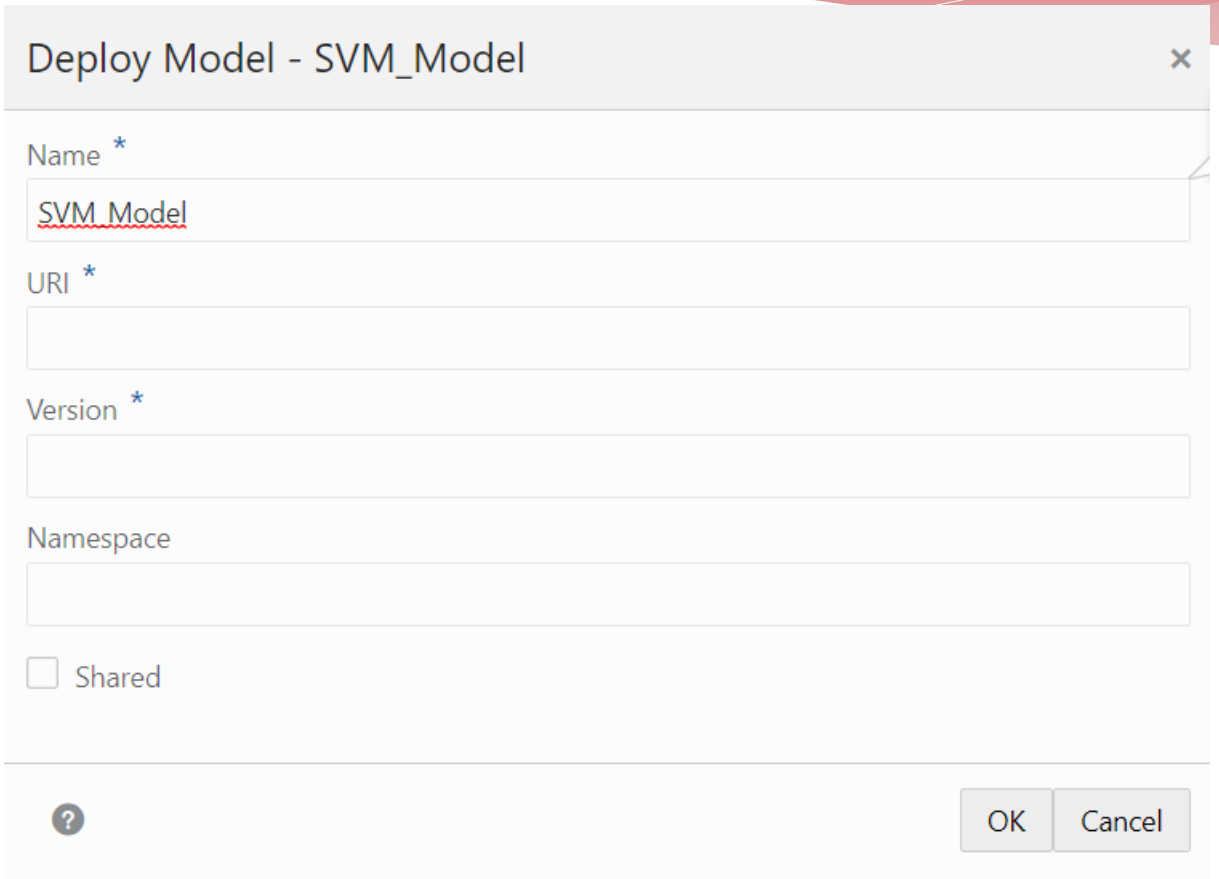
MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy



MIRACLE
Miracle Finland Oy

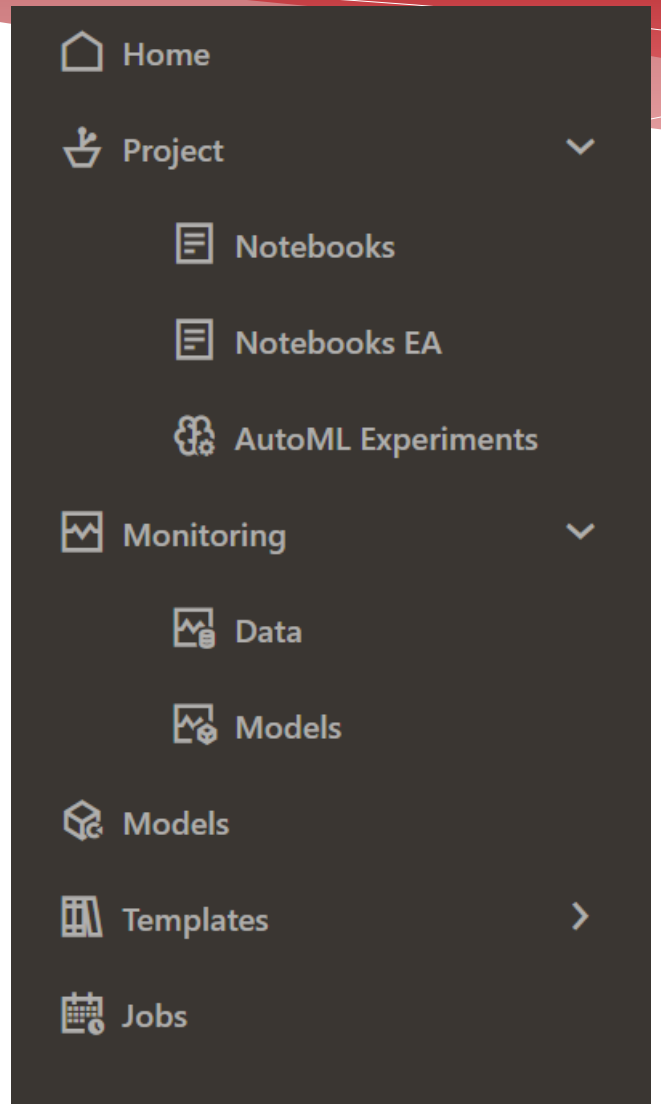
Create a Notebook

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

MIRACLE
Miracle Finland Oy



MIRACLE
Miracle Finland Oy

MyFirstAutoMLNotebook

default

Oracle Machine Learning AutoML UI - Experiment - Generated Notebook

READY

Get proxy object for selected data

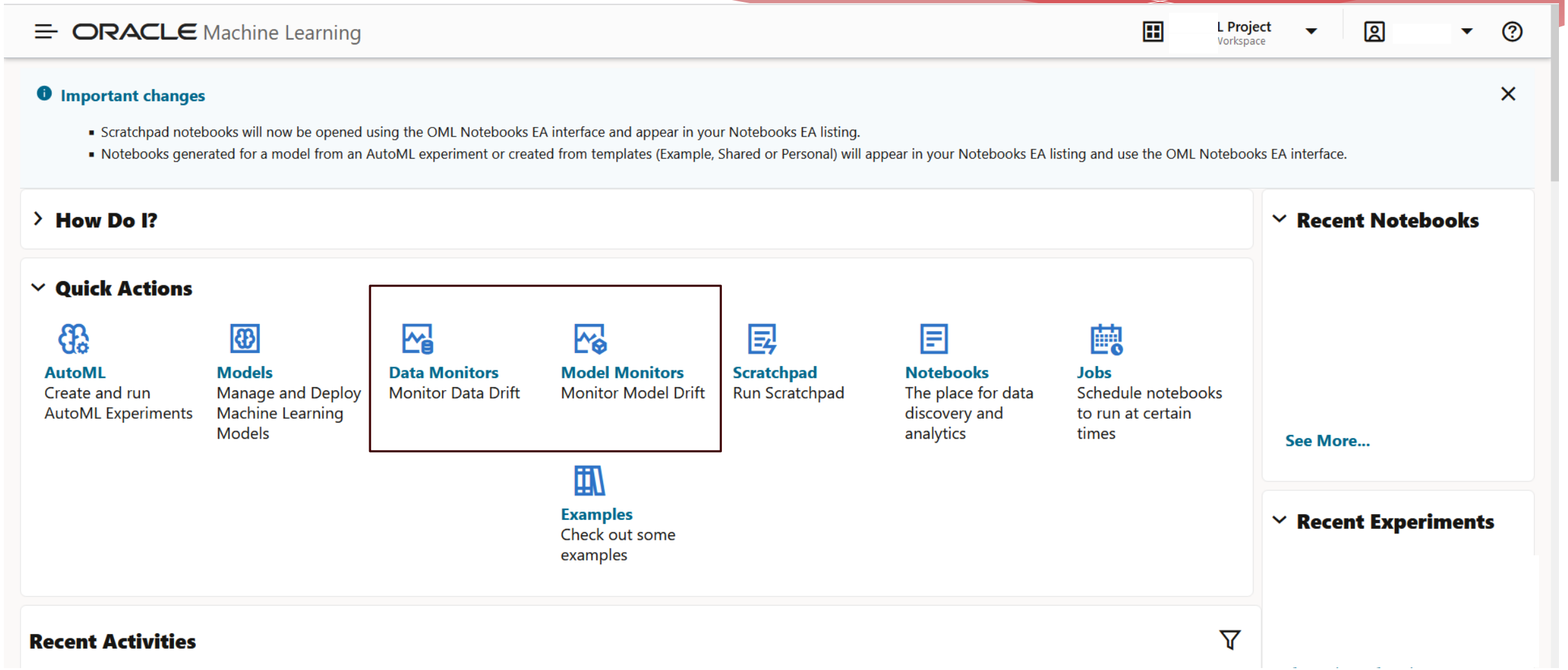
READY

```
%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='"HELIML"'
table='"BREASTCANCER"'

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

build_data = oml.sync(query=query)
z.show(build_data)
```

Prepare training data

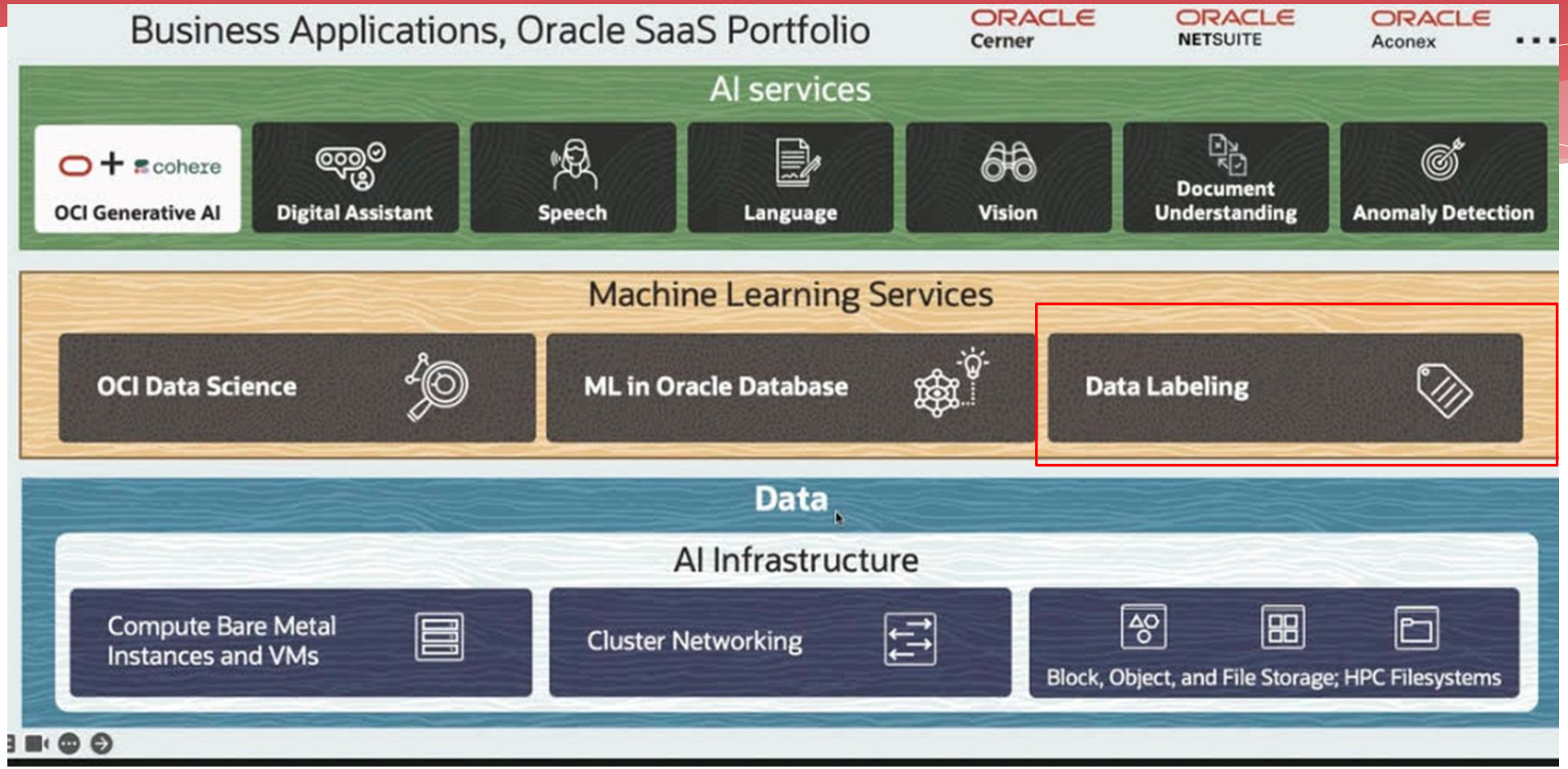
READY

```
%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']]
y_train = build_data[:,'TARGET']
```



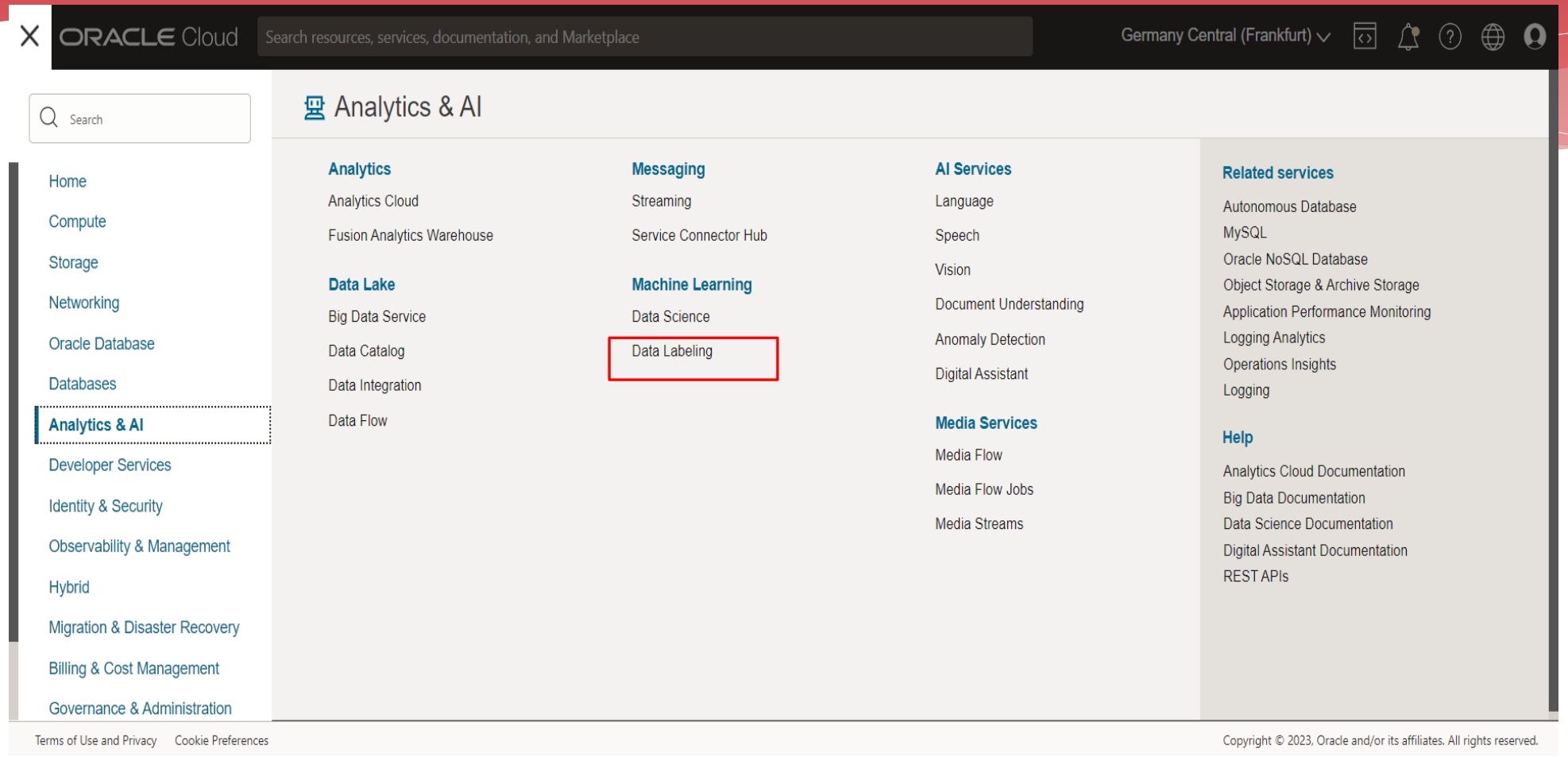
MIRACLE
Miracle Finland Oy

Machine Learning in OCI



MIRACLE
Miracle Finland Oy

Data Labeling



MIRACLE
Miracle Finland Oy

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

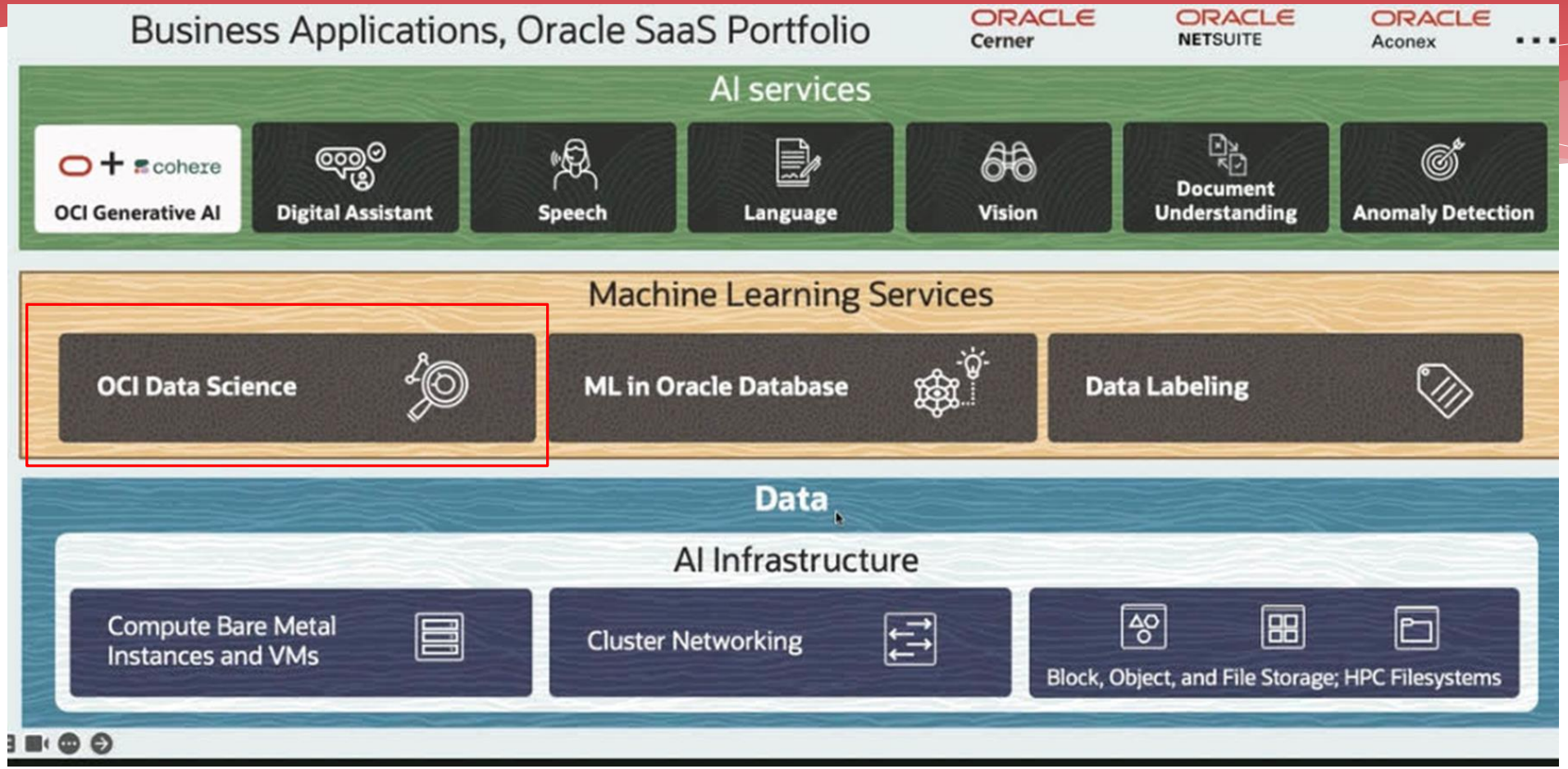
MIRACLE
Miracle Finland Oy

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

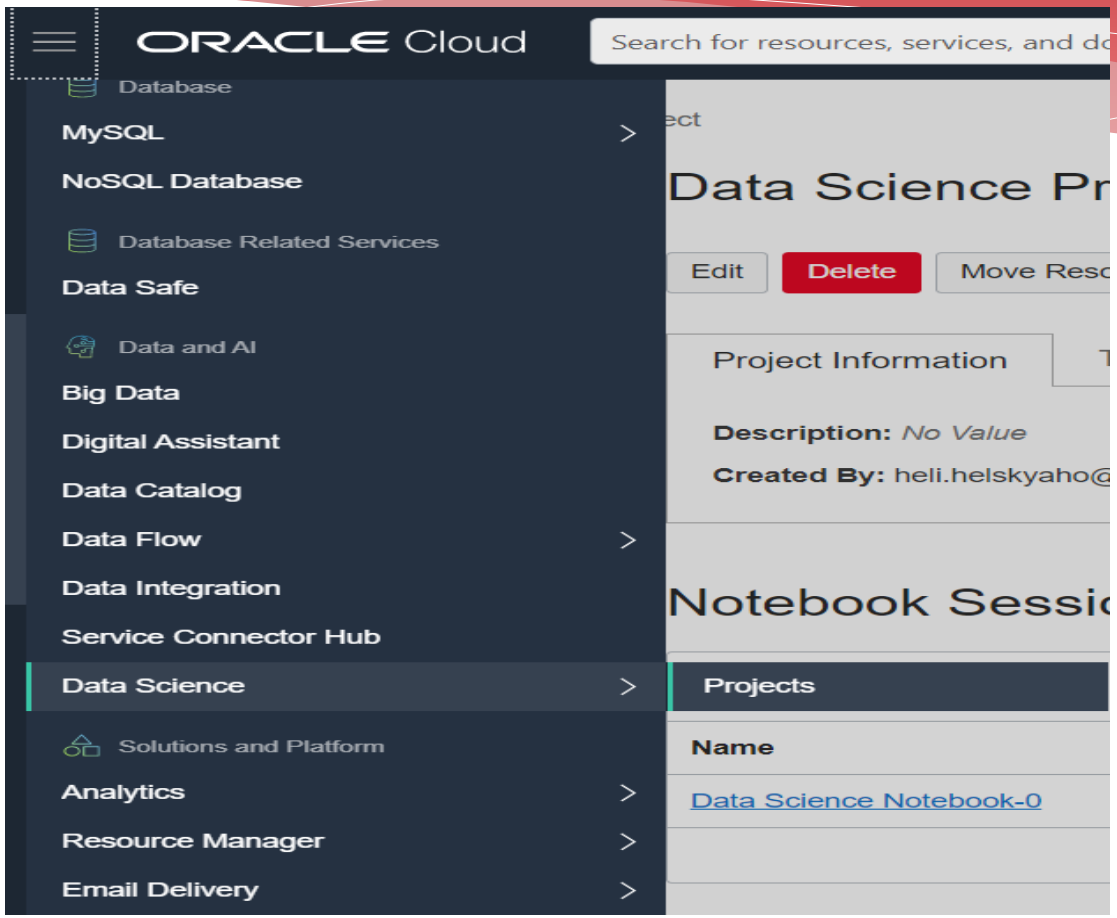
MIRACLE
Miracle Finland Oy

Oracle Machine Learning in OCI



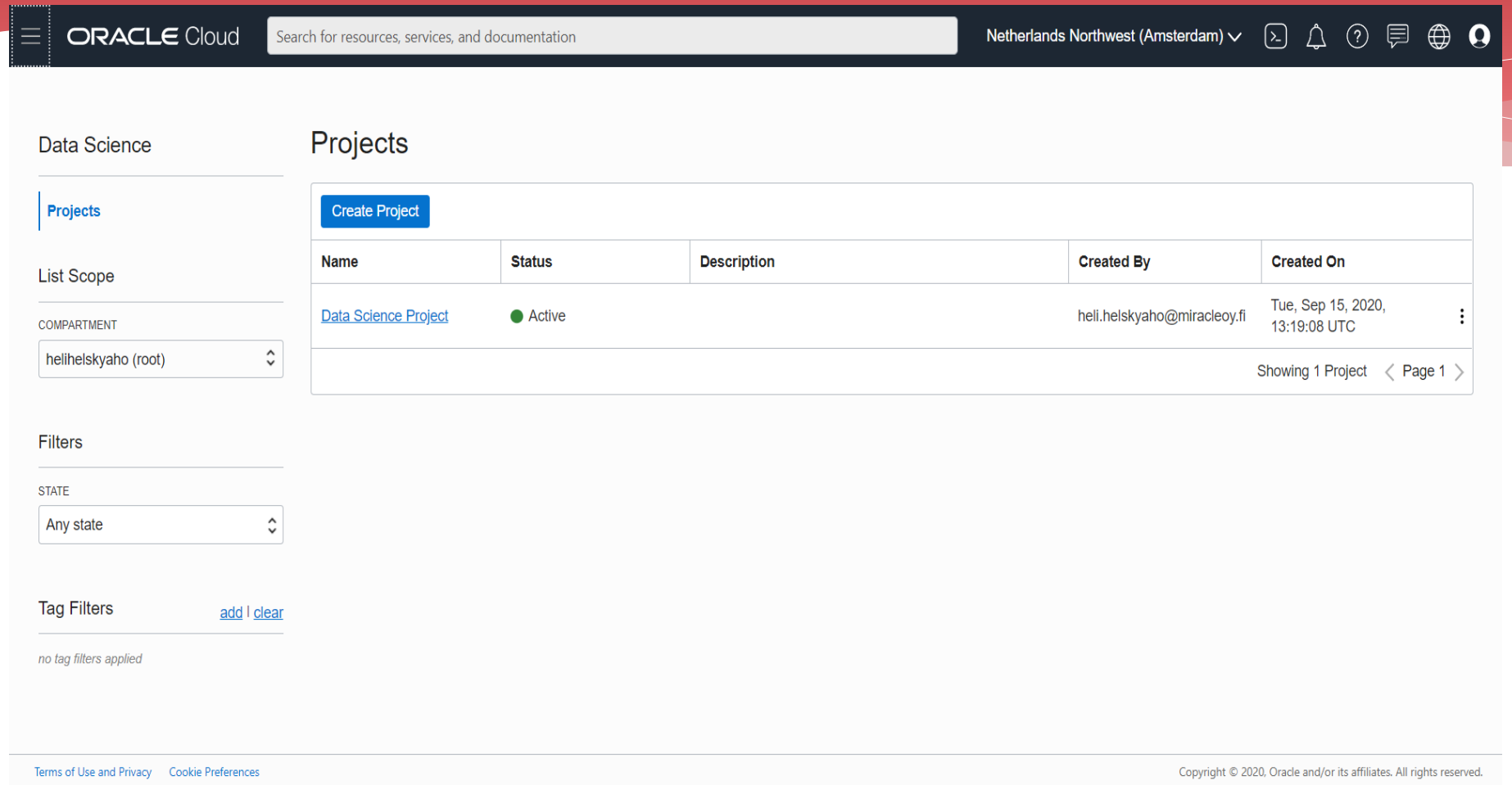
MIRACLE
Miracle Finland Oy

OCI Data Science



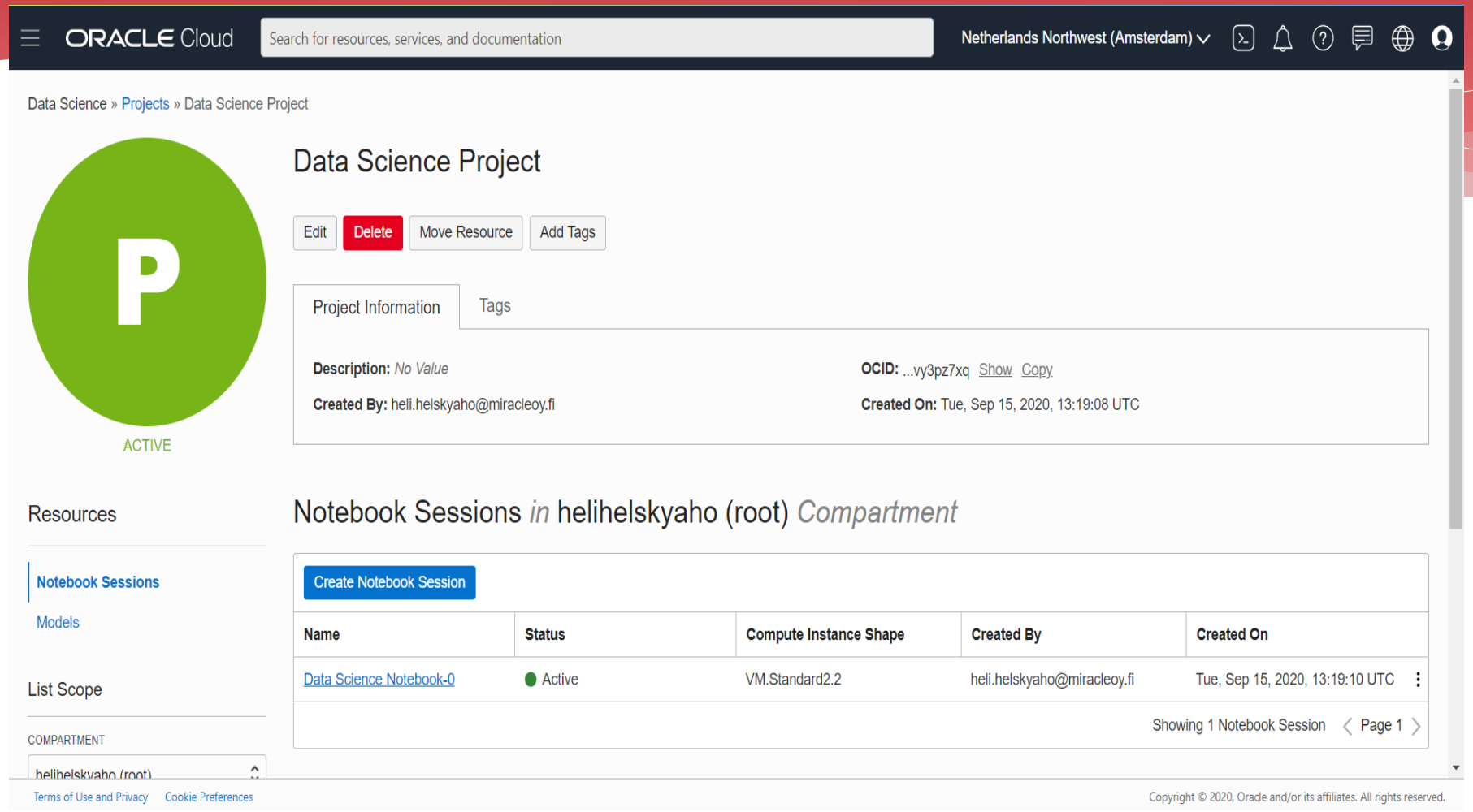
MIRACLE
Miracle Finland Oy

Project

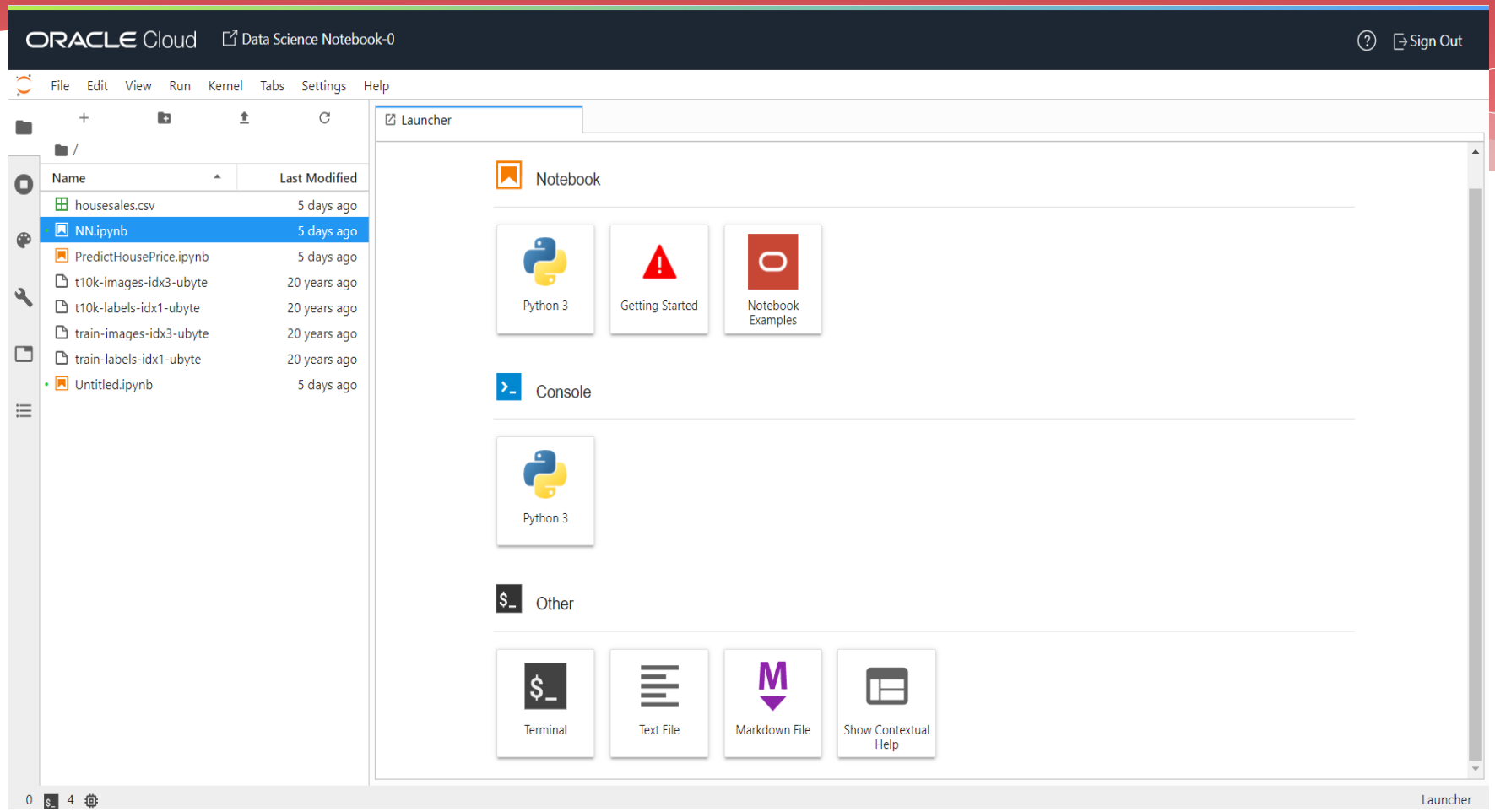


MIRACLE
Miracle Finland Oy

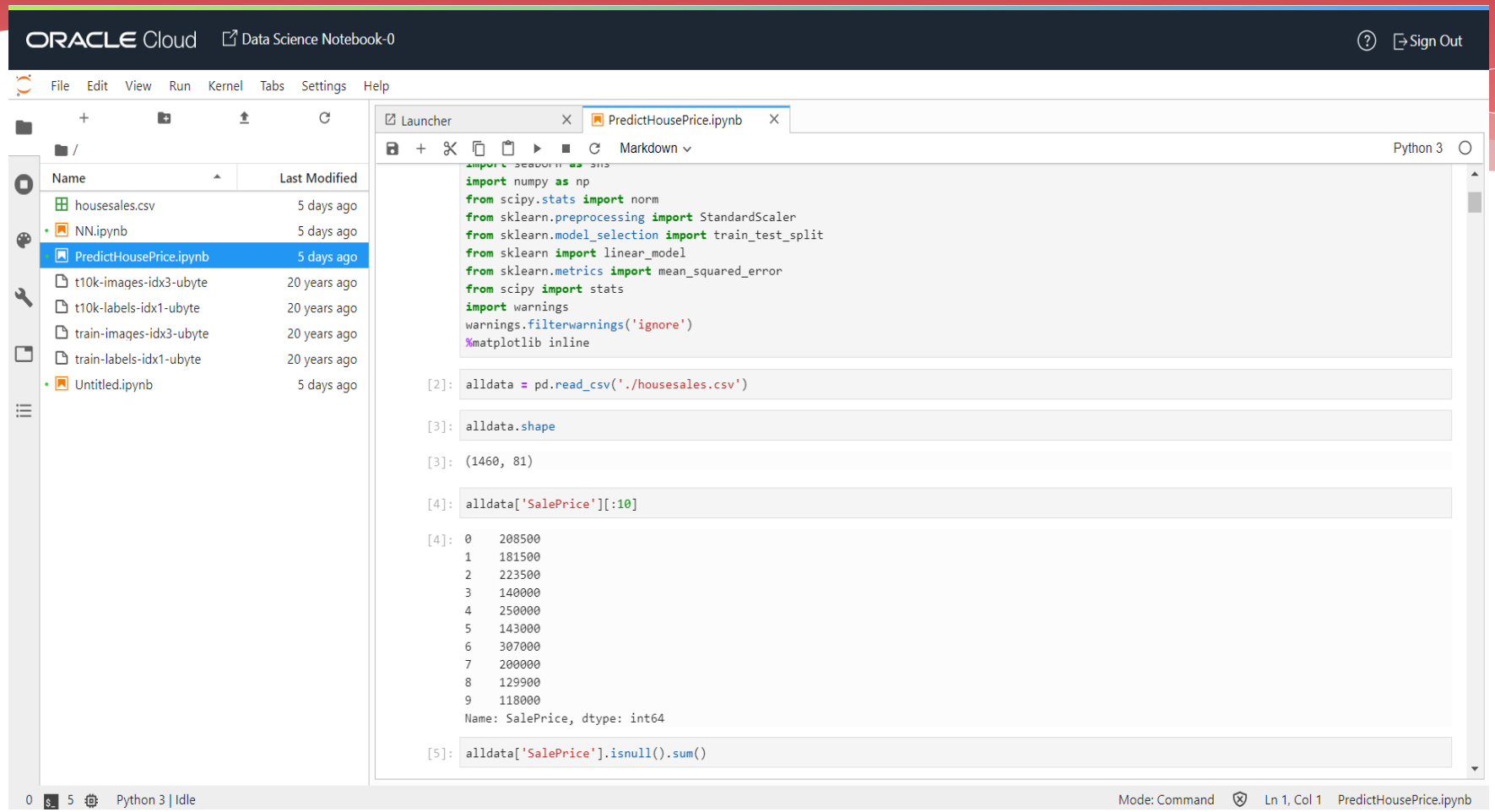
Notebook session



Notebook session



Notebook



MIRACLE
Miracle Finland Oy

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

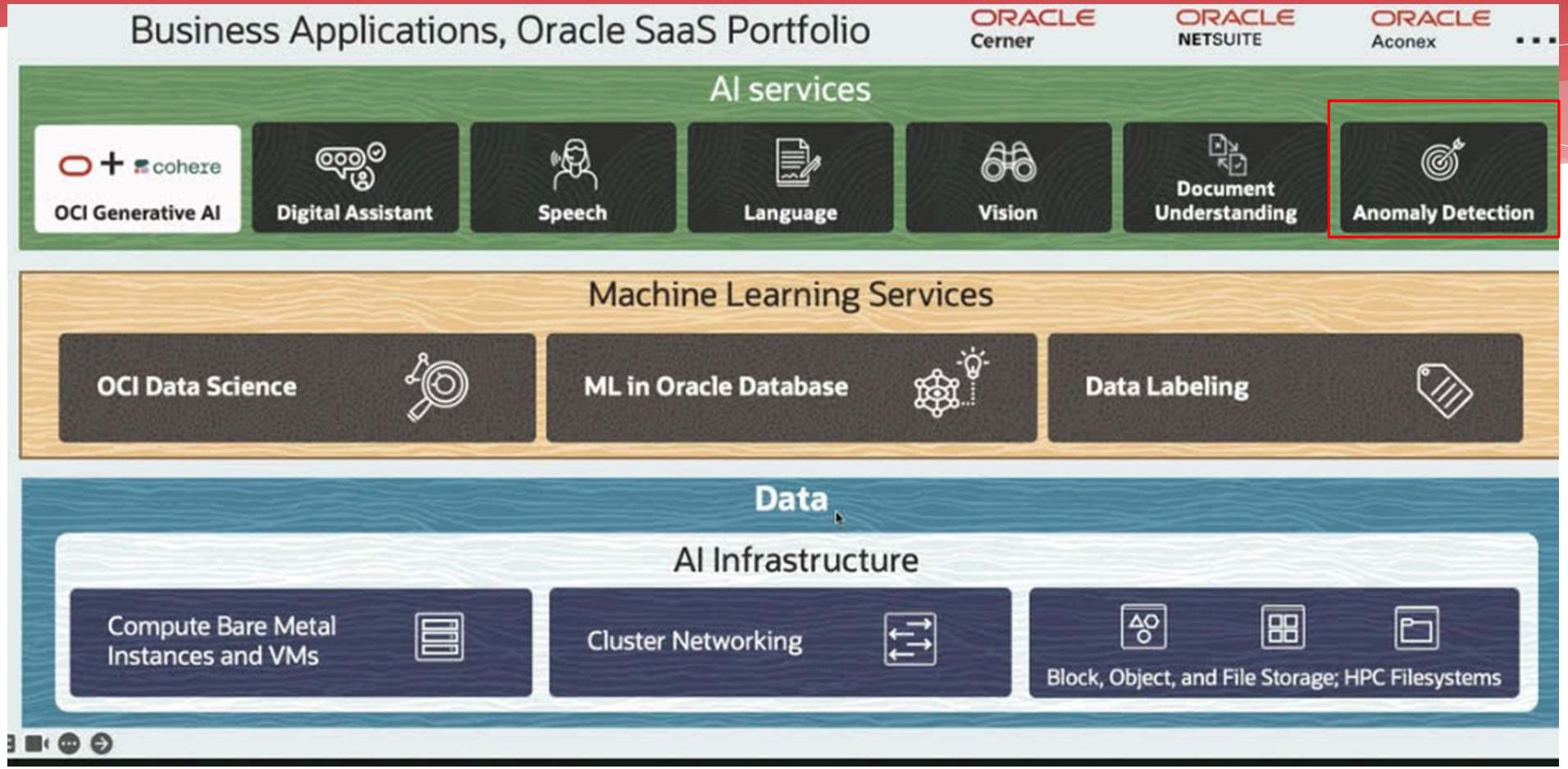
MIRACLE
Miracle Finland Oy

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 explainablility…

MIRACLE
Miracle Finland Oy

Machine Learning in OCI

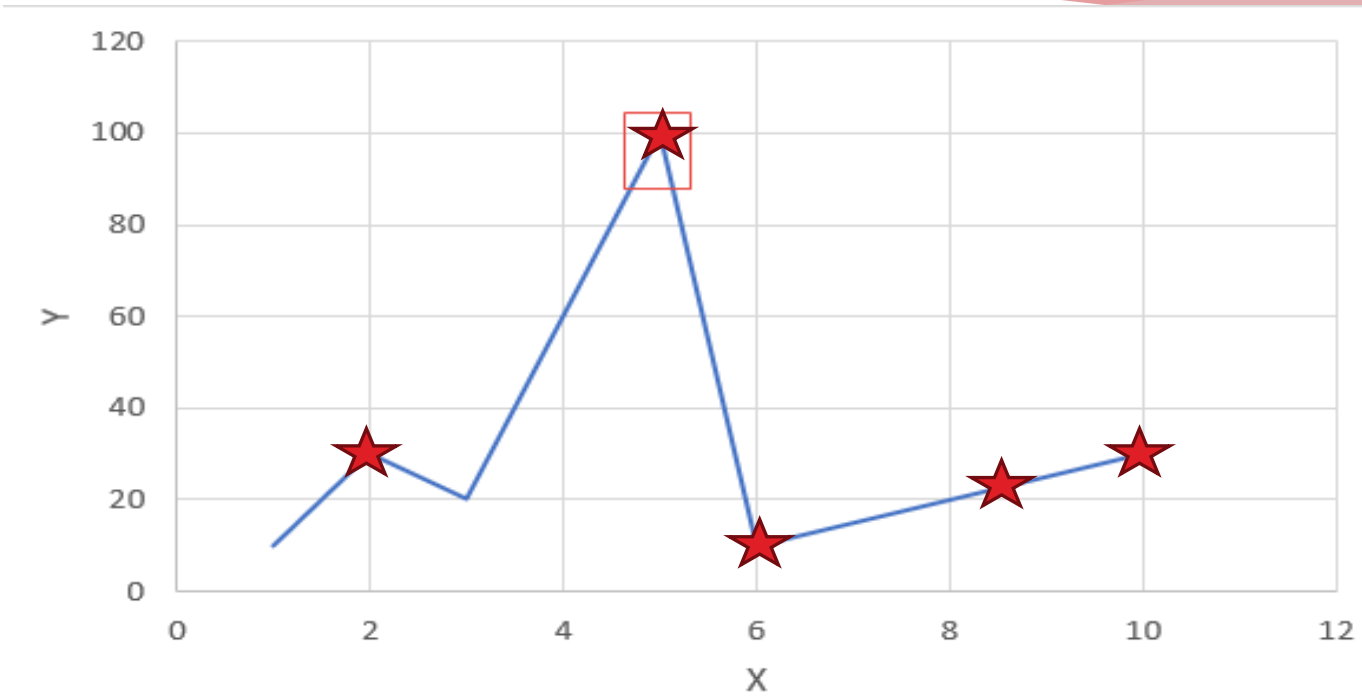


MIRACLE
Miracle Finland Oy

Anomaly Detection

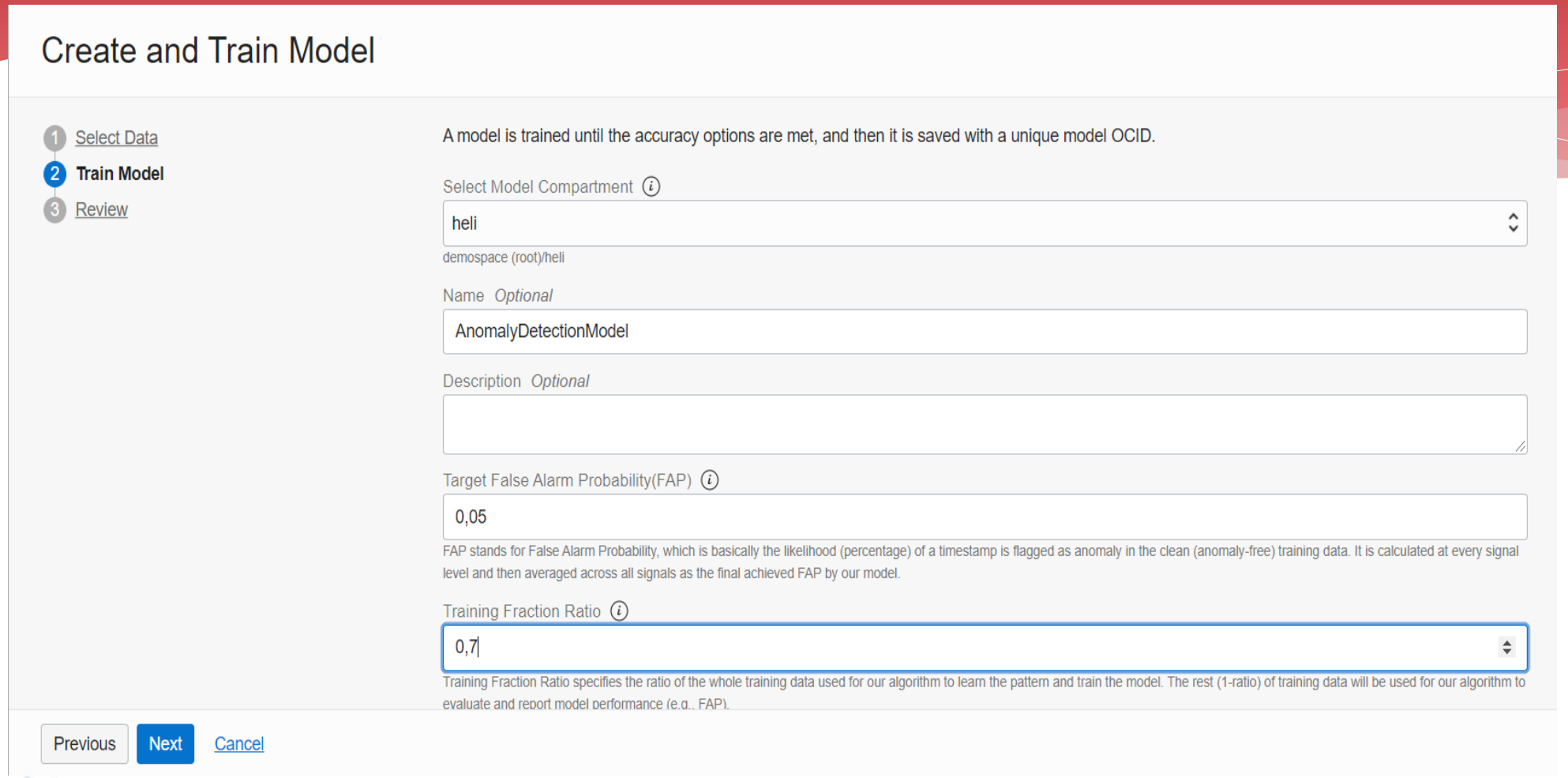
* Identificating rare events/observations, very different from majority of the data

Anomaly Detection



MIRACLE
Miracle Finland Oy

Anomaly Detection, Create and Train Model



Miracle Finland Oy

Anomaly Detection, Detect Anomaly

Anomalies

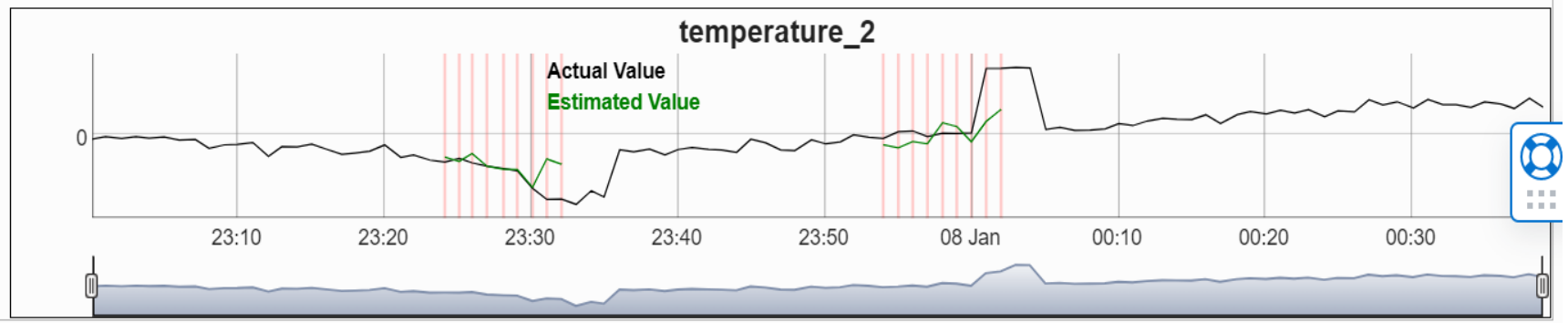
Detect the anomalies for the data contained in the request using the stored model

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.

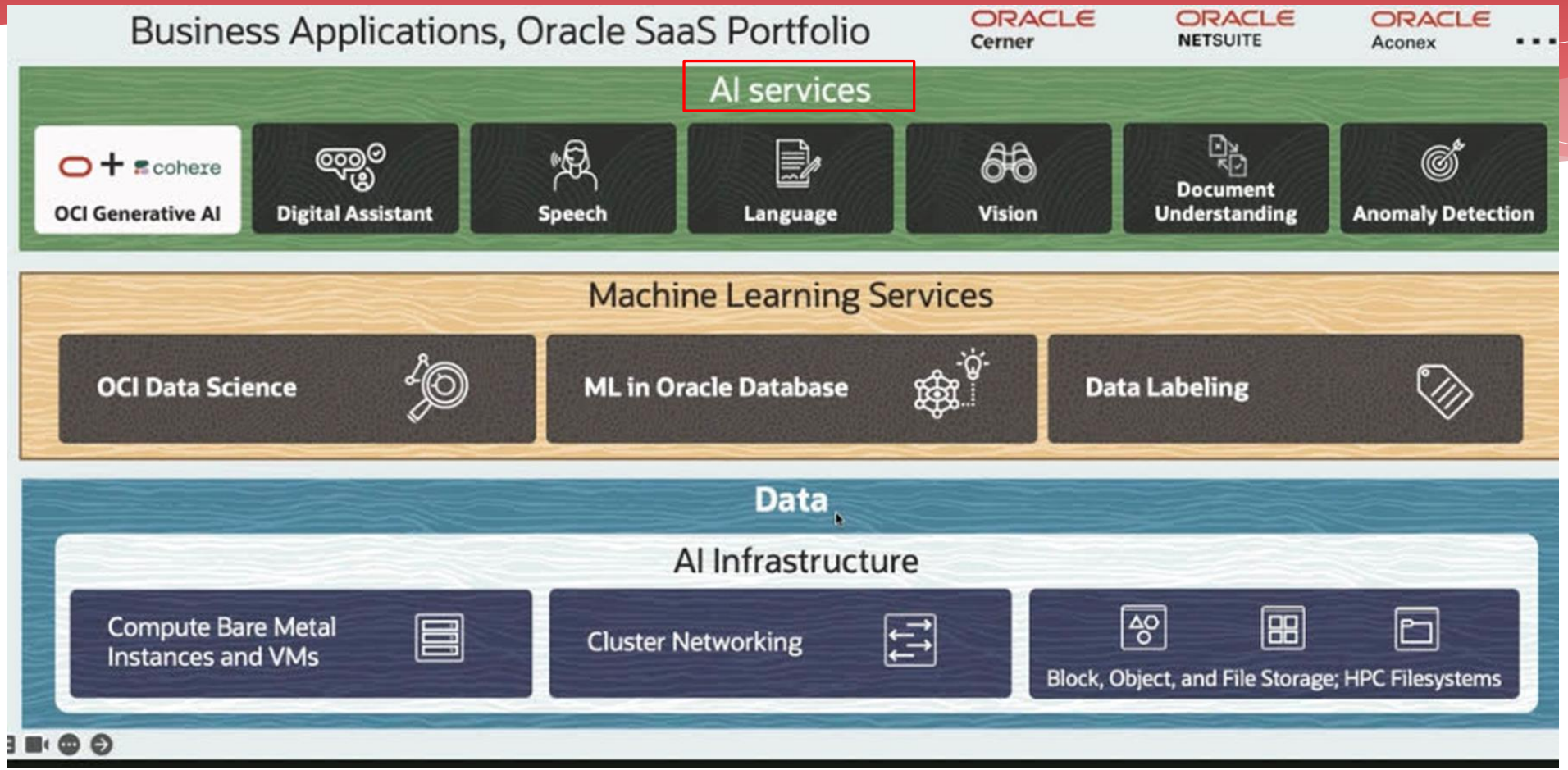
Select column labels(with anomalies) for visualization

Select column labels(with anomalies) for visualization

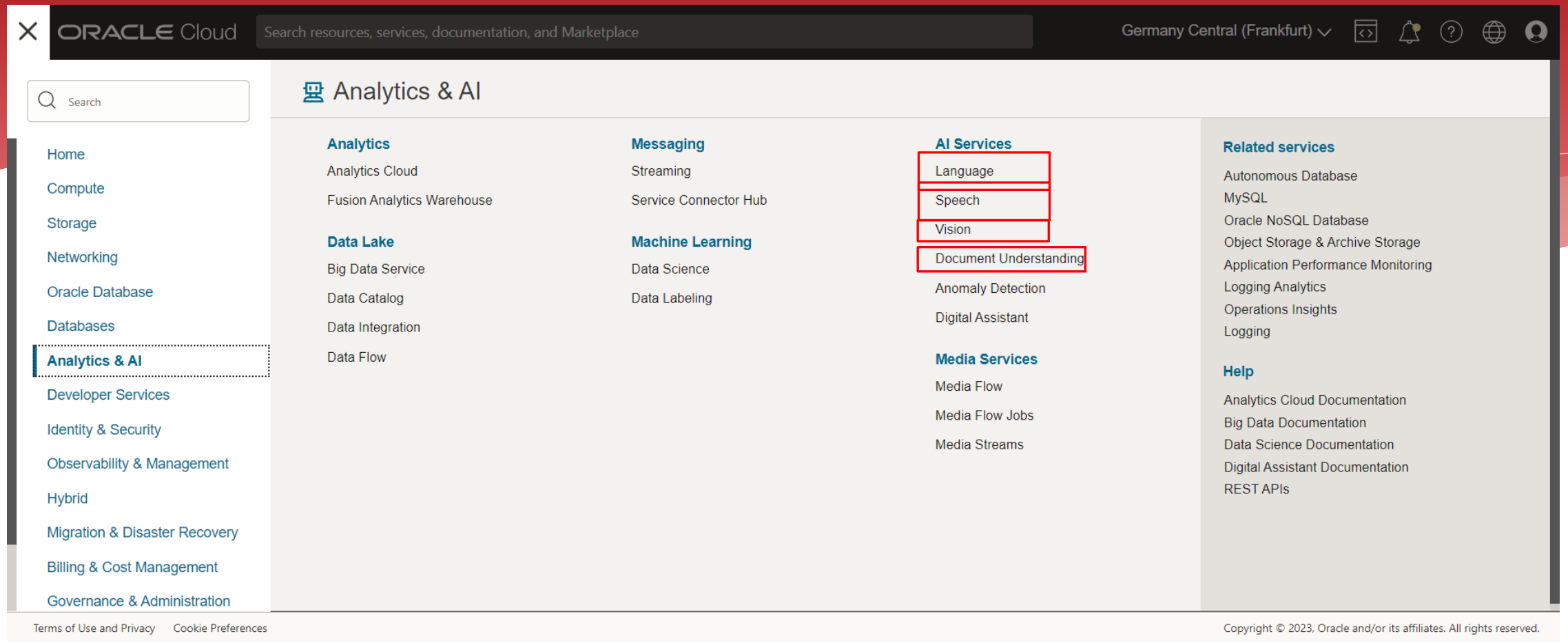


MIRACLE
Miracle Finland Oy

Machine Learning in OCI



MIRACLE
Miracle Finland Oy



AI Services

Generative AI

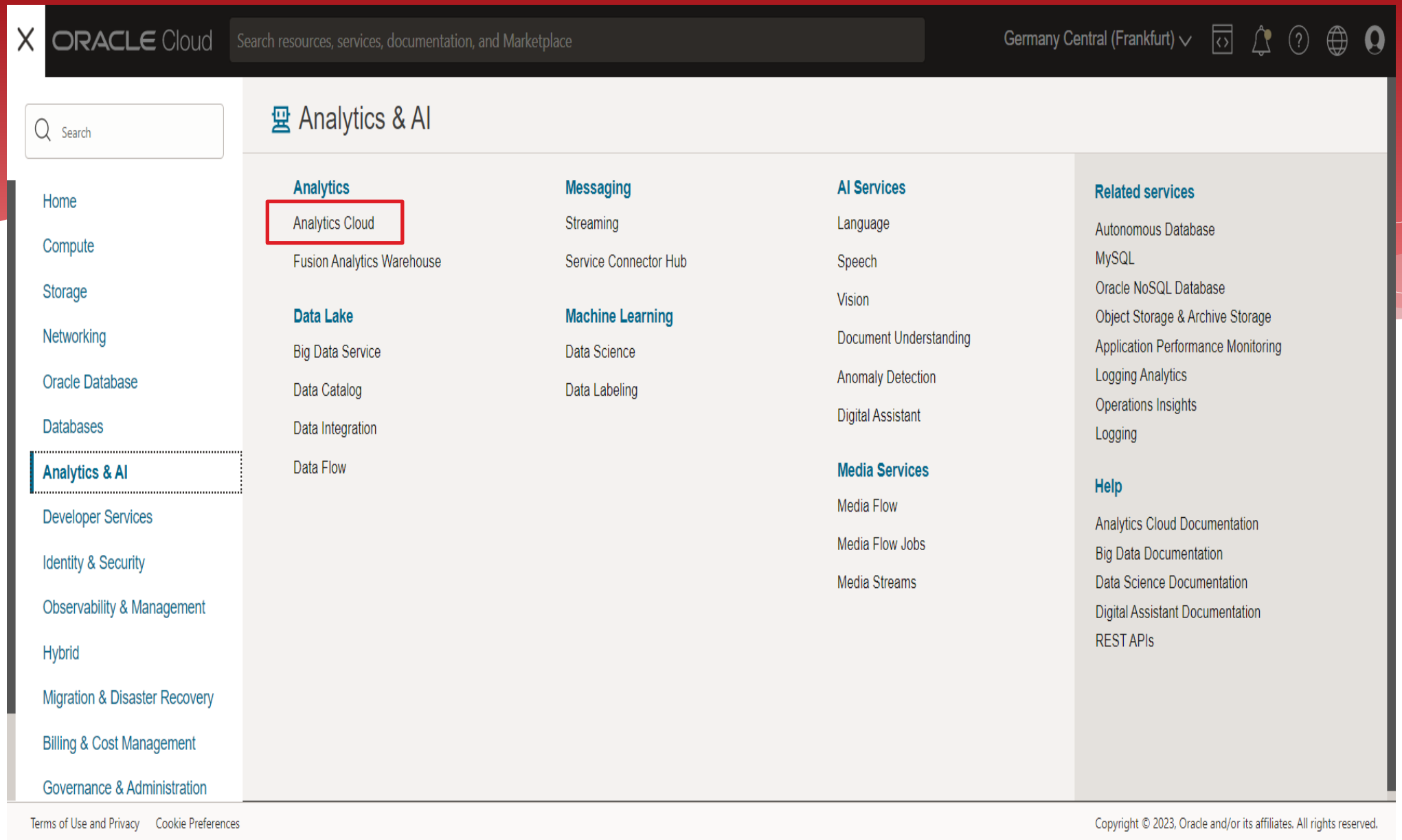
Oracle Analytics Cloud, OAC

MIRACLE
Miracle Finland Oy

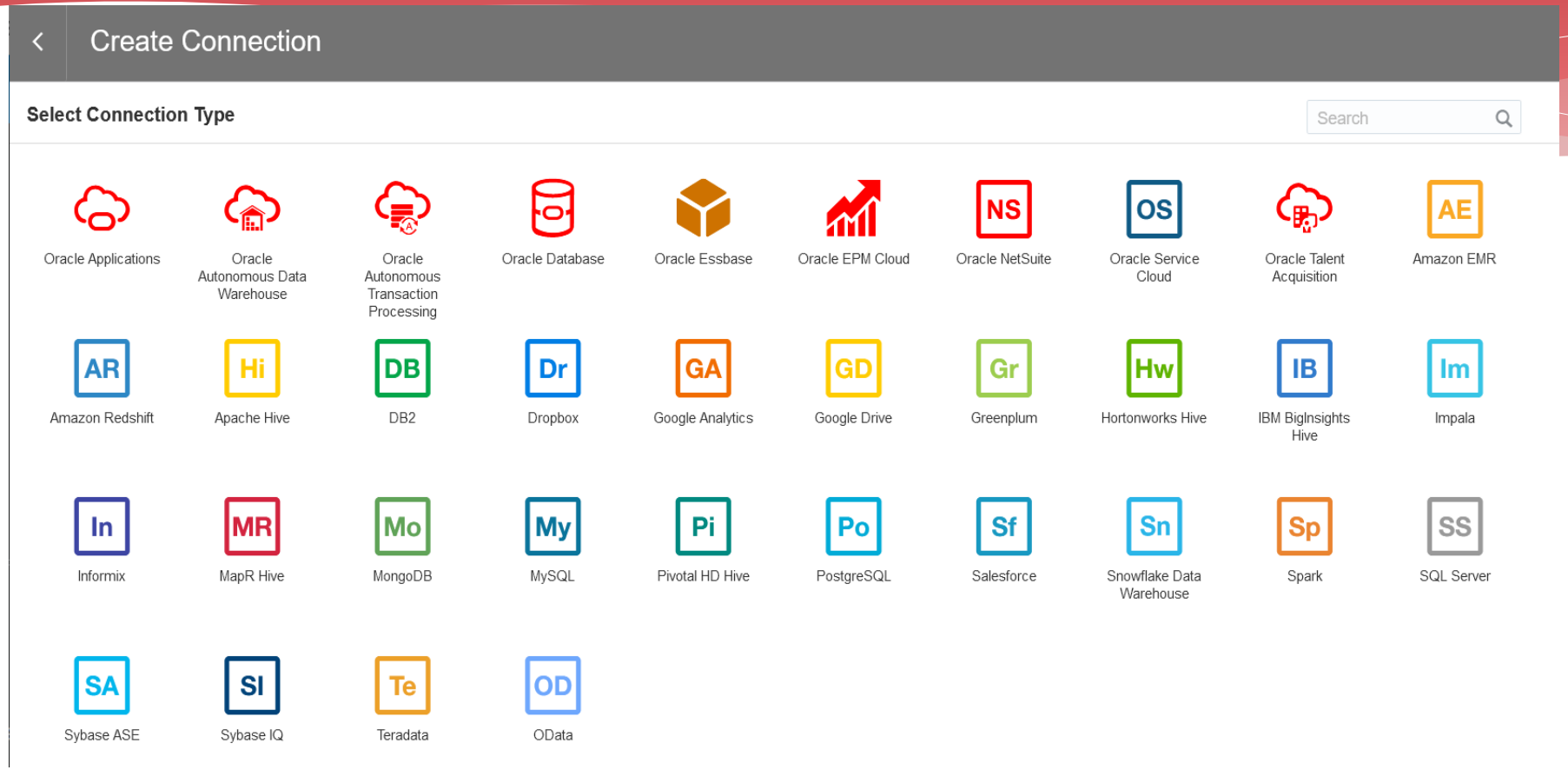
Oracle Analytics Cloud (OAC) is…

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

MIRACLE
Miracle Finland Oy



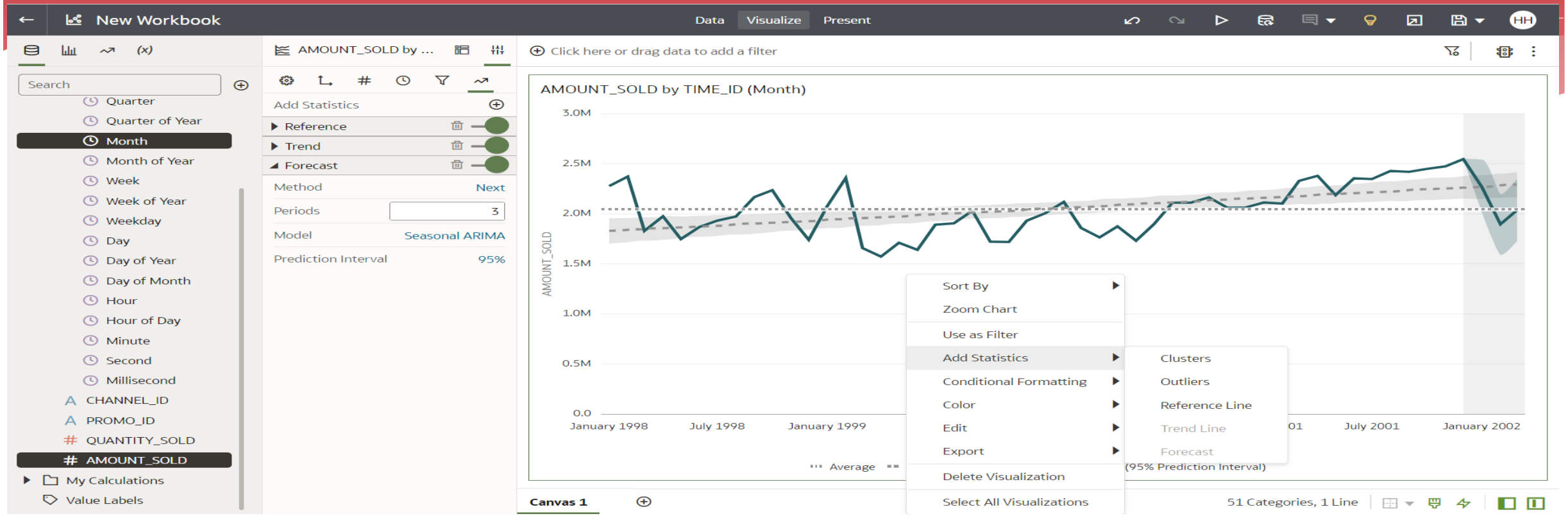
Plenty of data source candidates



Machine Learning?

MIRACLE
Miracle Finland Oy

ML build in OAC



Insights

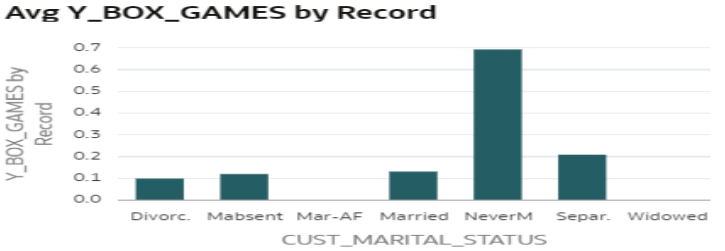
We found the following insights for this dataset

This table shows overall aggregation of metrics selected by Auto-Insights process. Columns that were selected from this dataset were, Metrics : CUST_CREDIT_LIMIT, Y_BOX_GAMES,...

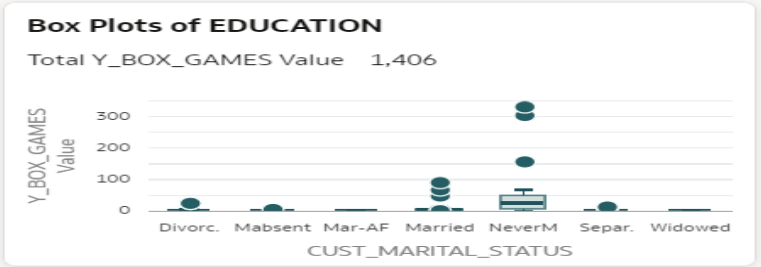
Measures Overview

Row Count	4,500
CUST_CREDIT_LIMIT Value	35,659,000
Y_BOX_GAMES Value	1,406
CUST_CREDIT_LIMIT by Record	7,924.22
Y_BOX_GAMES by Record	0.31
Ratio CUST_CREDIT_LIMIT, Y_BOX_GAMES	25,362.02

This visualization shows average value of Y_BOX_GAMES by record for each member of CUST_MARITAL_STATUS. We identified an interesting variance for these values between...



This Boxplot visualization helps understand comparative spread of EDUCATION individuals measured by Y_BOX_GAMES value, for each of the CUST_MARITAL_STATUS. We detected...



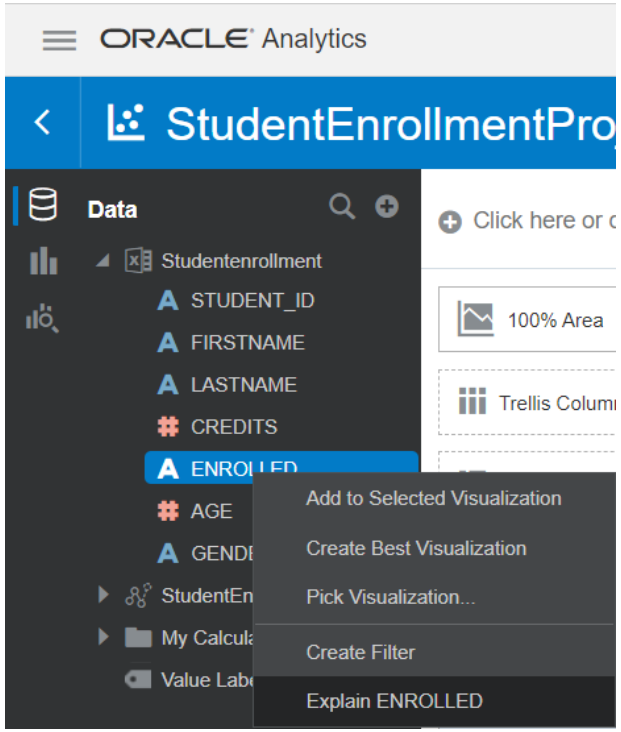
This heatmap shows Y_BOX_GAMES value (% of total) for each intersection of HOUSEHOLD_SIZE and AFFINITY_CARD dimensions. We detected that a few cells have significantly...

% of Y_BOX_GAMES

	1	2	3	4-5	6-
0	41.11%	26.60%	12.45%	1.99%	6.4
1	0.28%	0.85%	2.77%	0.43%	0.0
Grand Total	41.39%	27.45%	15.22%	2.42%	6.4

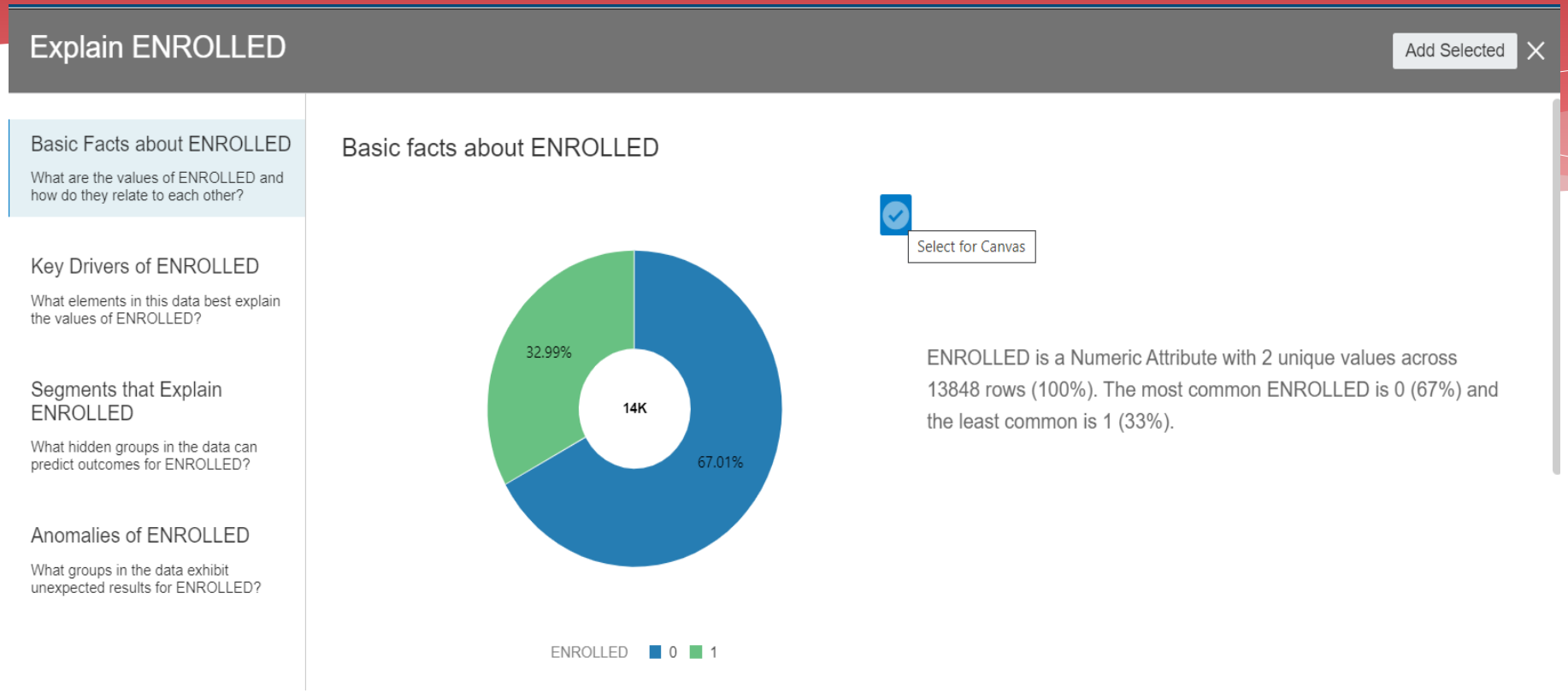
MIRACLE
Miracle Finland Oy

Explain



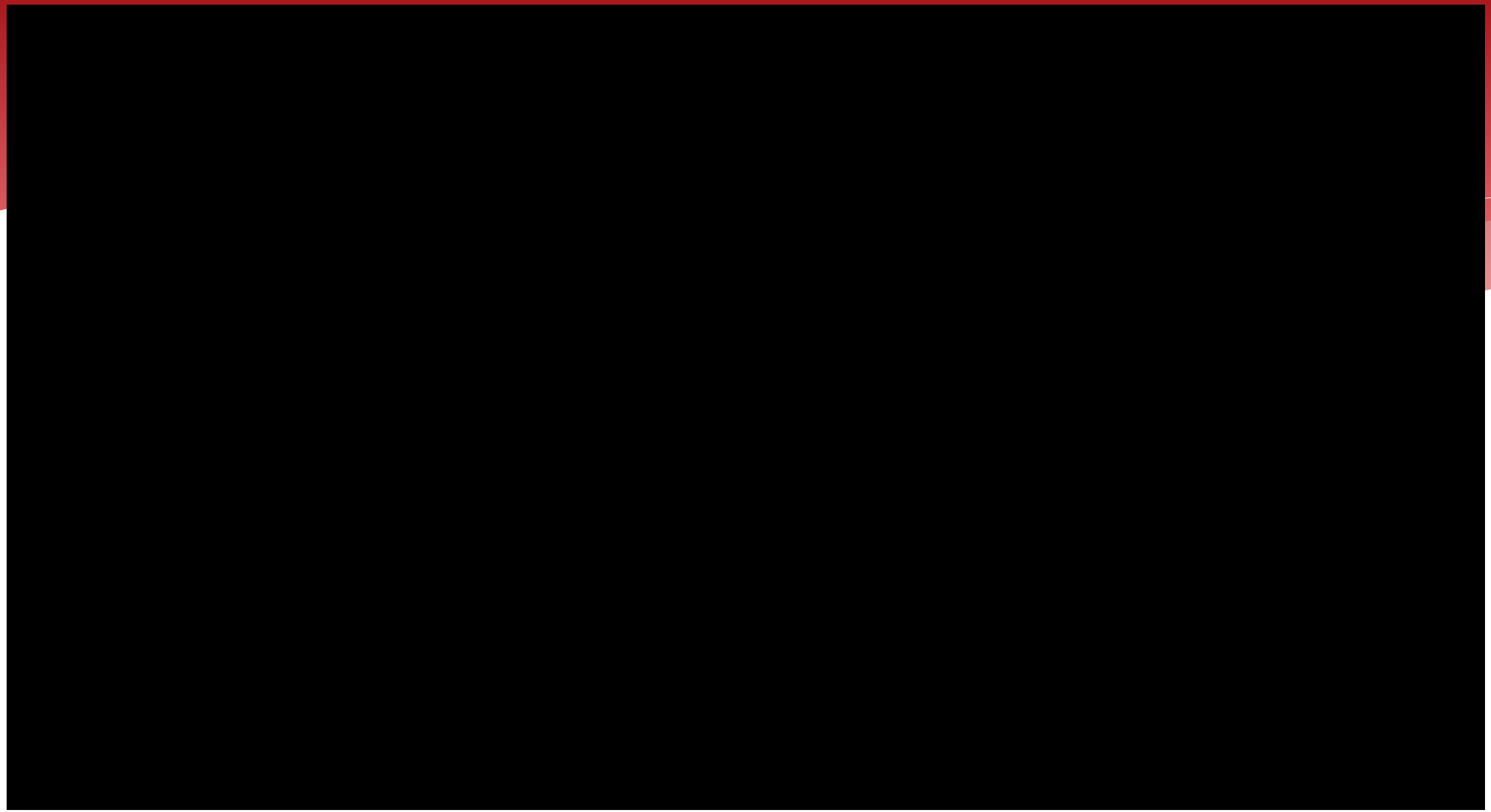
MIRACLE
Miracle Finland Oy

Explain

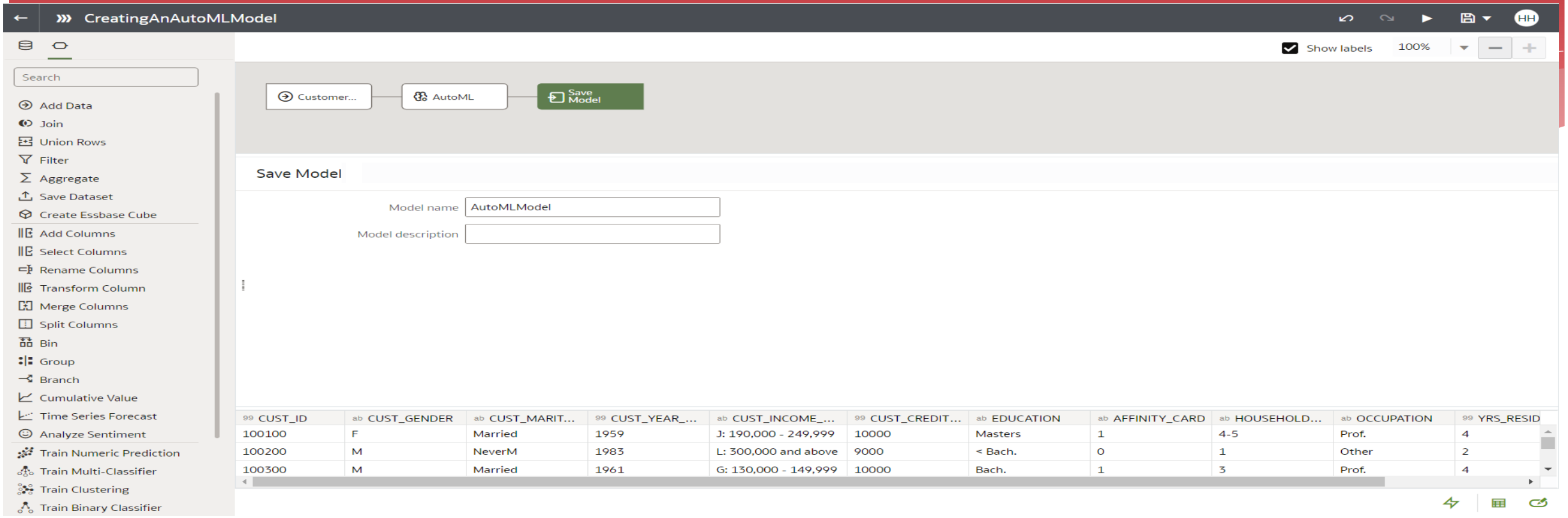


Creating Models in OAC, Classification

Miracle Finland Oy

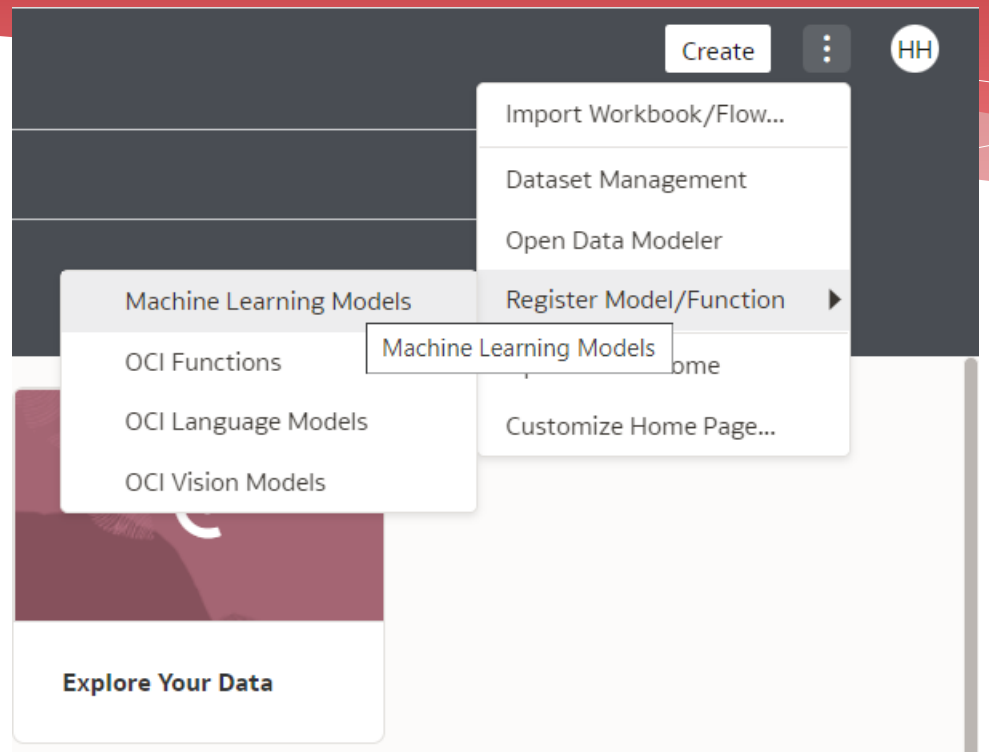


AutoML

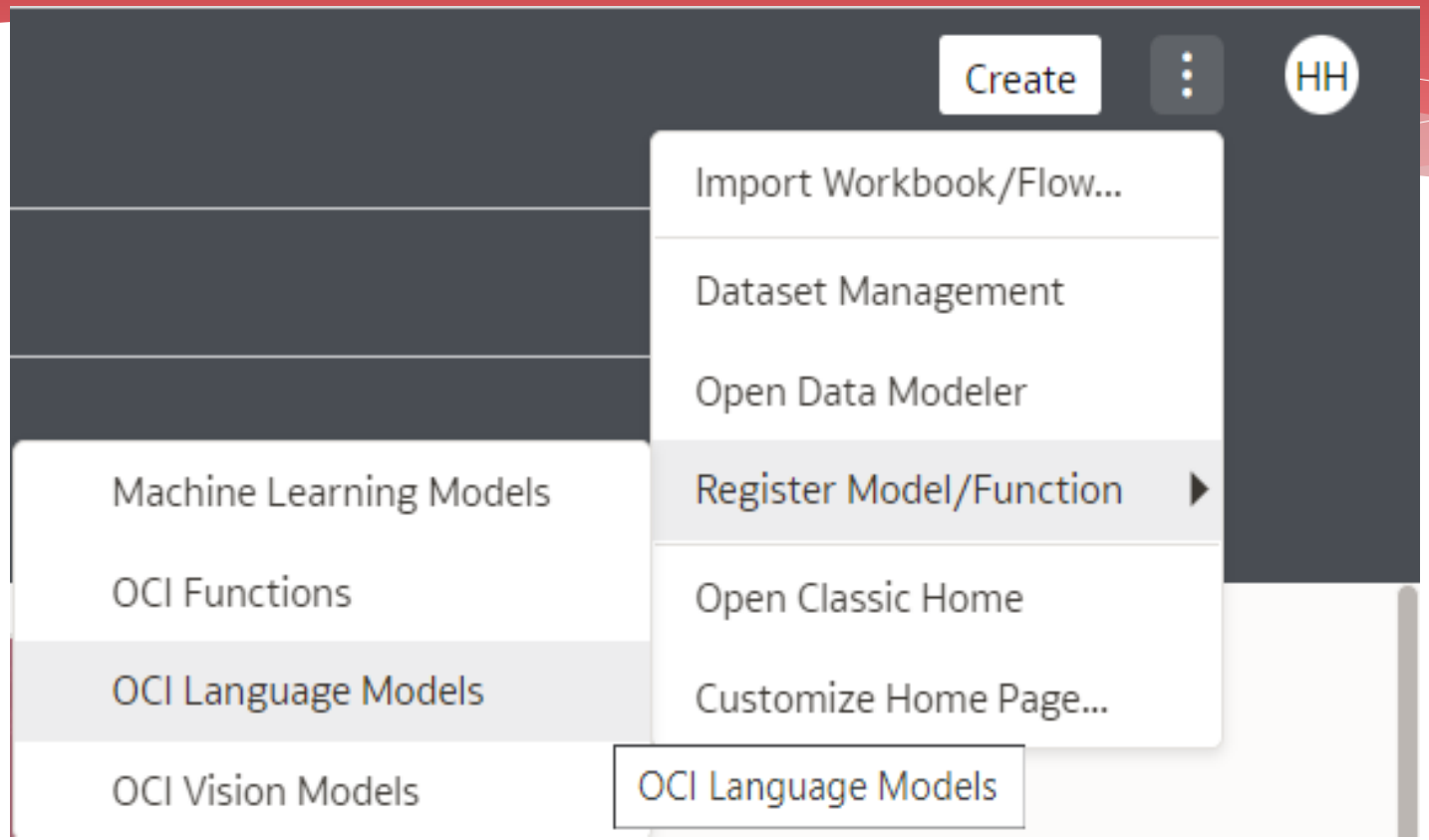


99 CUST_ID	ab CUST_GENDER	ab CUST_MARIT...	99 CUST_YEAR_...	ab CUST_INCOME_...	99 CUST_CREDIT...	ab EDUCATION	ab AFFINITY_CARD	ab HOUSEHOLD...	ab OCCUPATION	99 YRS_RESID
100100	F	Married	1959	J: 190,000 - 249,999	10000	Masters	1	4-5	Prof.	4
100200	M	NeverM	1983	L: 300,000 and above	9000	< Bach.	0	1	Other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

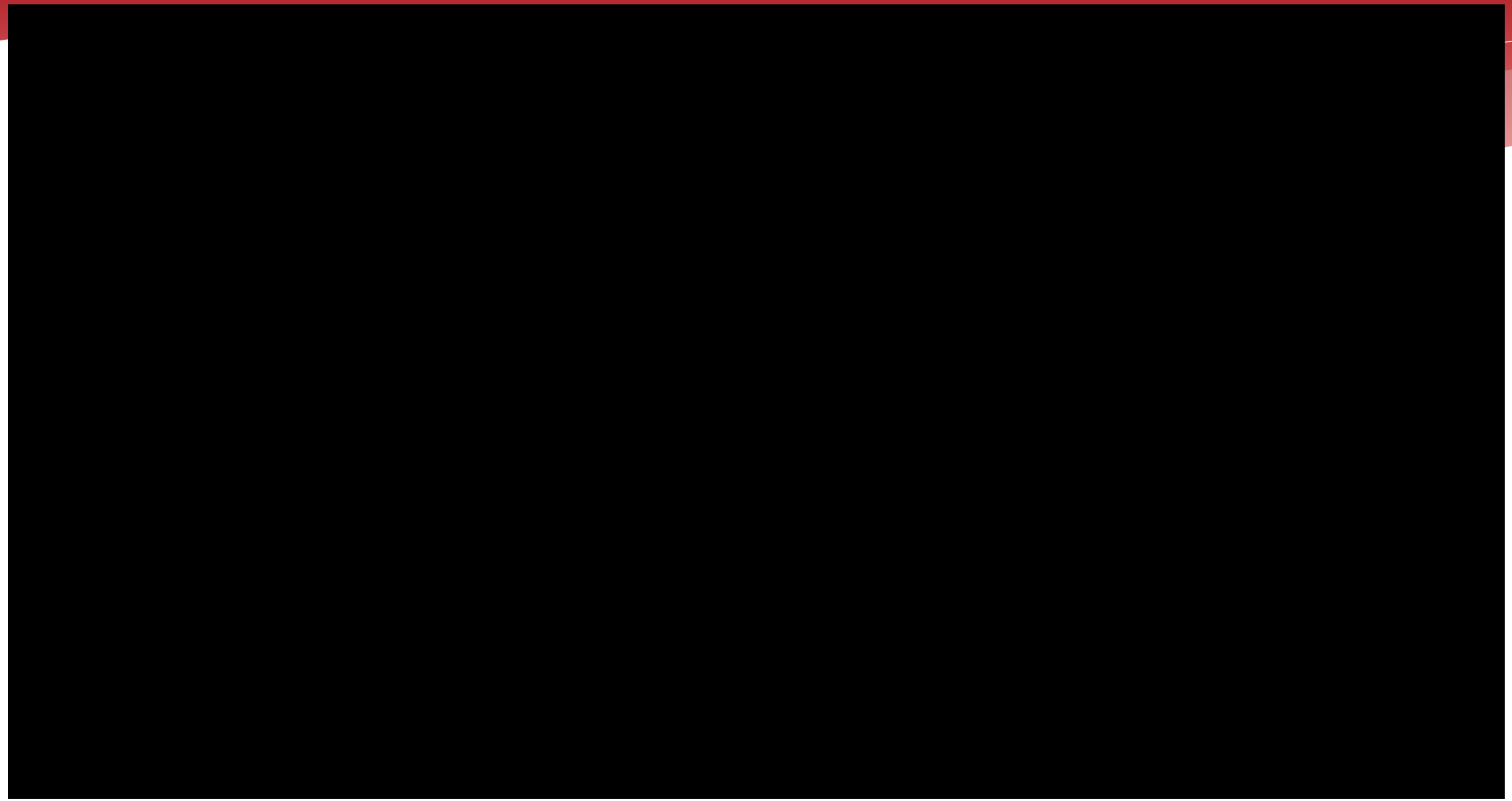
Use OML models from the database



AI Services



Using a model in OAC



MIRACLE
Miracle Finland Oy

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

MIRACLE
Miracle Finland Oy

Conclusions

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

MIRACLE
Miracle Finland Oy

THANK YOU!

QUESTIONS?

Email: heli@miracleoy.fi
Twitter: @HeliFromFinland
Blog: Helifromfinland.com

MIRACLE
Miracle Finland Oy