

Oracle Machine Learning AutoML User Interface on Autonomous Database

Automated machine learning for data scientists,
simplified machine learning for citizen data
scientists, and accelerated machine learning
model deployment for application developers

March 2021, Version 1.0
Copyright © 2021, Oracle and/or its affiliates
Public

Purpose statement

This document provides an overview of features included in Oracle Machine Learning AutoML User Interface. It is intended solely to help you assess the business benefits of OML AutoML User Interface and to plan your I.T. projects.

Disclaimer

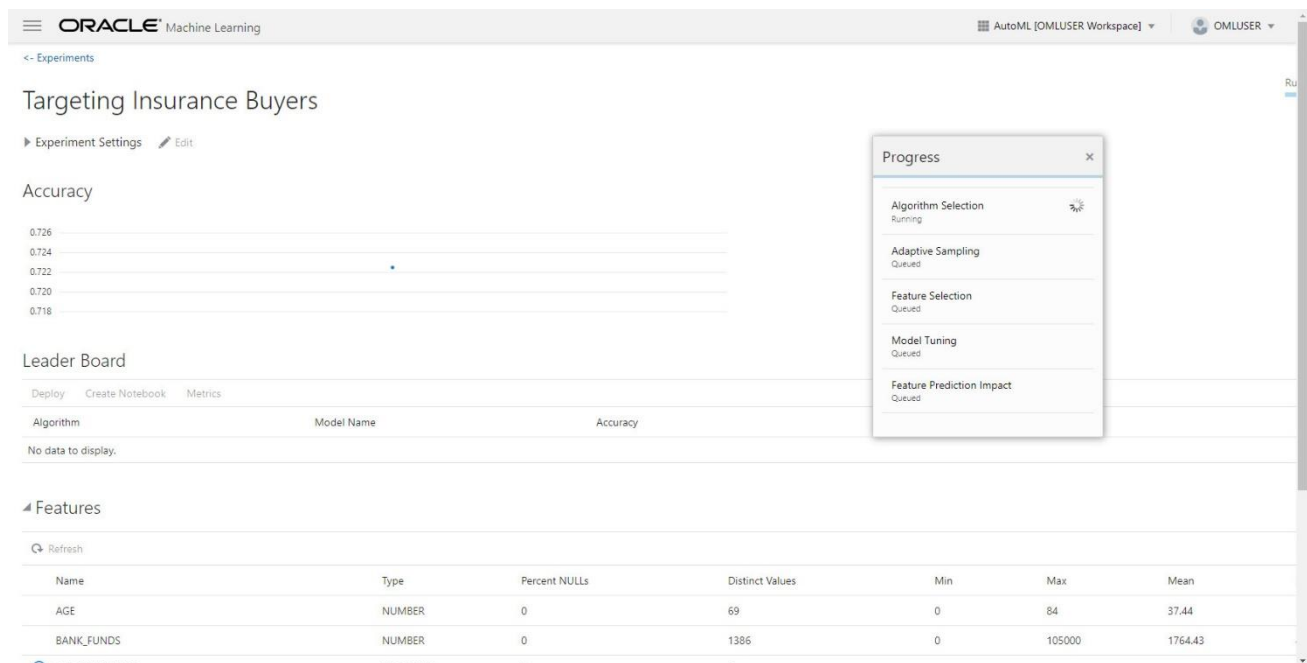
This document in any form, software or printed matter, contains proprietary information that is the exclusive property of Oracle. Your access to and use of this confidential material is subject to the terms and conditions of your Oracle software license and service agreement, which has been executed and with which you agree to comply. This document and information contained herein may not be disclosed, copied, reproduced or distributed to anyone outside Oracle without prior written consent of Oracle. This document is not part of your license agreement nor can it be incorporated into any contractual agreement with Oracle or its subsidiaries or affiliates.

This document is for informational purposes only and is intended solely to assist you in planning for the implementation and upgrade of the product features described. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described in this document remains at the sole discretion of Oracle. Due to the nature of the product architecture, it may not be possible to safely include all features described in this document without risking significant destabilization of the code.

Summary

Oracle Machine Learning (OML) in Oracle Autonomous Database provides access to in-database machine learning (ML) algorithms and functionality. OML AutoML User Interface (UI) makes machine learning easy—providing an easy to use interface that automates repetitive, time-consuming tasks normally taken by expert data scientists, while simplifying machine learning for non-expert users. OML AutoML UI accelerates the machine learning process from model building to model deployment.

With just a few “clicks”, users specify a data table and the target attribute and OML AutoML UI builds several models for you to consider. OML AutoML UI automatically preprocesses the data, picks the best in-database candidate algorithm(s) for the experiment, selects the right input data samples and features to improve model quality, and speeds up model tuning. OML AutoML UI builds OML models using the selected algorithms, tunes their hyperparameters, and displays accuracy metrics so users can select the model that best meets their needs.



Oracle Machine Learning AutoML User Interface is an easy-to-use tool that automates routine machine learning steps, including algorithm and feature selection, model tuning and deployment using in-database OML algorithms. OML AutoML UI enables citizen data scientist to build and select machine learning models, increases data scientist productivity, and accelerates model deployment.

What is Different about OML AutoML User Interface?

Oracle Machine Learning AutoML User Interface makes it easy for citizen data scientists and expert data scientists to build and deploy machine learning models. OML AutoML UI, a new component of Oracle Machine Learning on Oracle Autonomous Database, provides a no-code browser-based interface that automates the machine learning modeling process and simplifies deployment to just a few clicks.

While data scientists may prefer to write code, perhaps using the OML4Py AutoML feature, OML AutoML UI makes machine learning possible for a much broader set of users, who may “know enough ML to be dangerous” but are not familiar with specific algorithm details and the modeling process to refine usable models.

At the same time, OML AutoML UI delivers significant productivity improvements for data scientists as a modeling accelerator—allowing automation to produce an initial model, but then get the specific hyperparameters to continue tuning or to augment the model directly in a notebook generated by OML AutoML UI that reproduces OML4Py code for the specified model.

OML AutoML UI simplifies machine learning, automates many routine but time-consuming steps, and increases data scientist productivity while reducing the overall compute time required to deliver machine learning models. With OML AutoML UI:

- Citizen data scientists quickly build machine learning models in just a few clicks and generate notebooks that can be extended and scheduled to run while facilitating data science team collaborating
- Data scientists increase their productivity and easily create, explore, and evaluate multiple models that automate machine learning best practices while leveraging OML in-database algorithms and machine learning functionality
- Application developers accelerate machine learning model deployment via in-database scoring, modifying generated OML notebooks, and using REST endpoints with OML Services for application integration.

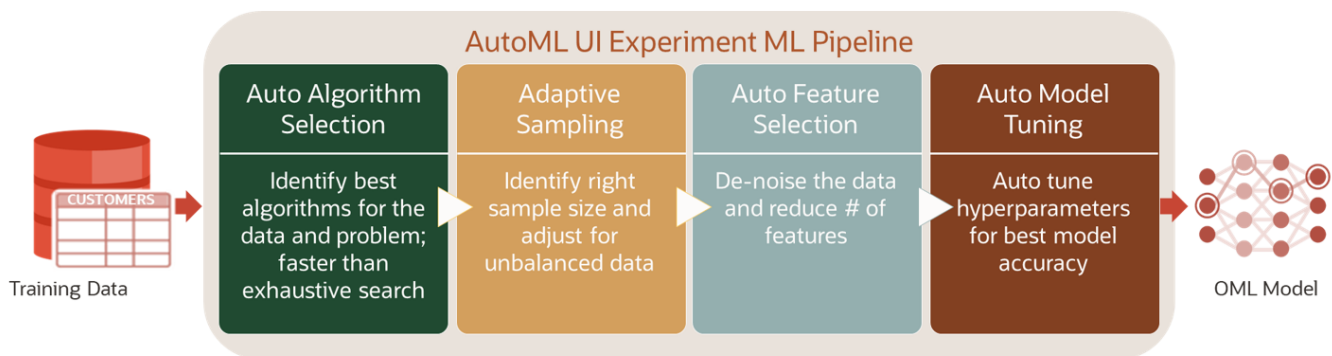
Artificial intelligence, machine learning, autonomous systems are so fundamentally different from what came before, it marks a new generation in computer technology.

Larry Ellison

Chairman and Chief Technology Officer
Oracle Corporation

Automated Machine Learning Pipeline

OML AutoML UI uses the concept of an *experiment* to create a machine learning pipeline that automates several time consuming and repetitive tasks taken by data scientists: algorithm and feature selection, data sampling, model building and evaluation, and hyperparameter tuning.



OML AutoML User Interface automates four major time consuming and tedious steps in the machine learning modeling process: algorithm selection, data sampling, feature selection and model tuning to deliver significant productivity improvements for data scientists and citizen data scientists.

Algorithm Selection

In each experiment, AutoML identifies the most promising algorithms for building models for the specified data and target. Using models built from a wide range of datasets, automated algorithm selection uses *meta-learning* – where based on the distribution of values or *meta-features* in the data, a pre-built model predicts which algorithms are most likely to produce the best results. The algorithms with the highest scores are later used for model tuning. This helps data scientists and non-expert users to find the best algorithm candidates faster than with exhaustive search. This can also reduce compute costs.

Adaptive Sampling

OML AutoML UI uses adaptive sampling to determine the right sampling percentage such that increasing sample size does not further improve model quality. This also speeds up model building. Further, adaptive sampling detects unbalanced datasets that can cause poor models to be built and employs stratified sampling as necessary to create balanced datasets for building better models.

Feature Selection

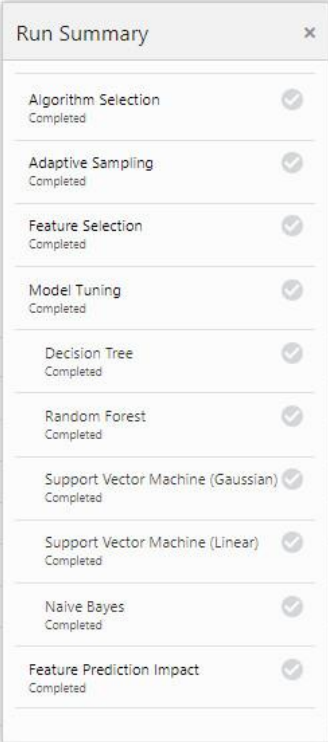
Attributes that have no correlation with the target attribute, have too many constants or missing values, or have too high cardinality can reduce model quality, while increasing model building and data scoring time. OML AutoML UI pre-processes the input data and automatically removes those attributes that contain little information, or worse, noise. OML AutoML UI first ranks the features and evaluates subsets based on these rankings, using several techniques:

- Correlation score computed by Pearson's correlation coefficient (SQL CORR-related functions)
- Attribute importance from a trained Random Forest model
- Coefficients from a trained General Linear model
- Attribute importance from the Minimum Description Length model

To support faster response times, automated features selection uses meta-learning internally to prune the number of possible rankings and feature subsets to be evaluated. This reduces model building and hyperparameter tuning time. OML AutoML UI employs this multifaceted approach for ranking and selecting candidate model input features. Users can manually override to add or remove input features.

Model Tuning

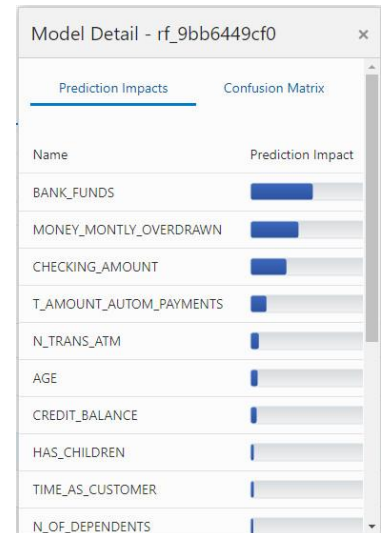
After feature selection, OML AutoML UI builds and tunes multiple machine learning models – taking advantage of OML's in-database processing. OML AutoML UI performs model tuning for each selected algorithm to find hyperparameter settings that significantly improve the model, but without resorting to exhaustive search. For example, in a decision tree, AutoML explores and optimizes the maximum tree depth and minimum split percentage for the algorithm. In a neural network, AutoML explores a range of hidden layers and neurons per layer. OML AutoML UI offers the user the option to select *Better Accuracy* which may take longer or *Faster Results* which will strive to build a good model more quickly. When building models, OML leverages in-database *Automatic Data Preparation* (ADP) for dealing with missing values, data binning, data normalization, and data scaling.



Run Summary	
Algorithm Selection Completed	✓
Adaptive Sampling Completed	✓
Feature Selection Completed	✓
Model Tuning Completed	✓
Decision Tree Completed	✓
Random Forest Completed	✓
Support Vector Machine (Gaussian) Completed	✓
Support Vector Machine (Linear) Completed	✓
Naive Bayes Completed	✓
Feature Prediction Impact Completed	✓

Feature Prediction Impact

For each model built, OML AutoML UI uses a model-agnostic global explanation method that provides insights into the model's behavior. It displays the top attributes and their relative influence on the target attribute. After a model has been built, Feature Prediction Impact shuffles the current values of each attribute assigning them to different records and estimates and ranks feature importance based on the impact each feature when shuffled has on the trained machine learning model's predictions.



Supported Machine Learning Functions, Algorithms and Model Metrics

OML AutoML UI supports the following machine learning functions, in-database algorithms, and score metrics:

Mining Function	Example Use Cases	Algorithms	Model Metrics
Classification (binary and multi-class) Assigns items in a collection to target categories or classes	Identify loan applicants as low, medium, or high credit risks Predict likely buyers Predict voluntary employee attrition	Naïve Bayes Generalized Linear Models Generalized Linear Models (Ridge Regression) Support Vector Machines (Linear) Support Vector Machine (Gaussian) Decision Tree Random Forest Neural Networks	Accuracy Balanced Accuracy F1 Precision Recall ROC AUC Recall Micro Recall Macro Recall Weighted Precision Micro Precision Macro Precision Weighted Confusion Matrix Prediction Impacts
Regression Predicts numeric values along a continuum	Predict the value of a house Estimate lifetime value (LTV) of a customer Predict the	Generalized Linear Models Generalized Linear Models (Ridge Regression) Support Vector Machines (Linear) Support Vector Machine (Gaussian) Neural Network	R-squared Negative Mean Squared Error Negative Mean Absolute Error Negative Median Absolute Error Prediction Impacts

Note: OML AutoML UI does not support unstructured text data, nested data resulting from aggregations of transactional data, or partitioned models.

Accelerate Model Deployment

Once OML AutoML UI has built the machine learning models, data scientists and application developers can use multiple strategies to deploy OML models.

Generate notebooks for use in Oracle Machine Learning Notebooks

When OML AutoML UI has finished building models, users can generate a notebook in OML Notebooks from each model produced in the experiment. Such notebooks can be customized, scheduled to run automatically as jobs, and used for data scoring. Data scientists and application developers can modify and extend the generated notebook using OML4Py and OML4SQL interfaces.

Model Repository

OML on Oracle Autonomous Database also supports a Models interface to enable model management and deployment for in-database models.

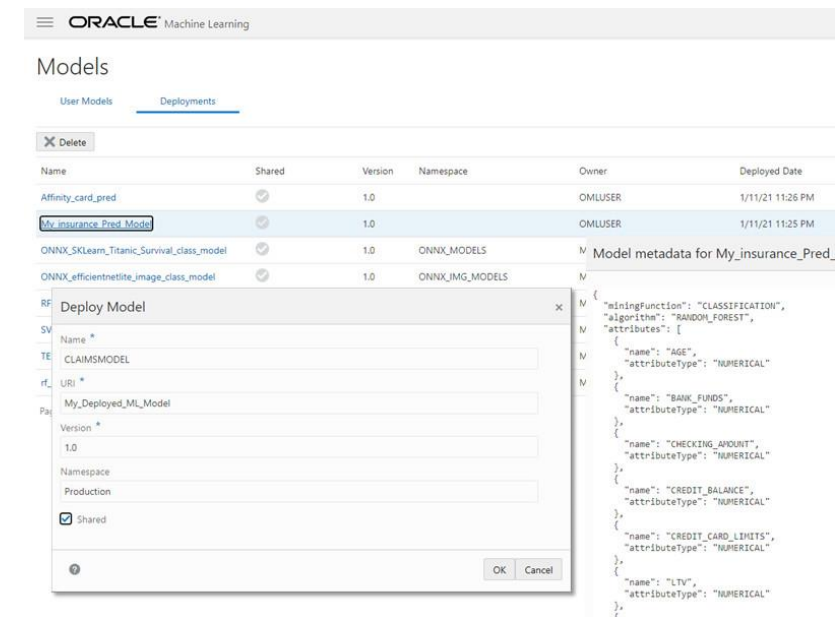
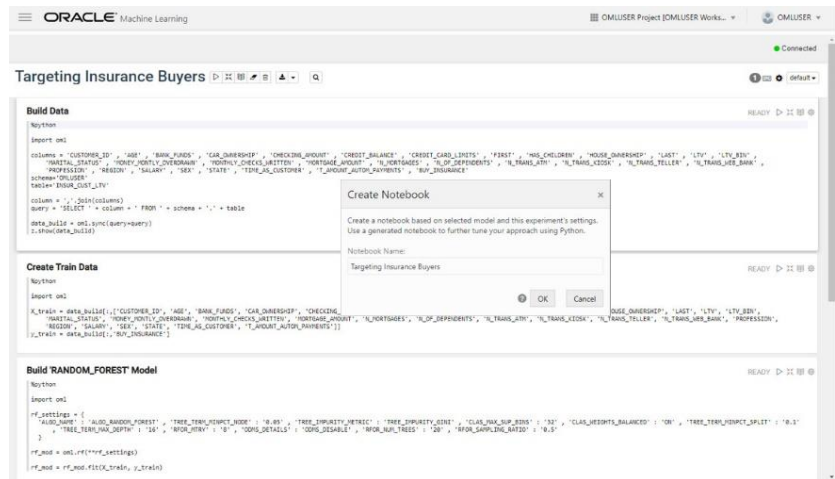
Model Deployment

Users can deploy in-database models in several ways. As stated above, with OML AutoML UI users can generate OML notebooks from experiment-generated models. In addition, in-database models can be used directly in SQL queries to score data from tools such as SQL Developer, SQL Developer Web, Oracle Applications Express (APEX), and Oracle Analytics Cloud. Use models in the same database where the model was built or in a different Oracle Database, whether on premises or cloud, for productionizing OML models using model export and import capabilities.

Additionally, in a few clicks, OML AutoML UI users can deploy models to Oracle Machine Learning Services, which supports model management and deployment with REST endpoints for OML models.

Summary

OML AutoML UI provides an easy-to-use interface that automates repetitive, time-consuming tasks typically taken by expert data scientists, while at the same time simplifying machine learning for non-expert users. OML AutoML UI accelerates the machine learning process from model building to model deployment for all users. OML AutoML UI delivers significant new advances in machine learning simplicity, power, and functionality for enterprises to transform data into actionable insights quickly.



For Further Reading

[OML AutoML UI User's Guide](#)

[OML User's Guide](#)

[OML4Py User's Guide](#)

[OML4SQL User's Guide](#)

[Oracle Machine Learning](#)

Connect with us

Call **+1.800.ORACLE1** or visit **oracle.com**. Outside North America, find your local office at: **oracle.com/contact**.

 blogs.oracle.com

 facebook.com/oracle

 twitter.com/oracle

Copyright © 2021, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document, and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

This device has not been authorized as required by the rules of the Federal Communications Commission. This device is not, and may not be, offered for sale or lease, or sold or leased, until authorization is obtained.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0120

Disclaimer: If you are unsure whether your data sheet needs a disclaimer, read the revenue recognition policy. If you have further questions about your content and the disclaimer requirements, e-mail REVREC_US@oracle.com.