Oracle Machine Learning Technical Brief

Move the Algorithms, Not the Data

April 2023, Version 1.0

Public

Purpose statement

This document provides an overview of features and enhancements included in Oracle Machine Learning. It is intended solely to help you assess the business benefits of leveraging the Oracle data management and analytics platform with Oracle Machine Learning to plan your data-driven data science, machine learning, and information technology projects.

Intended Audience

This technical brief is intended for executives, LOB managers, and practitioners who are looking to provide scalable machine learning capabilities to the data scientists, data analysts, data engineers, and application and dashboard developers across the enterprise.

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.

Table of contents

Introduction

Analyze data where it resides at scale—with minimal data movement. Simplify data science and machines learning solution architecture and deployment. Oracle's converged database places machine learning and other essential analytics technologies at your fingertips in a single, secure, integrated platform, while expanding data access and sharing for your team. As a key part of Oracle's converged database strategy, Oracle Machine Learning helps you quickly develop and deploy data-driven solutions.

As a strategic part of the Oracle AI platform, Oracle Machine Learning empowers data scientists and data analysts to extract knowledge, discover new insights, and make data-driven predictions—working directly with large data volumes in Oracle Database and Oracle Autonomous Database. Oracle Machine Learning empowers application developers by making it easy to deploy and use machine learning models and solutions supporting applications and dashboards. In this technical brief, we introduce Oracle Machine Learning and how it enables enterprise data science teams to achieve greater value from data.

You have choice and flexibility in how you interact with Oracle Machine Learning (OML) using popular languages such as SQL, Python, R, and REST, or using interfaces such as Oracle Machine Learning Notebooks, the Oracle Data Miner user interface, and Oracle Machine Learning AutoML User Interface. In addition, OML Services supports model management and deployment, as well as data and model monitoring with Autonomous Database from a REST interface.

On Oracle Autonomous Database, **OML Notebooks** helps you develop and run SQL, PL/SQL, Python, and R scripts, and use Markdown language for in-notebook documentation. OML Notebooks extends Oracle Autonomous Database as a platform for machine learning and data science, supporting data scientists, data engineers, data analysts, application developers, and IT professionals, including DBAs. OML Notebooks facilitates collaboration through real-time concurrent editing sessions, integrated graphics, easy sharing of notebooks and templates, as well as permission-based access, versioning, and job scheduling, while maintaining the security and enabling permissions inherited by the data from Oracle Database.

With **Oracle Machine Learning AutoML User Interface** on Oracle Autonomous Database, you can build in-database ML models with minimal input—the data and the target to be predicted—and the tool does the rest to produce models guided by machine learning *meta-models*. OML AutoML UI increases productivity for data scientists while making machine learning more accessible for non-experts. OML AutoML UI also helps you deploy in-database models to OML Services and generate editable notebooks for selected models using the OML4Py Python API. The companion OML Models interface helps you manage models and deployments.

Oracle Machine Learning Services supports model management, deployment, and monitoring while benefiting from system-provided infrastructure and an integrated database architecture. The model management and deployment services enable users to deploy in-database machine learning models from both on-premises Oracle Database and Autonomous Database for classification, regression, clustering, and feature extraction machine learning techniques. OML Services supports both in-database machine learning models as well as models exported in ONNX format. Such ONNX-format models may be produced separately from third-party packages such as Tensorflow, PyTorch, Scikit-learn, among others. In addition, OML Services supports cognitive text analytics, including extracting topics and keywords, sentiment analysis, and text summary and similarity. OML Services is optimized for scoring in support of streaming and real-time applications. And, unlike other solutions that require provisioning a VM for 24 - 7 availability, OML Services is included with Oracle Autonomous Database, so users pay only for the additional compute when producing actual predictions.

Oracle Machine Learning for SQL (OML4SQL) offers powerful in-database machine learning algorithms for model building using PL/SQL and scoring data using SQL queries to make predictions. These scalable algorithms are parallelized and scalable for multi-node clusters with memory optimizations. These same in-database algorithms are also exposed through Python and R interfaces, OML Notebooks, and Oracle Data Miner. This extends the database as an enterprise-wide analytical platform for data-driven problems such as churn prediction, customer segmentation, fraud and anomaly detection, cross-sell and up-sell opportunity identification, market basket analysis, and predictive maintenance, and many others.

Oracle Machine Learning for Python (OML4Py) enables data scientists and Python developers to take advantage of the Python environment on data managed by Oracle Database and Oracle Autonomous Database. Python provides a suite of software packages for data manipulation, data exploration, graphics, statistical functions, and machine learning algorithms. OML4Py extends Python's capabilities through four primary areas: (i) explore and prepare database data using overloaded familiar Python syntax and functions that transparently generate SQL, (ii) Python interfaces to in-database machine learning algorithms using a natural Python interface, (iii) ease of deployment of user-defined Python functions with embedded Python execution through REST and SQL APIs, and (iv) automated machine learning (AutoML). For transparent access, OML4Py overloads select Python functions, such as those on Pandas DataFrame objects, and produces SQL for in-database processing—transparently. With AutoML, OML4Py supports automatic algorithm selection, feature selection, and model tuning to enhance your productivity as well as model accuracy and performance.

Oracle Machine Learning for R (OML4R) makes the open source R statistical programming language and environment more scalable through integration with Oracle Database and Oracle Autonomous Database. R provides a rich ecosystem of software packages for data manipulation, graphics, statistical functions, and machine learning algorithms. OML4R extends R's capabilities through three primary areas: (i) explore and prepare database data using familiar R syntax and functions that transparently generate SQL, (ii) R interfaces to in-database machine learning algorithms using a natural R interface, and (iii) ease of deployment of user-defined R functions with embedded R execution through a SQL API.

Oracle Data Miner —an Oracle SQL Developer extension—automates many of the steps in the machine learning process using a drag-and-drop workflow user interface. Oracle SQL Developer is a widely used, free, integrated development environment that simplifies database development and management of Oracle Database and Oracle Autonomous Database. With Oracle Data Miner, you can share analytical workflows and generate SQL scripts to accelerate solution deployment. Oracle Data Miner also provides a PL/SQL API for workflow scheduling and automation.

To summarize, by taking advantage Oracle's converged database strategy and combining the functionality of multiple special-purpose databases—such as machine learning, in-memory, blockchain, JSON, text, graph, spatial, partitioning, sharding—you reduce complexity and management for an end-to-end optimized experience. Oracle Machine Learning benefits from this strategy with minimized data movement, cross-technology usage, reduced total cost of ownership, and enabling the fastest way to deliver enterprise-wide data-driven solutions.

Executive Summary

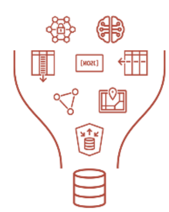
Oracle's mission is "to help people see data in new ways, discover insights, and unlock endless possibilities." As such, Oracle provides enterprises with a broad range of capabilities within the Oracle Machine Learning product family. This enables data scientists and citizen data scientists, data and business analysts, data engineers and IT, application and dashboard developers, and executives to maximize value from Oracle technology and discover insights and make predictions.

Large and small businesses alike know the value of applying machine learning technology to solve key business problems and achieve competitive, if not leadership positions. Oracle enables democratizing machine learning across the enterprise and enhances the ability of businesses to leverage powerful machine learning tools.

Move the algorithms, not the data

With Oracle Machine Learning, Oracle "moves the algorithms to the data" by delivering machine learning algorithms that process the data within your Oracle database instance, where the data resides, whether on premises or in the Cloud. This approach minimizes or eliminates data movement, leverages database scalability, preserves data security, and accelerates time from data acquisition to model deployment. Oracle delivers parallelized and scalable in-database implementations of machine learning algorithms and integrates with the open-source R and Python environments - delivering performance, scalability, security, and automation required by enterprise data science projects. In-database algorithms specifically leverage the parallelism, scalability, and security features of the industry-leading Oracle Database.

Oracle Database - a converged database



You shouldn't have to provision and manage multiple databases and tools, which add complexity and cost, to work with different types of data and analytical functionality. Instead, all such functionality should exist in a single converged, multi-model database, bringing together a broad set of analytical capabilities.

> *In the world of databases, the Oracle Converged Database can support all types of data and processing. Relational Data, Document Data, Spatial Data, Text Data and Graph Data can all be efficiently stored and processed in the same database. OLTP, Real-Time Analytics, Machine Learning, IoT, and Blockchain workloads can all efficiently run in the same database. There is no longer any need to deploy, manage, and synchronize many specialized databases to use these technologies.*
>
> *Like Smartphones, the Oracle Converged Database not only supports these data types and workloads, it does it BETTER than the specialized databases. Further, the synergy of having all these capabilities together in one database enables new possibilities like real-time Machine Learning based fraud detection directly in the OLTP system. Also, using declarative open SQL rather than the bespoke languages needed by specialized databases makes it dramatically simpler for developers to implement these technologies.*
>
> –Juan Loaiza, Oracle

SQL, R, and Python

Polls[1] show that in 2019, Python, R, and SQL were among the most popular languages used for machine learning projects. Different business problems and their underlying data require different analytical techniques and algorithms to be successful. Similarly, different teams and team members have a variety of skills often focused on a particular language. Oracle Machine Learning supports all three popular data science and machine learning languages so you

[1] https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html

can leverage the language that best suits your team and the problem at hand. Oracle Machine Learning supports a natural evolution from Oracle data professionals to data scientists within the same platform. In-database machine learning models are first class database objects with many of the same data management features such as permissions, backup, recovery, and auditing. OML enables enterprises to deploy machine learning models and results for use in dashboards and applications built with tools such as Oracle Analytics Cloud and Oracle APEX, as well as custom server and mobile applications, and leverage convenient SQL or REST interfaces.

By supporting R and Python, data scientists and other R and Python users access Oracle Database as a high-performance compute engine for scalable data exploration and preparation with overloaded R and Python functions and syntax that transparently generates SQL. We expose the in-database algorithms through well-integrated R and Python interfaces – the same algorithms provided through the SQL API. You can deploy user-defined R and Python functions to the database script repository and invoke those functions from SQL and, on Oracle Autonomous Database, REST APIs. User-defined R and Python functions can include third-party packages on Oracle Database and Oracle Autonomous Database.

Automation for increased productivity and model quality

A significant part of the machine learning process involves iteration – performing feature engineering and selection, trying multiple algorithms, and tuning algorithm hyperparameters to find the “best” model, and comparing and selecting the models for a given business problem.

To increase data scientists’ productivity and enable non-expert users, Oracle Machine Learning introduces *automated machine learning*, or AutoML, in both OML4Py and the no-code OML AutoML UI. AutoML can reduce the overall compute time required to derive a high-quality model. AutoML also makes powerful in-database algorithms more accessible to *citizen data scientists* as they do not need to know algorithm hyperparameter details, algorithm-specific data transformations, nor write code to compare and select models according to various score metrics.

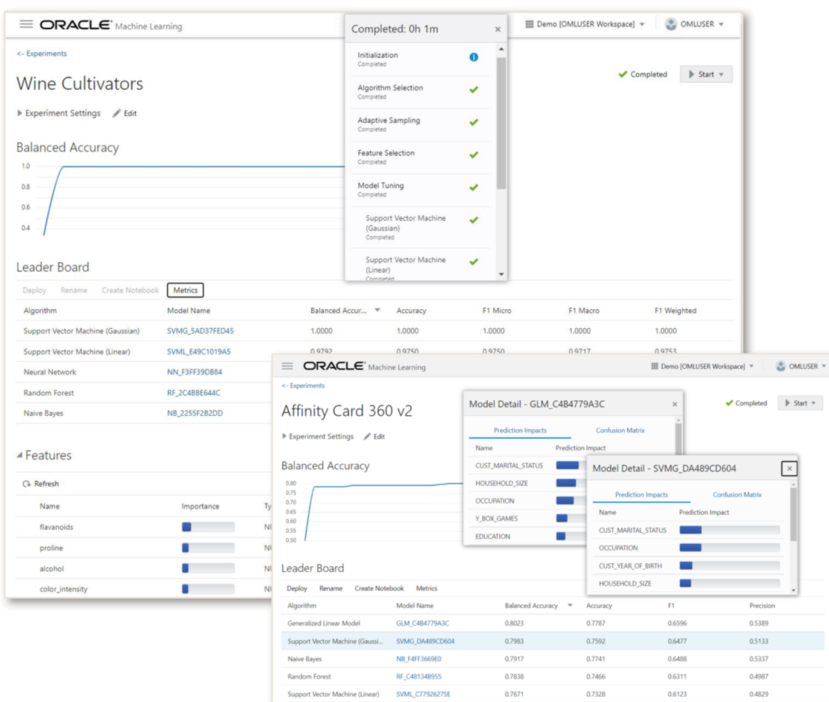


Figure 1: Oracle Machine Learning AutoML UI aids data scientist productivity while enabling non-experts to develop in-database machine learning models for immediate use via SQL queries or deploy them with a few “clicks” to OML Services.

OML4Py AutoML supports the explicit construction of the modeling pipeline using three main capabilities: *algorithm selection*, *feature selection*, and *model tuning*. Oracle Machine Learning AutoML UI builds on OML4Py's AutoML functionality to provide a no-code, easy-to-use interface that requires minimal specification, just the data table and target column to predict. You benefit from built-in intelligence to produce and deploy better models more quickly and easily through a fully assembled AutoML experiment pipeline.

Oracle Machine Learning in-database algorithms also support automation at the algorithm level: (i) automatic data preparation for algorithm-specific transformations of numeric and categorical data, (ii) integrated text mining for token and theme extraction, and (iii) automated partitioned models that produce a convenient-to-use ensemble model.

Oracle Data Miner facilitates data assembly and transformation, and model building and evaluation using multiple algorithms in an easy-to-use drag-and-drop graphical workflow representation.

Production deployment - the critical step

Some might refer to production deployment as the Achilles' heel of data science projects. Back in 2019, VentureBeat.com noted that 87% of data science projects never make it to production. Further, Gartner noted that "Through 2022, only 20% of analytic insights will deliver business outcomes." When you try to integrate machine learning models or open-source R and Python scripts in production, you're faced with the realities of provisioning environments, configuration containers, and addressing the tedious and error-prone tasks associated with ensuring adequate backup, recovery, security, and scalability. Having the algorithms and models in the database virtually eliminates these tasks. With Oracle Machine Learning, deployment is immediate – in-database ML models exist in the database and can be used from SQL queries. You can store your user-defined R and Python functions in the database and run them in database-environment spawned and controlled R and Python engines, respectively. As such, Oracle simplifies production deployment by providing the "plumbing" so enterprise teams can focus on machine learning solutions.

OML Services provides an alternative to in-database model deployment by supporting application developers with REST endpoints on Oracle Autonomous Database. OML Services supports model deployment and model management via a REST API for both in-database OML models and models produced by third-party machine learning algorithms. OML Services manages and deploys classification, regression, clustering, and feature extraction models that were exported from Oracle Database, Oracle Database Cloud Service, and Oracle Autonomous Database. These model management and deployment services extend to third-party models exported in Open Neural Networks Exchange (ONNX) format. You may have produced such third-party models separately, for example, in Oracle Cloud Infrastructure Data Science, from packages such as Scikit-learn, TensorFlow, PyTorch, among others.

OML Services also supports data and model monitoring, where you can be alerted to issues in both data quality and native in-database model quality. In data-driven enterprises, you need to know if there are significant changes in your data over time, as well as whether machine learning models built from your data are performing as expected. Monitoring in OML Services expands support for the machine learning lifecycle and MLOps. Data monitoring helps maintain organizational data quality standards in support of enterprise application and dashboard integrity. For example, upstream data collection may have introduced a new sensor the reports data in a different unit, or a bug in an application that collects customer data may have swapped data in similar fields resulting in erroneous reports. Model monitoring helps ensure that models continue to meet solution success criteria or indicate that rebuilding or re-evaluating a model is warranted. For example, for an enterprise to maintain profitability, a model to predict likely campaign responders must be at least 75% accurate. If accuracy drops below this level, LOB managers' quarterly targets will not be met and a problem that may have been caught and corrected early instead shows up on the CFO's dashboard.

OML Services also supports cognitive text analytics for extracting topics and keywords, and performing text summary and similarity in English, Spanish, French and Italian, in addition to supporting sentiment analysis in English.

Enterprise success

As enterprises amass greater data volumes and supplement corporate data with external data sources, you need to be able to integrate and prepare data at scale for machine learning and streamline the time from business problem definition to solution deployment to meet business objectives. Oracle Machine Learning customers have achieved impressive results, including:

- UK National Health Service (NHS), the system of public healthcare providers in the United Kingdom, were able to assemble and analyze billions of data points on prescriptions, medicines, medical exemptions, doctor relationships, and call center services from across the organization and use this to reveal potential new efficiencies and effective treatments to provide better outcomes. NHS achieved very fast ROI – 581 million British pounds or $717 million USD in the first 2 years – and most recently over 1.5 billion British pounds or nearly $2 billion USD within 5 years and counting. Nina Monckton, head of Information Services, NHS Business Services Authority, states "We chose Oracle because the solution could cope with very large data volumes running into billions of rows and could scale as volumes increase. In addition, the Oracle solution required no IT team support to run the queries, which enables our team of data analysts to be self-sufficient. Oracle's engineered systems accelerated deployment and reduced risk. Working with Oracle has been a very positive experience. The team has been incredibly responsive and provided a number of experts to help us get up and running as quickly as possible. With one vendor providing the whole solution, it's very easy for us. If we need help, we know where to go."
- CaixaBank, a major financial services company in Spain, integrates data from bank branches, ATMs, and internet and mobile banking to gain a complete understanding of customers and offer personalized banking solutions. CaixaBank gains customer loyalty and business competitiveness with improved messaging to reach customers more effectively, better informing them of new bank services and products to boost sales.
- Certegy, a leader in check payment services of more than $2B for over 4,000 clients and 23 of top 50 US retailers, chose Oracle Autonomous Data Warehouse to provide the repository for 850M records for statistical modeling, risk assessments and fraud detection by data scientists. Certegy uses Oracle Analytics, Oracle Machine Learning, and Oracle APEX for improved integration, data quality, and performance than Microsoft PowerBI and Python. Using Oracle Machine Learning, Certegy achieves more accurate, real-time risk scores with an expected 10% reduction in fraud and improved customer service.
- StubHub, the world's largest ticket marketplace, uses Oracle Machine Learning in-database models and integrated R capabilities to run real-time fraud detection models in their database. With this solution, StubHub reduced online fraud by 90%. This not only saves money, but also significantly improves the customer experience. Of course, fraudsters notice when they are shut down and so change their way of operating. This requires reacting quickly, or fraud goes up. With Oracle Machine Learning, StubHub can react very quickly; updating their predictive model happens with the application and database still running. If administrators notice a problem in the morning, they are able to fix it in an hour, deploy the new model immediately, with no downtime.

For more details about these and other customer stories, see OML Customers.

What is Machine Learning?

Machine learning uses algorithms and statistical models to *automatically* process potentially large volume data to find hidden patterns, discover new insights, and make predictions for data-driven problems including:

- Predicting customer behaviors, identifying cross-selling and up-selling opportunities
- Anticipating customer churn, employee, and student attrition
- Detecting anomalies and combating potential tax, medical or expense fraud
- Understanding hidden customer segments and understanding customer sentiment
- Recommending when equipment should be brought in for maintenance
- Identifying key factors that drive outcomes
- Delivering improved customer satisfaction and product quality

Machine Learning, also referred to as *predictive analytics* or *data mining*, has been delivering measurable value for decades. Today, machine learning solutions are even more pervasive—being implemented and deployed across enterprises globally. As analytics technologies and user adoption matures and expands, machine learning use cases and "intelligent" applications emerge daily.

Move the Algorithms, Not the Data

Data is big; algorithms are small. Hence, it makes sense to move the algorithms to the data rather than move the data to the algorithms. Oracle realized this challenge of big data and the algorithms required to make sense of it when it acquired machine learning technology and the development team from Thinking Machines Corporation in 1999. Oracle capitalized on core Oracle Database strengths – specifically aggregation, memory management, parallelism, scalability, and the database architectural framework to build models and score data with all computations inside the database kernel. Neither the data, the machine learning models, nor the predictions and insights need to leave the database.

Today, Oracle Machine Learning supports over 30 in-database algorithms, integrated text mining, automated data preparation, partitioned models, and automated machine learning (AutoML), among other innovative features.

What is Oracle Machine Learning?

Oracle Machine Learning provides support for data driven problems by offering a wide range of powerful data exploration and preparation capabilities, as well as machine learning algorithms implemented inside Oracle Database and Oracle Autonomous Database. We expose the in-database algorithms directly through SQL, as well as Python and R APIs.

Oracle Machine Learning algorithms leverage the underlying SQL features and can even analyze data in its original star schema representation including standard structured tables and views, transactional data and aggregations, and unstructured data as found in character large object (CLOB) data types—using Oracle Text to extract *tokens or themes*. Through Oracle's converged database, you have a single environment for combining machine learning with both spatial and graph analytics.

Oracle Machine Learning in-database algorithms take advantage of database parallelism for both model building and scoring, honor security and privilege schemes, and adhere to audit tracking database features. On Exadata (on premises, Public Cloud, or Cloud at Customer) and Autonomous Database, scoring using OML in-database models occurs at the storage tier using Oracle's *smart scan* technology. One of the challenges with some other machine learning platforms is the need for data to fit in memory. With OML, data is brought into memory incrementally as needed. Further, models are cached and can be shared across queries when used for scoring. OML leverages disk-aware structures – relying on the database memory manager for efficient allocation in multi-user environments. When building or scoring partitioned models, not all component models need to be loaded at once.

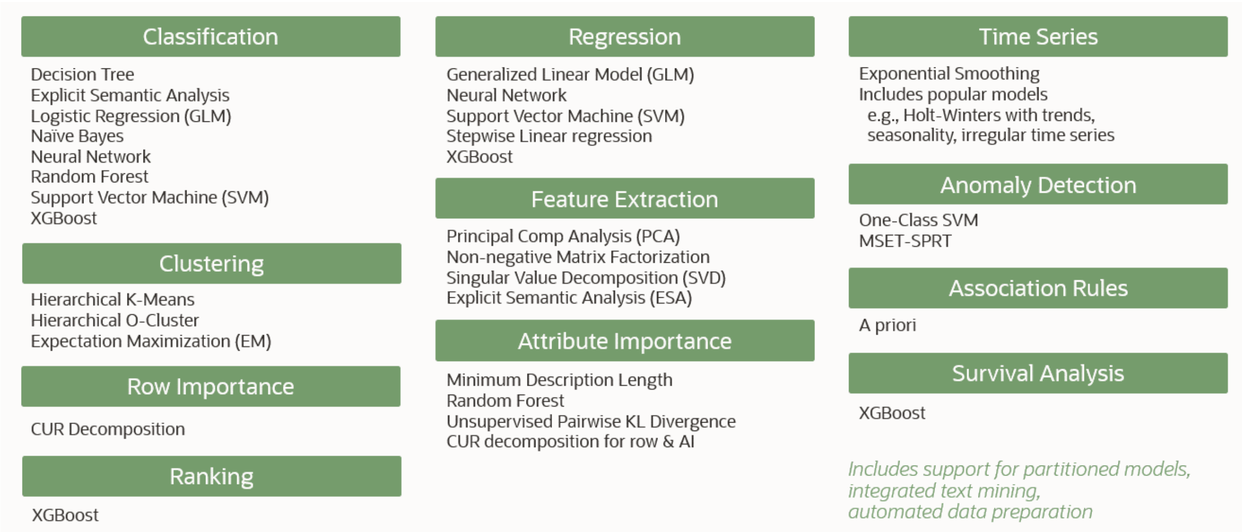


Figure 2: Oracle Machine Learning algorithms and analytics functionality support a wide range of critical business use cases.

Oracle Machine Learning supports a mixture of data analytics and machine learning methodologies. For example, you may want to combine transactional data, demographic data, customer service data, and customer comments to assemble a 360-degree customer view. You may decide to perform clustering on your customers to pre-assign them to customer segments, and then, for each segment, build separate classification, regression, or anomaly detection models for better accuracy.

Support for SQL and open-source R and Python

SQL has been the standard language for data management for over 40 years. For data analytics and machine learning, however, various languages compete—R, Python, Scala, Java, SAS, MATLAB, and others. These have been long time favorites, but in recent years, open-source Python and R have surged to the top of the pack – being taught in most data science educational programs. Both blogs and data science community polls cite Python, R, and SQL as the top languages and Oracle Machine Learning supports these three with OML4Py, OML4R, and OML4SQL.

In-Database Processing with Oracle Machine Learning

Oracle Machine Learning extends Oracle Database and Oracle Autonomous Database as a platform for data science and machine learning with powerful in-database analytics, where algorithms operate directly on database data—in the database. Oracle Machine Learning in-database models are first-class database schema objects, built by invoking a PL/SQL procedure or corresponding functions in OML4R or OML4Py. You can use these models to score data and explore model details (patterns and insights revealed by the model) via OML4SQL, OML4R, or OML4Py functions and database views.

When building models, Oracle Machine Learning uses standard scalable database technology, including parallel execution, bitmap column indexes, storage-level partitioning, query optimization, and aggregation techniques. Oracle Machine Learning also takes advantage of custom-built technologies, including parallel infrastructure, IEEE float, Intel Math Kernel Library (MKL), support for unstructured text data, and automatic data preparation for binning, one-hot encoding, and missing value handling.

The power of in-database algorithms as SQL functions is even more evident when scoring machine learning models, whether in batch or in *online transactional processing* (OLTP) environments. You get high-performance scoring – millions of records in seconds. Using the *smart scan* technology of Oracle Exadata Storage Cells, SQL predicates and OML models get pushed down to the storage layer for processing. In both cases, only records that satisfy the predicates are pulled from disk for further processing inside the database.

For example, here's a simple SQL query to find the US customers likely to churn:

```
SELECT cust_id
FROM customers
WHERE region='US' AND prediction_probability(churn_model,'Y' using *) > 0.8;
```

Reading this query, it equates to US customers with more than 80% probability to churn based on the model called *churn_model*. OML pushes the scoring computations to the storage tier via predicate functions using Oracle's proprietary Exadata Smart Scan technology for 2-5 times faster scoring performance than non-Smart Scan in-database scoring.

Whether building models or scoring data, in-database machine learning helps you to streamline data science projects from days and weeks to seconds, minutes, or hours.

Automation drives productivity

The data science process can be highly iterative. Popular methodologies, such as the *CRoss Industry Standard Process for Data Mining* (CRISP-DM), have iterative loops built in, especially for data preparation, and model building and evaluation. Fortunately, a significant part of this process can be automated. Such automation not only improves data scientist productivity by addressing some of the more time-intensive, iterative aspects of machine learning, but also helps non-experts to leverage machine learning even without understanding the finer algorithm details such as hyperparameters involved in model tuning. Still, such automation does not eliminate the need for broader machine learning skills, especially in the areas of problem definition, data preparation, and solution evaluation.

Oracle Machine Learning provides automation such as automatic data preparation (ADP), automated text processing, partitioned models, as well as a fully automated ML pipeline through *automated machine learning* (AutoML). AutoML provides automated algorithm and feature selection, and model selection and tuning. These features all help reduce the effort and expertise required to go from business problem to business solution using machine learning.

Integrating open-source R and Python with Oracle Database

Oracle Machine Learning for R (OML4R) and Oracle Machine Learning for Python (OML4Py) make the R and Python languages and environments more scalable through integration with Oracle Database and Oracle Autonomous Database. The strengths of these languages include drawing on their rich ecosystems with an extensive array of powerful and extensible graphics, statistical, and analytics packages. If you use R and Python, however, you likely face challenges including memory constraints, single-threaded processing, and difficulty deploying or translating R and Python model logic into enterprise applications or dashboards.

OML4Py integrates Python with Oracle Database and overloads select Python functions to equivalent SQL and in-database algorithms — enabling you to operate on database-resident data without using SQL. The overloaded functions form the *transparency layer*, which transparently converts requested Python functionality into equivalent SQL for in-database processing. Through the transparency layer, Python programmers can create pandas DataFrame proxy objects that reference database tables and views, enabling them to access, analyze, and manipulate data that resides in the database. The database automatically optimizes the corresponding SQL code to improve query efficiency, and takes advantage of column indexes, table partitioning, and database parallelism.

With embedded Python execution, you can run user-defined Python functions in one or more system-spawned and controlled Python engines. You can develop and refine your user-defined functions - even using custom installed packages - and then deploy them in the database script repository. You can invoke your user-defined Python functions in a non-parallel, data-parallel, or task-parallel manner using Python and SQL, and, on Autonomous Database, also REST APIs.

OML4Py is pre-installed on Autonomous Database and can be accessed through Oracle Machine Learning Notebooks. Using the standalone client that supports connecting to Oracle Autonomous Database and Oracle Database instances, you can also use third-party tools like PyCharm and Jupyter interfaces.

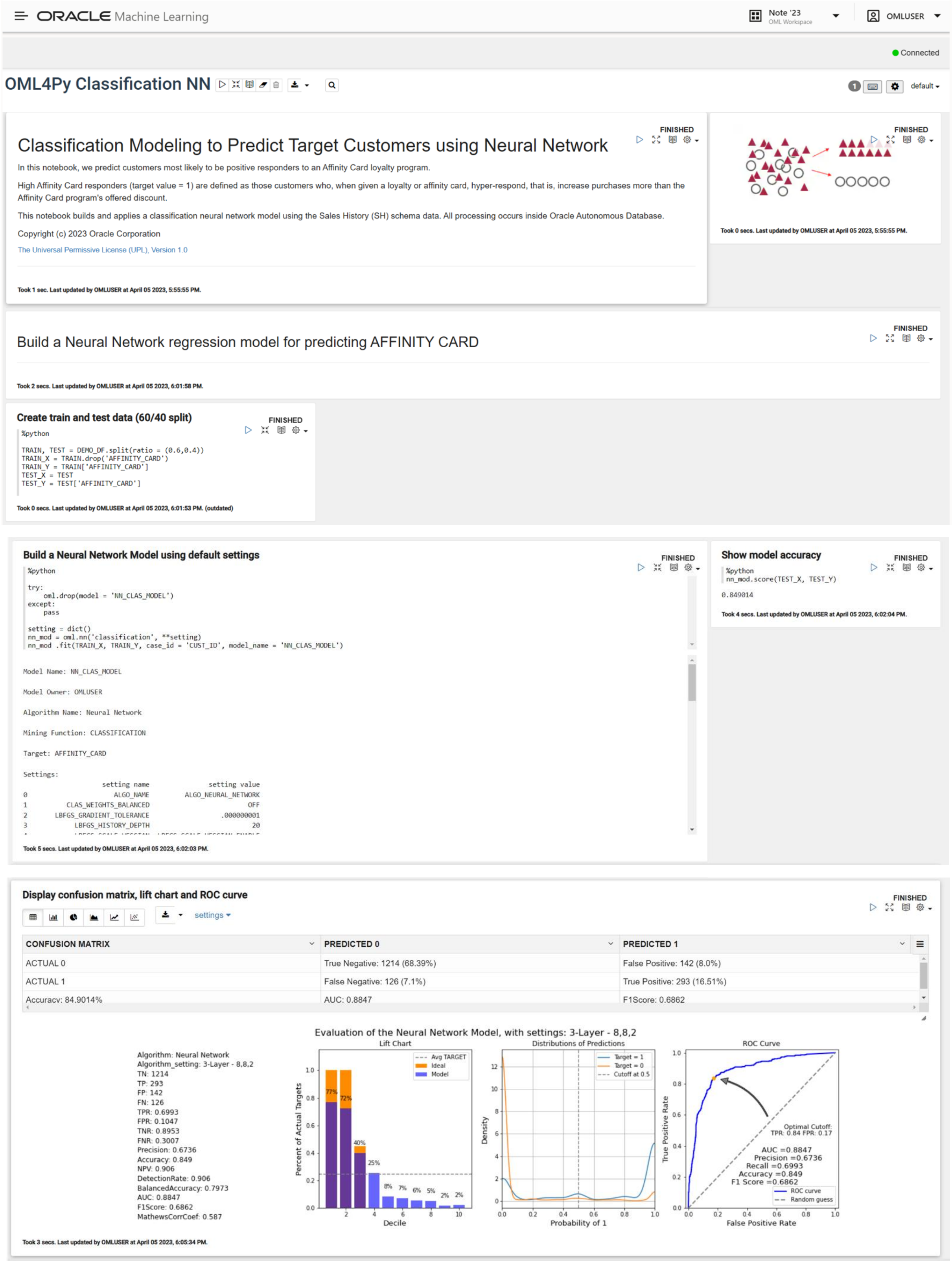


Figure 3: Oracle Machine Learning Notebooks supports Python, R, SQL, conda, and markdown interpreters. In this notebook, the Python developer leverages the Neural Network algorithm for classification, evaluating the model with visualizations leveraging the matplotlib library.

OML4R integrates R with Oracle Database, and like OML4Py, you take advantage of the R transparency layer for scalable in-database data processing and build models and score data using the R interface to in-database algorithms. With OML4R *embedded R execution* on Oracle Database, you can extend native functionality by creating user-defined functions that use third-party R packages that you or your administrator install at the database server machine with Oracle Database. These user-defined R functions can be invoked using both R and SQL APIs. You can also use your favorite R IDE, such as RStudio, to connect to an Oracle Database instance and use OML4R.

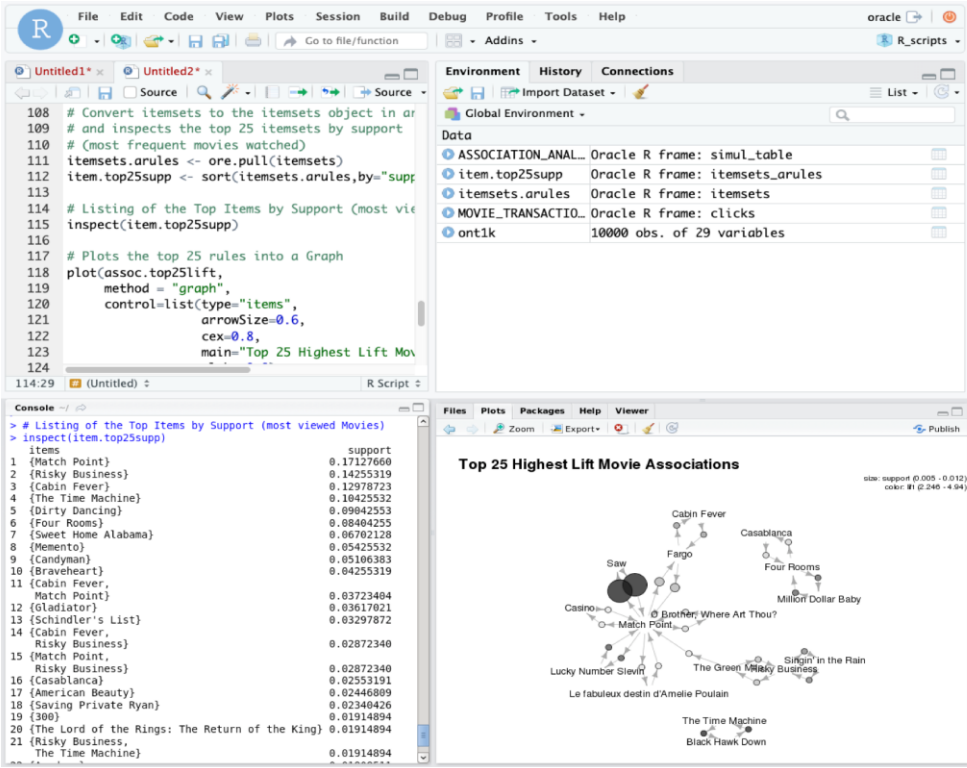


Figure 4: Oracle Machine Learning for R uses in-database algorithms (here, Apriori Association Rules) with local graphing capabilities to visualize the results, all within RStudio Server.

On Oracle Autonomous Database, you access OML4R through Oracle Machine Learning Notebooks. You define R functions, which can include native R functionality and that from third-party packages installed in conda environments and use Embedded R execution to invoke your functions from R, SQL, and REST APIs.

Oracle Machine Learning AutoML User Interface

Oracle Machine Learning AutoML User Interface provides a no-code user interface supporting automatic machine learning. OML AutoML UI supports data scientist productivity and provides non-experts access to powerful in-database algorithms for classification and regression.

OML AutoML UI automates model building with minimal input – just specify the data and the target in an “experiment” and the tool does the rest. If you want more control, use advanced settings to limit the set of algorithms tried and the number of models ranked, among other settings.

As shown in Figure 1 earlier in the document, running an experiment produces a leaderboard of models and their performance. You can select models for deployment in OML Services or use models directly through SQL and Python. You can also generate notebooks containing the OML4Py code that produces selected models.

Oracle Machine Learning Services

Oracle Machine Learning Services extends OML functionality to support model deployment and model management as well as data and model monitoring. OML Services uses a REST API for both in-database models and models exported in Open Neural Networks Exchange (ONNX) format built using third-party algorithms. OML Services provides REST endpoints available with Oracle Autonomous Database.

In-database models supporting classification, regression, clustering, and feature extraction can be deployed to OML Services. Third-party models supporting classification, regression, and clustering, including those from packages such as scikit-learn, TensorFlow, and PyTorch can be deployed to OML Services provided they can be exported in ONNX format.

OML Services supports data monitoring with data drift detection, where a flag is raised if the statistical properties of data significantly change over time. OML data monitoring enables comparing and tracking data changes over time. It's based on the idea that a business task relies on well-understood data referred to as the baseline dataset and periodic subsequent datasets that are compared against that baseline. Quickly and reliably identifying changes in the characteristics of the underlying data enables data stewards to take corrective action before such changes have a negative impact on the enterprise.

Model monitoring includes identifying when the user-specified model metric—like *balanced accuracy* or *mean squared error* (MSE)—significantly changes. Model monitoring can also detect when the distribution of predicted values deviates too much from initial values, also called *concept drift*. Such deviations can signify the need to rebuild models as well as consider investigating root causes of these deviations. For example, loan applicants who were considered attractive prospects last year may no longer be considered attractive because of changes in a bank's strategy or outlook on future macroeconomic conditions. Similarly, customers' interest in product categories may have changed over time, leading to supply issues, where retail prediction models were built on data reflecting interests from an earlier period. Like data monitoring, users provide new test datasets periodically to evaluate the model accuracy metric. When the metric falls below the specified threshold, such test data is flagged such that you may choose to rebuild your model or select a different model with a higher metric. You may want to craft scenarios that include building multiple models using, for example, different algorithms or settings. These models then compete based on their accuracy metric. This "champion score" can guide which model should be used by the application at a given point in time.

In addition, OML Services supports proprietary cognitive text capabilities, including topic discovery, keywords extraction, summarization, sentiment analysis, and similarity. Languages supported include English, Spanish, French and Italian (based on a pre-built Wikipedia-based Explicit Semantic Analysis model), with sentiment analysis available for English.

Through the ONNX format model deployment feature, you can also import cognitive image third-party models for image classification using images or tensors.

The set of REST API capabilities in OML Services is shown in Figure 5.

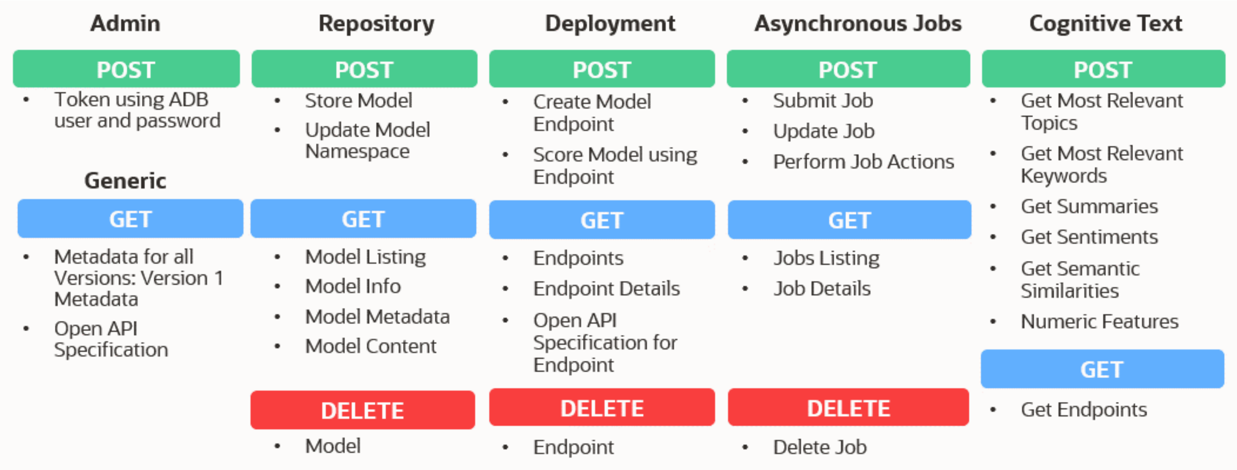


Figure 5: REST API capabilities provided by Oracle Machine Learning Services

Oracle Data Miner—a SQL Developer Extension

Oracle Data Miner, an extension to the desktop application Oracle SQL Developer, supports data scientists and citizen data scientists who may prefer a drag-and-drop user interface to work directly with data inside the database using a workflow paradigm. Oracle Data Miner workflows capture and document your machine learning methodology. You can save and share workflows with others to automate and publish machine learning methodologies. Oracle Data Miner supports a select set of in-database machine learning algorithms.

You can quickly visualize, explore, prepare, and transform data, build and evaluate machine learning models, and use model details and model evaluation viewers. Then, apply OML in-database models to new data or generate SQL and PL/SQL scripts to deploy analytical workflow.

These scripts can be passed to developers for deployment within the same or a different Oracle Database. Application developers can programmatically run workflows using the Oracle Data Miner PL/SQL workflow API or just run the appropriate SQL and PL/SQL scripts to easily integrate predictive methodologies into applications for wider enterprise use.

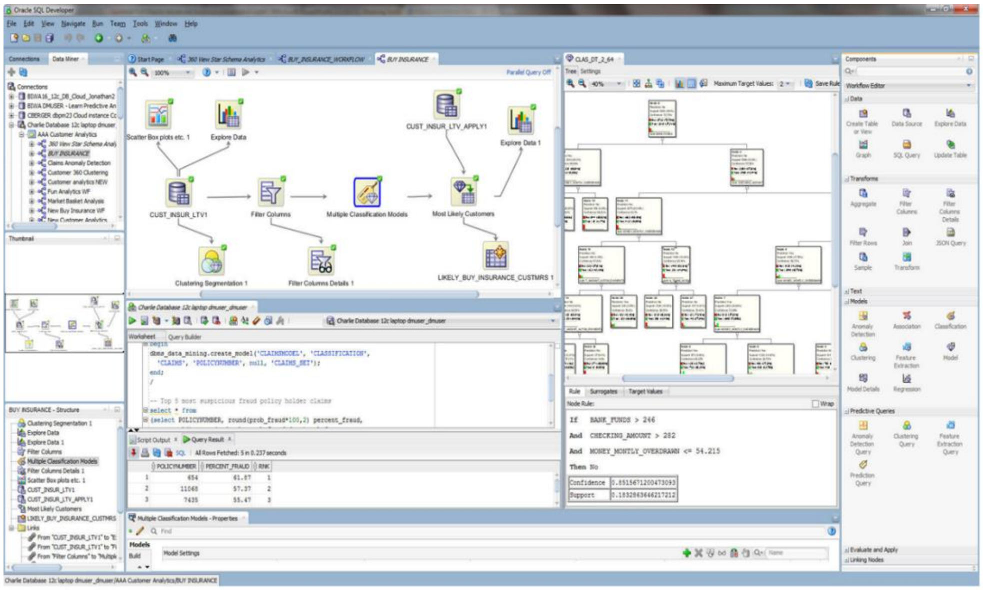


Figure 6: Oracle Data Miner, a SQL Developer extension, provides a drag-and-drop workflow user interface to explore data, build, evaluate and apply predictive models, and deploy analytics methodologies as SQL and PL/SQL scripts.

Oracle Machine Learning in Data Lakes

Big data usually represents structured and unstructured data that's more easily collected than analyzed, and then stored in *data reservoirs, data lakes*, or lakehouses. However, for the purposes of solving data-driven problems, oftentimes the main interest is not in the raw data itself, but on a summary of the data provided by counts, percentages, and "events" to be joined with other curated data inside Oracle Database. The resulting data can often be used to solve a business problem using machine learning. The idea is to "boil down" the data lake into new *engineered features* that capture domain expert knowledge for producing better models.

The data environment outside Oracle Database introduces new data management and data analysis challenges for DBAs. New Autonomous Database capabilities for data lakes addresses this challenge by extending SQL processing to data in Oracle Object Storage, where you can query the data lake using Oracle SQL – which means that any tool or application that can query Autonomous Database can now query the data lake as well. That means that Oracle Machine Learning extensive capabilities can now extend to data from both sources.

Data stored in Data Lakes and Object Storage is often large with sparse representations, such as transactional format. Given that much of the data may come from sensors, Internet of Things, "tweets", and other high-volume sources, data analysts can now more easily take advantage of new big data sources containing data of possibly unknown value, boil them down, and combine them with data of known value managed inside a database or data warehouse. You can aggregate this data at various levels, e.g., counts, maximum/minimum values, thresholds, averages, and sliding SQL window averages.

In this case, Oracle Machine Learning supports filtering "big data" by reducing it directly in the data lake, joining it to other database data, and then performing machine learning in-database.

Oracle Machine Learning—the Fastest Way to Deliver AI Applications and Dashboards

Oracle's strategy of simplifying data analytics makes it easier to develop, refine and deploy machine learning-enhanced, intelligent applications. Data, user access, security and encryption, scalability, applications development environment, and machine learning exist in the database ecosystem. With Oracle Machine Learning, you can easily add predictions and real-time actionable insights to your enterprise applications, dashboards, or tools using SQL, REST, or other APIs to Oracle Database.

Oracle has been developing intelligent applications for over a decade. Oracle provides next-generation predictive applications on premises and in the cloud, including:

- Oracle Human Capital Management Predictive Workforce
- Oracle Content and Experience
- Oracle Configure, Price, Quote
- Oracle Depot Repair
- Oracle Spend Classification
- Oracle Retail Science Platform Cloud Service
- Oracle Retail XBRi Loss Prevention
- Oracle Retail Customer Segmentation
- Oracle Adaptive Intelligence Foundation for Anti-Money Laundering

Figure 7: Oracle HCM Predictive Workforce application delivers pre-built Oracle Machine Learning predictive analytics for employee attrition, employee performance and "What if" analysis.

Through integration with Oracle Analytics Cloud and Oracle Analytics Server, users can seamlessly work with in-database machine learning models to include predictions and insights when analyzing data and producing dashboards.

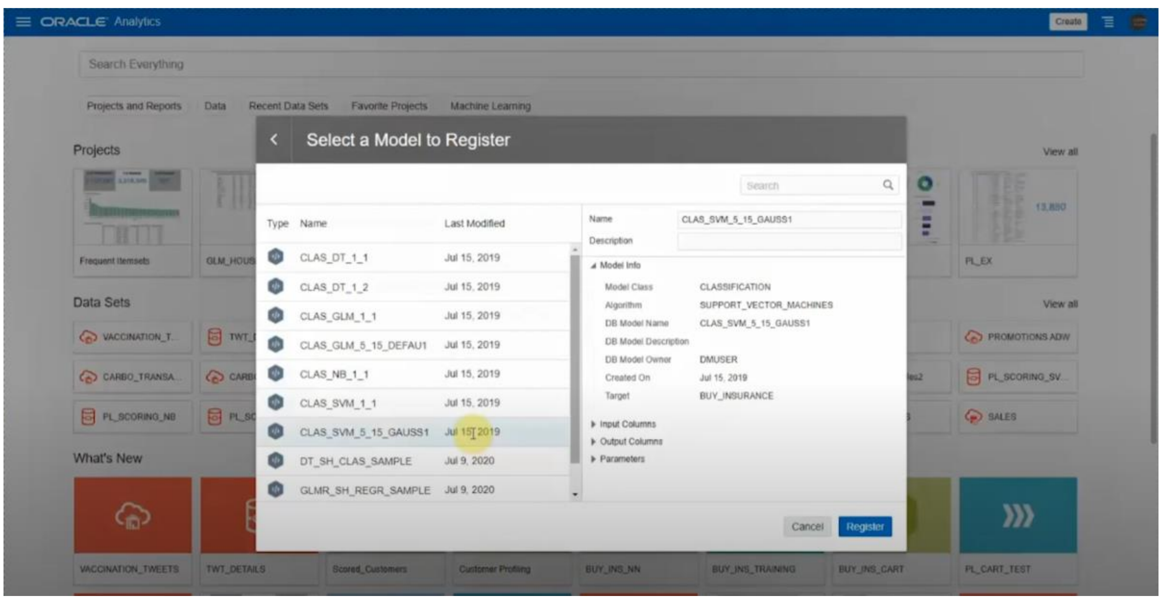


Figure 8: Using Oracle Analytics Cloud to select an in-database model to use in analytics and dashboards.

Conclusion

Oracle Machine Learning delivers scalable, parallelized, machine learning algorithms within Oracle Database and Big Data environments as part of Oracle's converged database strategy. Oracle Machine Learning provides machine learning algorithms supporting classification, regression, clustering, association rules, feature extraction, time series, ranking, row importance, and anomaly detection. You can use machine learning algorithms through multiple interfaces, including OML Notebooks, OML AutoML UI, popular language APIs for SQL, R, and Python, OML Services, and Oracle Data Miner.

Oracle Machine Learning native algorithms take full advantage of Oracle Database scalability, parallelism, security, and structured and unstructured data processing capabilities. This makes Oracle a powerful platform for developing machine learning solutions and embedding them in enterprise applications either on premises or Oracle Cloud. Oracle's multiple decades of leading-edge data management experience combines with the strategy of "moving algorithms, not data." By integrating data management and machine learning into the converged Oracle Database, you benefit from reduced cost of ownership, elimination of data movement, fast solution deployment.

For more information

Website

Machine Learning Blog

Documentation

Office Hours Sessions

Oracle Machine Learning LiveLabs

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