ORACLE DATABASE 11G
DATA MINING OPTION

KEY FEATURES AND ENHANCEMENTS

ORACLE DATA MINING

- In-Database Analytics
- Data Mining models are 1st class database objects
- Full Set of Algorithms
- Detect Anomalies
- Automatic Data Preparation
- Mining Activity Guides
- Data Preparation
- Text Mining
- Java API and PL/SQL API
- Code Generation
- Predictive Analytics
- The Oracle Database Platform

The Data Mining option to the Oracle Database 11g Release 1 enables enterprises to produce actionable predictive information and also to build integrated business intelligence applications. Using data mining functionality embedded in the Oracle11g Database, business analysts can find patterns and insights hidden in their data. Application developers can quickly automate the discovery and distribution of new business intelligence—predictions, patterns and discoveries—throughout the organization.

In-Database Analytics

Oracle Data Mining (ODM) provides comprehensive data mining functionality that is embedded in the Oracle Database. This eliminates the need to extract data from the database to external analytical engines for data analysis. All of Oracle Data Mining's functionality is embedded in the Oracle11*g* Database. With Oracle Data Mining, the data never leaves the database—the data, data preparation, model building, and model scoring activities—all remain in the Oracle Database. This also has significant advantages for security, scalability, manageability, application development, and user access.

Oracle Data Mining's embedded data mining in the database not only means that the data stays in the database but also that the mining tasks and data transformations are performed within the database. They can run automatically, asynchronously, and independently of any user interface.

The Oracle11*g* Database's scalability allows Oracle Data Mining to analyze large volumes of data to detect subtle patterns and relationships and extract new business intelligence hidden in the data. Oracle Data Mining's new insights and predictions are stored in database tables and are available for access by other Oracle and nonOracle query, analysis, and reporting tools and applications.

Full Set of Mining Algorithms

The Oracle Data Mining provides support for a wide range of data mining model building and evaluating functionality including: classification, regression, clustering, associations, anomaly detection, text mining, attribute importance and feature extraction.

Algorithms supporting solutions to Classification problems include Decision Trees, Naïve Bayes, Generalized Linear Models (GLM), and Support Vector Machines (SVM).

ORACLE®

Clustering can be performed using Enhanced K-means (based on distance metric) or O-cluster (based on density).

Regression problems can be solved using Generalized Linear Models or Support Vector Machines.

Text Mining, and Anomaly Detection use Support Vector Machines, Attribute Importance uses Minimum Description Length (MDL), Associations uses A Priori, and Feature Extraction uses Non-negative Matrix Factorization (NMF).

Anomaly Detection

Normally, classification requires knowledge of all target classes. A version of SVM can build a profile of one class and when applied, flag cases that are somehow different from that profile (that is, "abnormal" or "suspicious"). This allows for the detection of rare cases not necessarily related to each other that are nearly impossible to identify using Classification.

Automatic Data Preparation

Ordinarily, data transformations required by the data mining model would have to be carried out manually, or as part of a Mining Activity. In Oracle 11g Data Mining, all data transformations required by a particular algorithm are completed automatically as part of the execution of the model.

Mining Activities

The Oracle Data Miner graphical user interface (GUI) employs Mining Activities that not only prescribe the correct order of operations and perform all algorithm-required data transformations, but also provide intelligent settings and optimizations for all parameters; however, the expert can expose all parameters in order to override default values.

The model build activity includes the evaluation of the model when appropriate, and several methods of testing, including Receiver Operating Characteristics (ROC) Analysis for Classification and Residual Plot for Regression.
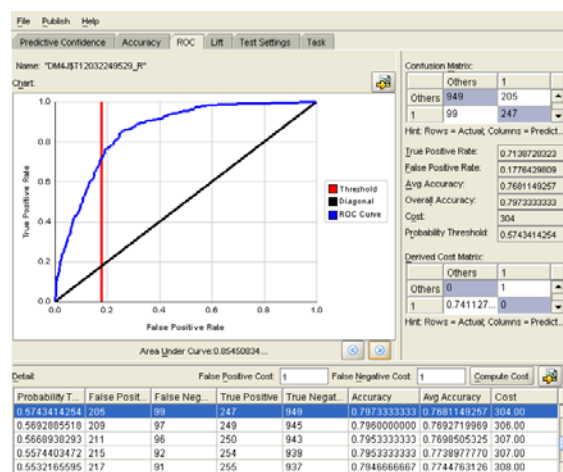

Figure 1. Receiver Operating Characteristics (ROC Analysis)

ORACLE®

The Model Build activity "remembers" all data transformations and parameter settings, so when it comes time to score data with the optimal model, the Build metadata is passed seamlessly to the Model Apply activity for automatic execution.

Data Preparation

Oracle Data Miner can accept as input multiple tables or views and perform the appropriate joins and transformations necessary for modeling. ODM can mine transactional data and nested data tables. Managing the data aggregation, transformations, and data preparation inside the database helps accelerate model deployment and application development.

Text Mining

The Support Vector Machine, Association Rules, K-Means Clustering, and Nonnegative Matrix Factorization algorithms can all accept text (unstructured data) as an input attribute, so that a column containing, for example, a physician's notes, technical paper, or a police report can be treated like any other input variable to enhance the value of the predictive model.
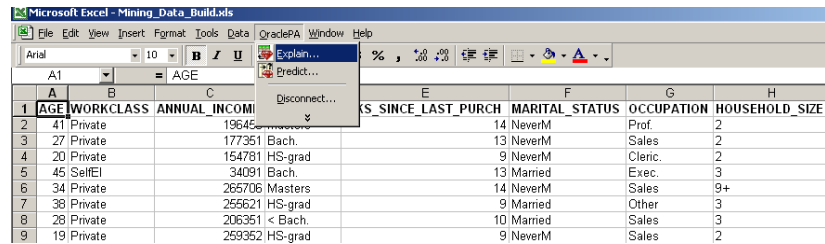
Java or PL/SQL API

Application developers can use Oracle Data Mining's Java and/or PL/SQL Application Programming Interface (API) to integrate data mining's insights and predictions into business applications. Developers can build a data mining model using one API, for example PL/SQL, and then apply the model using either API. Sample programs illustrating the coding required for most data mining operations are included with the Oracle Database.

Code Generation

When the Oracle Data Miner GUI executes the operations of an activity, PL/SQL code is generated, allowing the building, testing, or applying of a predictive model to be packaged and executed in a different Oracle Database environment. The code can be accessed through JDeveloper or SQLDeveloper and used in the building of an application. Thus, a model built and optimized on one system can be applied to data as a component of an application on another system.

Predictive Analytics

The SQL functions PREDICT, EXPLAIN, and PROFILE are completely self-contained packages for building a classification model or an attribute importance model. All parameters are assigned optimized values and intermediate data is not saved. The results are the predicted scores (PREDICT), the ranked list of attributes (EXPLAIN), or the scores and rules (PROFILE), which can be used as part of an operational pipeline, or displayed on the command line or in a spreadsheet.

ORACLE®

Figure 2. The PREDICT function executed within a spreadsheet

The Oracle Database Platform

With Oracle Data Mining, enterprises benefit from a completely integrated Oracle Data Warehouse, Business Intelligence environment. All Oracle Data Mining functions are integrated with the Oracle Database's industry leading security, scalability, and data management platform.