Oracle9*i* Data Mining

Oracle9i Data Mining enables companies to build integrated business intelligence applications. Using data mining functionality embedded in the Oracle9i Database, application developers can automate the extraction and distribution of business intelligence for integration into other business applications.

Data Mining

ORACLE9*i* DATA MINING

Oracle9*i* Data Mining provides comprehensive data mining functionality that is embedded in the Oracle9*i* Database. Oracle9*i* Data Mining allows companies to build advanced business intelligence applications that mine corporate databases, discover new insights, and integrate that information into business applications.

Oracle9*i* Data Mining is a priced option to the Oracle9*i* Database Enterprise Edition (EE) that embeds classification, prediction, associations, and clustering natively within the database. All model building and scoring functions are accessible through a Java API.

Oracle9*i* Data Mining enables application developers and data analysts to extend the benefits from data mining to many users throughout an organization. By delivering applications driven by data mining, customers, independent software vendors (ISVs), and Oracle Applications can deliver far greater value than is possible from a data mining tool for a single data analyst.

Oracle9*i* Data Mining helps companies build advanced business intelligence applications that find patterns in data that can help you discover new insights and make predictions. With Oracle9*i* Data Mining, you can implement strategies to:

- prevent customer attrition
- cross-sell to existing customers
- acquire new customers and identify profitable customers
- detect fraud
- identify customer clusters or segments
- profile customers with greater accuracy

Oracle9*i* Data Mining helps companies tap information hidden in corporate databases to reveal new customer insights. At every stage of the customer life cycle, Oracle9*i* Data Mining delivers value that goes straight to your bottom line.

Oracle9*i* Data Mining can also detect hidden patterns in life science, scientific, government, manufacturing, and other applications, such as:

- finding "target" genes and promising "leads" to attack diseases
- predicting the quality of a manufactured part
- finding associations between patients, drugs, and outcomes
- identifying patterns of anomalies

Insights discovered by Oracle9*i* Data Mining can be revealing, significant, and invaluable.

DATA MINING EMBEDDED IN ORACLE9*i* DATABASE

Oracle9*i* Data Mining automates the process of extracting business intelligence from large amounts of data. It eliminates off-loading data to external special-purpose analytic servers for data mining and scoring. All of Oracle9*i* Data Mining functionality is embedded in the Oracle9*i* Database. With Oracle9*i* Data Mining, the data never leaves the database—the data, data preparation, model building, and model scoring activities all remain in the database. This also has significant advantages for data security, manageability, back-up, and user access.

Embedding in the database not only means that the data stays in the database but also that the mining tasks can run automatically, asynchronously, and independently of any GUI-driven user interface. This plus automated capabilities provides a powerful and scalable platform to quickly and efficiently build advanced business intelligence applications.

Oracle9*i*'s scalability allows Oracle9*i* Data Mining to analyze large volumes of data to detect subtle

patterns and relationships and extract new discovered business intelligence. Oracle9*i* Data Mining's new insights and predictions are available for access by other query, analysis, and reporting tools and applications. This allows businesses to build applications that are driven by data mining results.

Because the Oracle9*i* Database delivers unrivaled performance and scalability, Oracle9*i* Data Mining provides the ideal infrastructure for building advanced business intelligence applications. Companies can score large data tables without extracting the data to external dedicated data mining servers.

By automating the discovery of new business intelligence, Oracle9*i* Data Mining significantly reduces information latency time from *data* to *information* and provides results that translate directly into higher profits and lower costs.
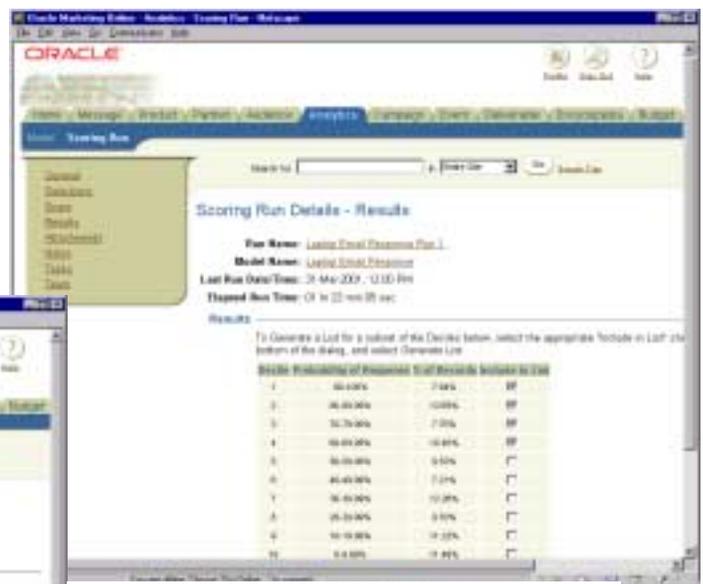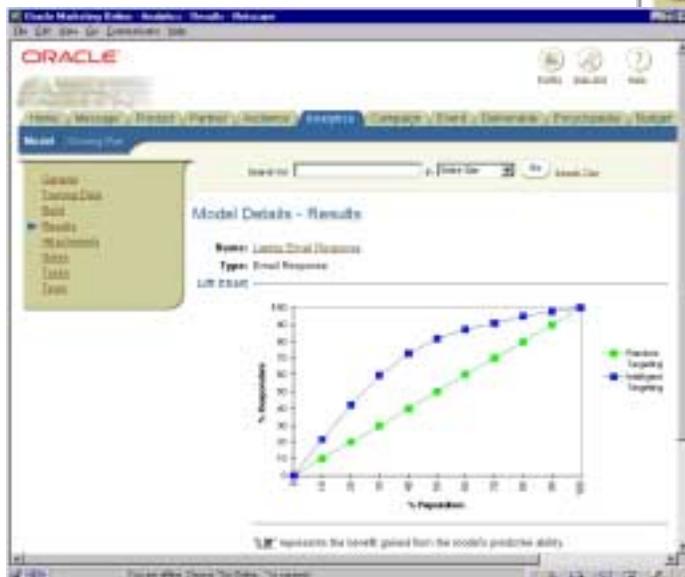
Data mining insights can be integrated into other applications, such as this Oracle CRM/Oracle Marketing Online campaign management application.

AN ANALYTICAL INFERENCE ENGINE

Oracle9*i* Data Mining enables companies to systematize the discovery and integration of new business intelligence within their operations. Application developers can use Oracle9*i* Data Mining's Java API to add data mining insights and predictions to enhance business applications such as Customer Relationship Management (CRM), Call Centers, Enterprise Resource Planning (ERP), Web portals, and even wireless applications. Companies, for example, can use Oracle9*i* Data Mining to build churn applications that identify customers that are likely to churn before they leave for a competitor. Oracle9*i* Data Mining's predictions can help anticipate and proactively manage customer behavior in mutually beneficial 1-to-1 relationships.

Retailers and database marketers can use Oracle9*i* Data Mining to build marketing campaign applications that target those prospects that are most likely to respond to offers. Oracle9*i* Data Mining can integrate data mining results into these applications. Examples include predicting a customer's likelihood to churn, to respond to a special offer, to be a profitable customer, to file a claim, or to spend large amounts of money. E-businesses and Web sites can enhance Web searches using Oracle9*i* Data Mining to present other documents or items that are related or "associated" in use or content.

Once the data has been mined and the predictive models built, Oracle9*i* Data Mining can apply the models to score other data to make predictions. Scoring of data occurs in the database and the scores are available for use by other applications. Data mining models can provide insights and predictions on demand to interactive applications, such as call centers, that suggest "recommendations." For example, a call center application could use a customer's historical data together with responses from a call in progress to rate the customer's preferences and make personalized cross-sell recommendations.

PREDICTION AND CLASSIFICATION

Oracle9*i* Data Mining provides Naive Bayes and Adaptive Bayes Networks algorithms for making predictions and classifications. These algorithms are applicable to a variety of data mining problems and provide high accuracy. By finding patterns in data, companies can make predictions about the future behavior of customers with similar characteristics — using the past as a predictor of the future. Typical prediction applications estimate the probability of an outcome, such as "0, 1" or "yes, no" or "A, B, C, or D." Consider the following example:

Question: Will Customer #4567332 respond to my special offer?

Answer: "Yes," with a likelihood of 92%.

Customer ID	Response	Probability
4567332	YES	0.92
5463433	YES	0.92
1123444	YES	0.85
9565333	YES	0.85
2346557	YES	0.72
2257673	YES	0.55

Oracle9i Data Mining's predictive models return predicted outcomes and their associated probability, so companies can proactively manage their business.

Oracle9*i* Data Mining's Adaptive Bayes Networks also provide human readable "rules" that are useful for explanations and understanding the model, such as:

IF RELATIONSHIP = "Husband" AND
    EDUCATION_NUM = "13-16"
    THEN CHURN= "TRUE"

Results of models can be combined to provide valuable business intelligence. For example, Oracle9*i* Data Mining could build a model to predict the life-time value (LTV) of a customer and another model to predict the likelihood that a customer will churn. Multiplying the two expected results (P(LTV) x P(Churn)) can provide insights on how to spend your marketing budget.
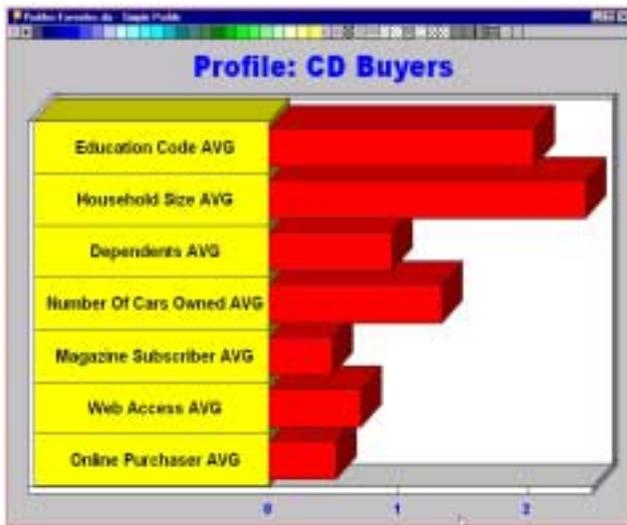
Oracle9i Data Mining's predictions and classifications can be examined using other software and applications, such as Oracle Discoverer shown here.

CLUSTERING AND ASSOCIATION RULES

Oracle9*i* Data Mining provides Clustering and Association Rules algorithms to detect naturally occurring clusters and associated or co-occurring items hidden in the database.

Cluster analysis is popular for discovering groupings within the data that may reveal some additional insight. Clustering can also be used to

assign "cluster member values" that can be used as input to supervised learning techniques. An example output of clustering might reveal, for example:

- the average AGE of Cluster 1 members is 20% higher than Cluster 2 members

Association Rules analysis is often used to find popular product bundles (e.g., market basket analysis) of products that are related for customers, such as "milk" and "cereal" being associated with "bananas." Association Rules can also be used to identify co-occurring items or events such as:

- What manufactured parts and equipment settings are associated with failure events?

- What patient and drug attributes are associated with which outcomes?

- Which items or products is a person who has purchased item A most likely to buy?

Oracle Discoverer showing the results of Oracle9i Data Mining's Association Rules analysis.

Associations Rules can be used to predict the next item placed into the shopping basket, which can

be helpful to satisfy customers and increase average order value.

MODEL SEEKER

Model Seeker allows a user to asynchronously build multiple classification models, then evaluates them and selects a best model. The models to be built and evaluated can be a combination of Naive Bayes and Adaptive Bayes Network models.

ATTRIBUTE IMPORTANCE

Often, corporations are swamped with data and feel they need to store, manage, and mine *all* the data. Oracle9*i* Data Mining's attribute importance feature allows users to rank fields on their relative importance or influence on a specified target field. For example, Attribute Importance may be used to reduce the 650 fields or attributes about customer loyalty to 50 attributes that most affect it.

JAVA API

Application developers access Oracle9*i* Data Mining's functionality through a Java API. Programmatic control of all data mining functions enables automation of data preparation, model building, and model scoring operations.

Java Data Mining (JDM) is an emerging data mining standard, following Sun's Java Community Process as a Java Specification Request (JSR). JDM has participation from Oracle, Sun, IBM, and many other companies that recognize the need for a Java-based standard for specifying and using data mining. JDM leverages several evolving data mining standards, including Object Management Group's Common Warehouse Metadata (CWM), the Data Mining Group's Predictive Mining Markup Language (PMML), and

International Standards Organization's SQL/MM for Data Mining.

Oracle9*i* Data Mining's API provides an early look at concepts and approaches being proposed for JDM. Ultimately, Oracle9*i* Data Mining will comply with the standard after it is published.

TECHNICAL SPECIFICATIONS

DATA PREPARATION

- Oracle9*i* Data Mining' provides data transformations that are accessed using a Java API. Additionally, SQL and PL/SQL can be used for a wide range of data preparation tasks.

- Oracle9*i* Data Mining supports both aggregated single-record case format or multi-record case (transactional) format data.

- Automatic data binning (discretization) function for creating Oracle9*i* Data Mining bin boundary tables.

SUPERVISED LEARNING

- All model building and scoring functions occur within the Oracle9*i* Database.

- Default mining function settings for all Oracle9*i* data mining operations. Ability to override and specify settings.

- Prediction of binary (e.g., yes or no) or multi-class outcomes with supporting confidence (e.g., assuming A, B, C, or D as possible outcomes, B will occur with 60% confidence.

- Support for prior probabilities and cost matrix.

- Real-time scoring performance for individual records.

Attribute Importance

- Identification of the attributes as important for predicting a target field. Statistics give an idea of how correlated each predictor is with the target attribute.

- Attribute Importance is implemented using the Predictor Variance algorithm. Predictor Variance estimates the variances of the predictor target combinations and the variance with respect to the other predictors.

- Automatic attribute importance preprocessing is included in the Adaptive Bayes Network algorithm

Naive Bayes Algorithm

- Naive Bayes algorithm makes predictions using Bayes' Theorem, which derives the probability of a prediction from the underlying evidence, as observed in the data.

- Naive Bayes affords fast model building and scoring.

- Naive Bayes cross-validation permits the user to test model accuracy on the same data that was used to build the model, rather than building the model on one portion of the data and testing it on a different portion. This is important when the number of cases available to build a model is small.

Adaptive Bayes Networks Algorithm

- Adaptive Bayes Networks (ABN) build models based on counts observed in the database. ABN models support both prediction and classification.

- ABN provides human readable rules, e.g.:

  IF RELATIONSHIP = "Husband" AND EDUCATION_NUM = "13-16" THEN CHURN= "TRUE"

- Advanced ABN features:

  - Maximum Network Feature Depth: Network features are like individual decision trees. This parameter restricts the depth of any

individual network features in the model.

- Maximum Number of Network Features: Controls the maximum number of features included in this ABN model.

- Maximum Build Time: Allows the user to build quick, possibly less accurate models for immediate use or simply to get a sense of how long it will take to build a model with a given set of data.

Model Seeker

- Oracle9*i* Data Mining provides a Model Seeker productivity feature to automatically run multiple Naïve Bayes and Adaptive Bayes Networks model and recommend the *best* model

Model Evaluation

- Confusion matrix for Naive Bayes and Adaptive Bayes Networks models

- Lift calculation for Naive Bayes and Adaptive Bayes Networks models

UNSUPERVISED LEARNING

Association Rules Algorithm

- Find the occurrence and likelihood of co-occurring events—for example, Q, R, and S are associated with Z, 555 times, with 78% confidence.

- Find all combinations of items, called frequent itemsets, whose support is greater than a specified minimum support.

- Finds all rules with support greater than a

minimum support and confidence greater than a minimum confidence.

Clustering Algorithms

- Discover naturally occurring groupings within the data to reveal additional insights.

- *k*-means Clustering: The *k*-means algorithm is a distance-based clustering algorithm that partitions the data into a predetermined number of clusters. The *k*-means algorithm works only with numerical attributes. Oracle9*i* Data Mining implements a hierarchical version of the *k*-means algorithm.

- O-Cluster Algorithm: Creates a hierarchical, grid-based clustering model. The resulting clusters define dense areas in the attribute space. The clusters are described by intervals along the attribute axes and the corresponding centroids and histograms. O-Cluster works with both numerical and categorical attributes and supports hierarchy.

- Clustering techniques are useful as a data-preprocessing step to identify homogeneous groups on which to build better predictive models.

- Example results: Records 12, 15, and 25 are members of Cluster_1. Output includes: cluster centroids per attribute and cluster rules.

JAVA API

- All Oracle9*i* Data Mining functions for data preparation, model building and scoring are accessed via a Java API.

- Mining tasks and mining functions paradigm.

- Asynchronous execution of the mining tasks.

- Oracle9*i* Data Mining applications can be built using Oracle Jdeveloper.

PMML SUPPORT

- The Predictive Model Markup Language (PMML) specifies data mining models using an XML DTD (document type definition). PMML provides a standard representation for data mining models to facilitate model interchange among vendors. PMML is specified by the Data Mining Group (http://www.dmg.org).

- Oracle9*i* Data Mining is both a producer and consumer of PMML models. Oracle9*i* Data Mining is a producer and consumer of two model types: Association Rules models and Naive Bayes classification models. Oracle9*i* Data Mining consumes only models that use features supported by Oracle9*i* Data Mining.

PLATFORM REQUIREMENTS

Oracle9*i* Data Mining runs in Oracle9*i* Database on all supported platforms.

Oracle9*i* Partitioning is recommended for large data mining problems.