

Oracle Data Mining 11g

Know More, Do More, Spend Less

An Oracle White Paper
June 2007

NOTE:

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. 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 for Oracle's products remains at the sole discretion of Oracle.

Oracle Data Mining

Know More, Do More, Spend Less

Note.....	2
Oracle Data Mining.....	5
What is Data Mining?	8
Independent Reviews of Oracle Data Mining	9
Oracle data mining: not only good but affordable too!.....	9
Know More.....	11
Use ODM to Discover More Information From Your Data.....	11
ODM's Supported Algorithms	11
The Data Mining Process	13
Supervised Learning Algorithms	14
Naïve Bayes.....	15
Decision Trees	15
Generalized Linear Models (Logistic and Multiple Linear Regression)	16
Support Vector Machines	16
Attribute Importance.....	16
Unsupervised Learning Algorithms	16
Clustering.....	17
Association Rules	17
Feature Selection	17
Anomaly Detection	18
Text Mining	18
Spatial Data Mining.....	18
Predictive Analytics	19
Do More	19
Moving from a “Tool” to an “Enterprise Application”	19
Fewer Moving Parts.....	22
Model Apply and Model Scoring.....	25
Build Models and Applications Simultaneously	25
Oracle CRM 11 <i>i</i>	25
Customer Relationship Management Application That Integrates Data Mining	25
Spend Less.....	26
Eliminate Redundant Data, Servers, Software and Personnel —And Pay Less for Data Mining Software.....	26
Conclusion.....	27

ORACLE DATA MINING

Oracle Data Mining is powerful software embedded in the Oracle Database that enables you to discover new insights hidden in your data. Oracle Data Mining helps businesses to target their best customers, find and prevent fraud, discover the most influential attributes that affect Key Performance Indicators (KPIs), and find valuable new information hidden in the data. Oracle Data Mining helps technical professionals find patterns in their data, identify key attributes, discover new clusters and associations, and uncover valuable insights.

With Oracle Data Mining, everything occurs in the Oracle Database—in a single, secure, scalable, platform for business intelligence. Oracle Data Mining represents a breakthrough in business intelligence. Oracle Data Mining moves the analytical functions into the database—with the data. Traditional alternatives force you to extract the data out of the database to separate, unsecured and costly dedicated statistical, analytical or mining servers.

Oracle Data Mining enables companies to:

KNOW MORE—leverage your data and discover valuable new information and insights that were previously hidden.

DO MORE—build applications that automate the extraction and dissemination of new information and insights.

SPEND LESS—Oracle Data Mining is significantly less expensive compared to traditional approaches and, as a component of your investment in Oracle technology, significantly reduces your total cost of ownership.

Key benefits

- Seamlessly query, summarize, analyze, and mine the same data—with fewer moving parts
- Natural integration with database-driven applications
- Increased Security—data doesn't leave the database, leverage database security options
- State-of-the-art algorithms

- Real-world scalability—parallel model building and scoring, leverage Oracle Real Application Clusters (RAC)
- Lower total cost of ownership—purchase cost, model and result deployment, software development, maintenance
- Leverage 30+ years of experience with ever advancing database technology

Oracle Data Mining enables you to go beyond standard query and reporting tools and Online Analytical Processing (OLAP). Query and reporting and OLAP tools can tell you who are your top customers, what products have sold the most, and where you are incurring the highest costs. Oracle Data Mining helps you go beyond a manual search and query for information approach towards a new methodology where data mining automatically digs through your massive amounts of data to help you predict, understand, and develop new insights.

In today's competitive marketplace, companies must manage their most valuable assets — their data and the valuable information that lies hidden within it. Moreover, they must exploit their data. If they don't, their competitors will beat them using new insights, discoveries, and strategies developed by extracting more information from their data. That's where Oracle Data Mining can help. Data mining can sift through massive amounts of data and find new information — valuable insights that can help you find patterns, make predictions, and discover new, previously hidden, information.

With Oracle Data Mining, you can implement strategies to:

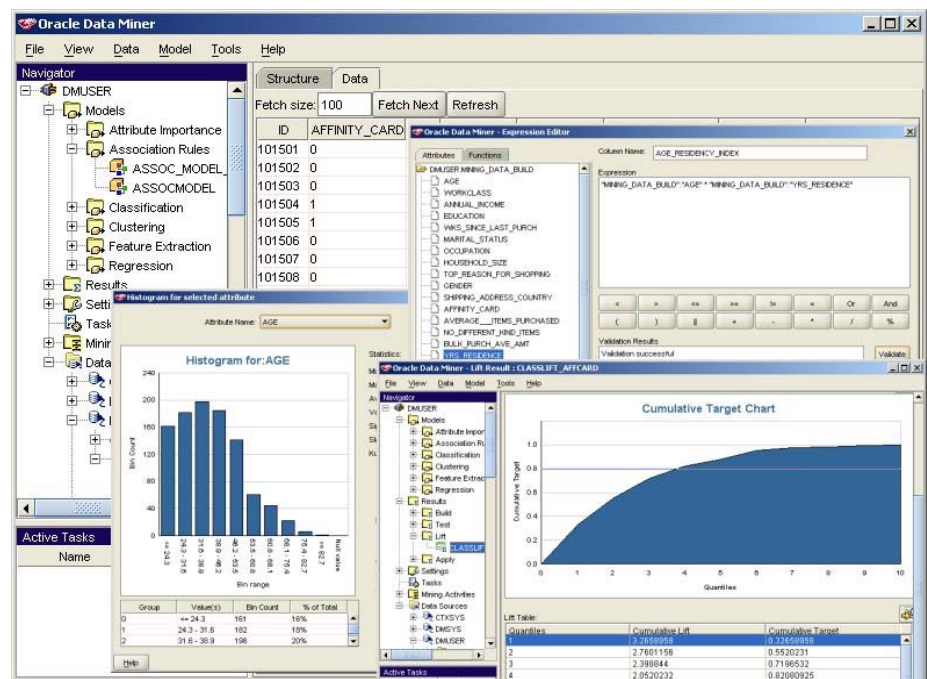
- Develop profiles of targeted, e.g. high value, customers
- Anticipate and prevent customer attrition
- Acquire new customers and identify the most profitable customers
- Identify promising cross-sell opportunities
- Detect noncompliant and fraudulent activities
- Discover new clusters or segments
- Develop customer profiles
- Identify likely targets and promising leads in drug discovery
- Find association relationships of co-occurring items and/or events
- Mine unstructured data, that is, text and spatial data

Traditional business intelligence (BI) tools such as reports, interactive query and reporting only report on what has happened in the past. They report on historical sales figures, quantities, and “current status” values. Online Analytical Processing (OLAP) provides rapid drill-down for fast, more detailed information, roll ups,

forecasting and trend analysis but usually for only for averages, sums, trends, and group-by aggregates. None of these approaches can provide the deeper insights and views to the future like data mining. Data Mining sifts deeper into your data to discover information—patterns, factors, clusters, profiles, and predictions—that remain “hidden” in the data.

Oracle Data Mining (ODM) allows you discover new insights, segments and associations, make more accurate predictions, find the variables that most influence your business, detect anomalies, and in general, extract more information from your data. For example, by analyzing the profiles of your best customers, ODM enables you to build data mining models and integrated applications to identify customers who are likely to become your best customers in the future. These customers may not represent your most valuable customers today, but may match profiles of your current best customers. Moreover, with ODM you can do more and transform a predictive model into a regular production application that distributes lists of your most promising customers to your Sales force every Monday morning. Knowing the “strategic value” of your customers — which are likely to become profitable customers in the future and which are not, or predicting which customers are likely to churn or likely to respond to a marketing offer — and integrating this information into your operations is the key to proactively managing your business.

Oracle Data Miner provides an easy to use graphical user interface (GUI) to Oracle Data Mining's data mining algorithms. Oracle Data Miner enables data analysts to build predictive models and generate the components (Java and PL/SQL code) necessary to build applications. Application developers can then integrate the models into enterprise applications that automate and integrate data mining.



Oracle Data Mining (ODM) is a priced option to the Oracle Database 11g Enterprise Edition (EE). With ODM, data mining and scoring functions reside natively in the Oracle Database—the data and data mining activities never leave the database. ODM embeds classification and regression, associations, and clustering

models, attribute importance, and feature selection algorithms in the Oracle Database. Additionally, most of these algorithms can mine both structured and unstructured (text) data. ODM model building and model scoring functions are accessible through both Java and PL/SQL application programmer interfaces (APIs), the Oracle Data Miner graphical user interface (GUI) and through the Spreadsheet Add-In for Predictive Analytics. The combination of graphical user interfaces and ODM APIs provide an infrastructure for application developers and data analysts to integrate data mining’s predictions and insights seamlessly with other database applications.

WHAT IS DATA MINING?

Now let’s describe what data mining is and how it both differs from and complements other business intelligence tools — query and reporting, Online Analytical Processing (OLAP), and statistical tools. Let’s also look at some common definitions of business intelligence tools.

“Simply put, data mining is used to discover [hidden] patterns and relationships in your data in order to help you make better business decisions.” -- Herb Edelstein, Two Crows Corporation

Query and reporting tools help you to get information out of your database or data warehouse. These tools are good at answering questions such as “Who purchased a mutual fund in the past 3 years?”

OLAP tools go beyond this and allow users to interactively “drill down” into the data for summaries, comparisons, analysis, and forecasts. OLAP is good at drill-downs into the details to find, for example, “What is the average income of mutual fund buyers by year by region?”

Business Intelligence

Oracle Data Mining differs from query, reporting, and OLAP tools by discovering new information that was previously hidden in the data.

Query and Reporting	OLAP	Data Mining
Extraction of detailed and roll up data	Summaries, trends and forecasts	Knowledge discovery of hidden patterns
“Information”	“Analysis”	“Insights & Prediction”
Who purchased mutual funds in the last 3 years?	What is the average income of mutual fund buyers, by region, by year?	Who will buy a mutual fund in the next 6 months and why?

Statistical tools are used to draw conclusions from representative samples taken from larger amounts of data. Statistical tools are useful for finding patterns and

correlations in “small to medium” amounts of data but fall short when the amount of data begins to overwhelm the tool. Typically, when you deal with more than 25 input variables and tens of thousands of records, traditional statistical techniques struggle. Because statistical tools cannot analyze all the data, they force data analysts to use representative samples of the data and to eliminate input variables from the analysis. By throwing out variables and using samples of the data, you are throwing away valuable “information.”

Note: Oracle provides a wide range of basic statistical functions in the Database including descriptive statistics, hypothesis testing, correlations analysis, ANOVA, and tests for distributions. For more information, see OTN’s Statistical Functions web page: http://www.oracle.com/technology/products/bi/stats_fns/index.html

Query and reporting, OLAP, and statistical tools are good at allowing the user to drill down and understand what has happened in the past. With these tools, if you know what you are looking for, have a good analytical tool, are a good analyst, and have a lot of time, you should be able to eventually find the information you seek.

Oracle Data mining goes deep into the data and finds patterns from the data. Oracle Data Mining uses state-of-the-art machine-learning algorithms to automatically sift through each record and attribute to uncover patterns and information that may have been hidden. Data mining goes beyond traditional business intelligence tools and analyzes the details of the past, for example, whether an individual purchased item “A”. Data mining builds models and uses the models as predictors of the details in the future, for example, the likelihood that a customer will purchase “A” in the future. Data mining is good at providing detailed insights and making individual predictions, such as “Who is likely to buy a mutual fund in the next six months and why?”

Independent Reviews of Oracle Data Mining

In November 2006, Bloor Research published a review of Oracle Data Mining that contained the following excerpts.

Oracle data mining: not only good but affordable too!

(Published: 30th November, 2006, Bloor Research 2007)

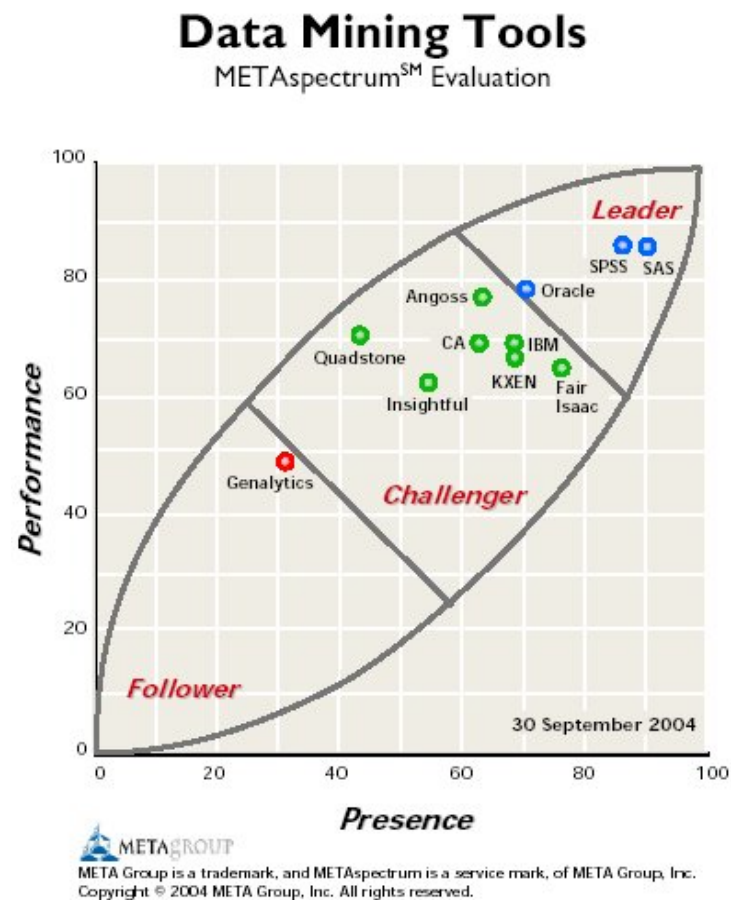
“...the Oracle data mining option is one of the **great bargains** available today because it is affordable and ... is a real **Rolls Royce of capability** and features. ...redesigned from scratch and **put the algorithms into the database** to ensure that, not only is the execution of the algorithms efficient, but the vast amounts of data handling that typifies traditional datamining is minimised. ... Oracle are leaving the database in situ and mining it there, which **saves a lot of effort and will greatly increase productivity**.

This is a **fully featured, highly sophisticated data mining capability** to enable professionals to operate against Oracle data sets with productivity and precision. Oracle data mining has a broad range of available algorithms, which enable it to undertake virtually every kind of business and scientific analysis that one can think of. ...

...Oracle are giving the data mining professional a real alternative to SAS and SPSS with an offering that is equally as well featured, but which promises to outperform any standalone offering.”

Meta Spectrum Evaluation

META Group, a leading provider of IT research, advisory services and strategic consulting, published its METASpectrum report for Data Mining, and ranked Oracle Data Mining a "leader," citing the maturity and stability of Oracle Data Mining and its large market share relative to the competition. META Group measured 15 categories in its evaluation. Among other things Oracle was noted for its ability to execute and pricing.



KNOW MORE

Use ODM to Discover More Information From Your Data

Oracle Data Mining allows companies find new information from their data using a wide range of state-of-the-art algorithms. Data mining algorithms are machine-learning techniques for analyzing data to discover patterns and relationships. Different algorithms are good at different types of analysis.

ODM's Supported Algorithms

Oracle Data Mining supports supervised learning techniques (classification, regression, and prediction problems), unsupervised learning techniques (clustering, associations, and feature selection problems), attribute importance techniques (find the key variables), and can mine both structured and unstructured (text) data.

Oracle Data Mining provides the following supervised learning algorithms for targeted business problems:

- Naïve Bayes for classification and prediction
- Decision Trees for classification and prediction and generating descriptive, human readable “rules”
- General Linear Models (Logistic and Multiple Linear Regression), a classical statistical workhorse
- Support Vector Machines, a state-of-the-art algorithm for classification, prediction and regression

To find naturally occurring patterns in the data, Oracle Data Mining provides the following unsupervised learning algorithms:

- Enhanced k-Means Clustering
- Orthogonal Partitioning Clustering
- Association Rules or Market Basket Analysis for finding patterns of co-occurring events
- Nonnegative Matrix Factorization (NMF) for feature creation and reducing the number of attributes.

Additional ODM algorithms include

- Anomaly Detection (one-class Support Vector Machines) for finding suspicious activity, fraud, or rare events
- Attribute Importance, that is, finding the attributes most influential on a dependent field or specified business problem is provided by the Minimum Description Length (MDL) algorithm

Collectively, Oracle Data Mining's algorithms can address a broad range of business, technical, and scientific data mining problems.

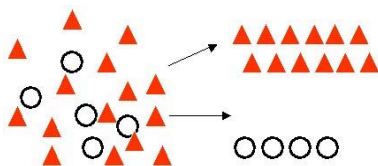
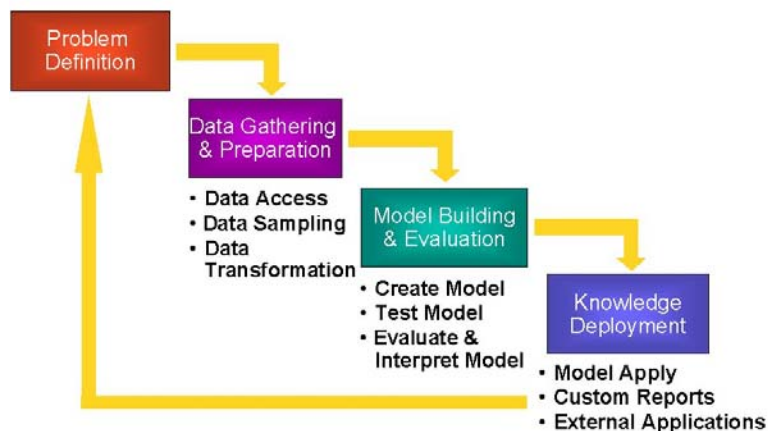
The Data Mining Process

To be effective in data mining, successful data analysts generally following a four step process:

1. **Problem definition** This is the most important step and is where the domain expert decides the specifics of translating an abstract business objective e.g. “How can I sell more of my product to customers?” into a more tangible and useful data mining problem statement, for example, “Which customers are most likely to purchase product A?” To build a predictive model that predicts who is most likely to buy product A, we first must have data that describes the customers who have purchased product A in the past. Then we can begin to prepare the data for mining.
2. **Data gathering and preparation** In this step, we take a closer look at our available data and determine what additional data we will need to address our business problem. We often begin by working with a reasonable sample of the data, e.g., hundred of records (rare, except in some life sciences cases) to many thousands or millions of cases (more typical for business-to-consumer cases). Some processing of the data to transform for example a “Date_of_Birth” field into “AGE” and to derive fields such as “Number_of_times_Amount_Exceeds_100” is performed to attempt the “tease the hidden information closer to the surface of the data” for easier mining. The fact that Oracle Data Mining’s data mining functionality is embedded inside the Oracle Database makes this data intensive process simpler.
3. **Model building and evaluation** Once steps 1 and 2 have been properly completed, this step is where the data mining algorithms sift through the data to find patterns and to build predictive models. Generally, a data analyst will build several models and change mining parameters in an attempt to build the best or most useful models. With ODM, unlike all competitors, there is no need to extract the data outside of the database to be mined.
4. **Knowledge deployment** Once Oracle Data Mining has found a useful model that adequately models the data, you want to distribute the new insights and predictions to others—managers, call center representatives, and executives. ODM’s embedded data mining algorithms eliminate any need to move (rewrite) the models to the data in the database or to extract huge volumes of unscored records for scoring using a predictive model that resides outside of the database. Oracle Data can generate the PL/SQL code for a Mining Activity. Because of this, Oracle Data Mining is the ideal platform for building business intelligence applications that automatically extract new information from your data and distribute it to where and when it is needed most.

The data mining process involves a series of steps to define a business problem, gather and prepare the data, build and evaluate mining models, and apply the models and disseminate the new information.

The Data Mining Process



Supervised learning techniques use historical data to build models that can be used to classify new data and make predictions about class membership (e.g. 0 or 1) or class value (numerical value).

Supervised Learning Algorithms

Most data mining algorithms can be separated into “supervised learning” and “unsupervised learning” data mining techniques. Supervised learning requires the data analyst to identify a target attribute or dependent variable. The supervised-learning technique then sifts through data trying to find patterns and relationships between the independent attributes (predictors) and the dependent attribute.

In supervised learning applications, a target attribute is defined to describe, for example, which customers have recently purchased a new car—for example, a “1” for “YES” and a “0” for “NO.” The supervised data mining algorithm sifts through the data finding patterns and builds a “model” that best represents the functional relationships. Typically, for the data mining process, the data is separated into two parts — one for training and another for testing. The initial model is built using the first, typically larger, sample of the data and then the model is applied to the second sample to evaluate the accuracy of the model’s predictions. Because we already know the outcome — who purchased a car and who didn’t — we can evaluate the model’s accuracy and make decisions about the usefulness of the model. Models with acceptable prediction capability often have high economic value. Binary and multi-class predictions together cover a huge range of important business problems, including campaign response, sales offers, fraud detection, profitability prediction, customer profiling, credit rating, churn anticipation, inventory requirements, failure anticipation, and many others. Oracle Data Mining also provides utilities for evaluating models in terms of model accuracy and “lift” — or the incremental advantage of the predictive model over the naïve guess.

Oracle Data Mining provides the following supervised data mining algorithms:

Naïve Bayes

ODM's Naïve Bayes (NB) algorithm is a supervised-learning technique for classification and prediction. ODM NB builds models that predict the probability of specific outcomes. The Naïve Bayes algorithm finds patterns and relationships in the data by counting the number of times various conditions are observed. It then builds a data mining model to represent those patterns and relationships. After ODM NB builds a model, the model can be used to make predictions. ODM NB enables application developers to build data mining models to classify and predict for a variety of purposes, such as:

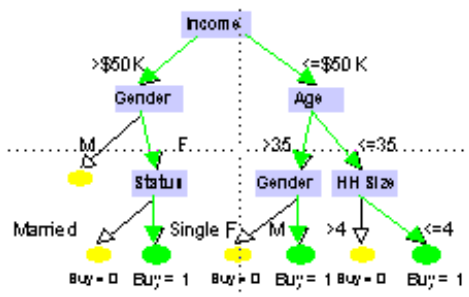
- Identify which customers are likely to purchase a certain product or who is likely to respond to a marketing campaign
- Identify which customers most likely to spend greater than \$3,000
- Identify customers likely to churn
- Predict the likelihood that a component will fail

ODM's Naïve Bayes algorithm makes predictions using Bayes' Theorem that assumes that each attribute is conditionally independent of the others. NB affords fast model building and scoring and can be used for both binary and multi-class classification problems.

Decision Trees

Oracle Data Mining offers a classical Decision Tree feature, based on the popular Classification & Regression Tree algorithm. The ODM Decision Tree model contains complete information about each node, including Confidence, Support, and Splitting Criterion. The full Rule for each node can be displayed, and in addition, a surrogate attribute is supplied for each node, to be used as a substitute when applying the model to a case with missing values.

Decision trees are popular because they are so universally applicable, easy to apply and easy to understand. Decision trees sift through each potential input attribute searching for the best "splitter", that is, attribute cut point e.g. AGE > 55 that splits the downstream data records into more homogeneous populations. After each split decision, ODM repeats the process growing out the entire tree and creating terminal "leaves" that represent "similar" populations of records, items, or people. Looking down from the root tree node, i.e. the total population, decision trees provide human readable "rules" of "IF A, THEN B" statements. These decision tree rules also provide the support and confidence for each tree node. Decision Trees provide full detailed rules for each splitting decision. Decision Trees are also useful for developing detailed "profiles" of, for example, your best customers, healthy patients, factors associated with fraud, etc.



Decision Trees are popular because they are easy to use and easy to understand.

Terminal tree nodes i.e. "leaves" provide detailed information about that segment of the population. For example:

Prediction: Buy = 1
Rule: INCOME < 50K, AGE > 35,
GENDER = Male
Confidence: = 72%
Support = .12

Generalized Linear Models (Logistic and Multiple Linear Regression)

ODM supports the classical statistical algorithm, Generalized Linear Models, as two mining functions: classification (Binary Logistic Regression) and regression (Multivariate Linear Regression). ODM supports GLM with the added capability to handle many hundreds to thousands of input attributes. Traditional external statistical software packages typically are limited to typically 10-30 input attributes. ODM logistic regression supports categorical target attributes, e.g. 0, 1 or A, B, or C while ODM multivariate linear regression supports numerical target attributes, e.g. loan amount.

Support Vector Machines

ODM's Support Vector Machines (SVM) algorithm supports binary and multi-class classification, prediction, and regression models, that is, prediction of a continuous target attribute. SVMs are particularly good at discovering patterns hidden in problems that have a very large number of independent attributes, yet have only a very limited number of data records or observations. SVMs can mine unstructured data, for example customer comments, email, or physician's comments.

SVM models can be used to analyze genomic data with only 100 patients who have thousands of gene expression measurements for each patient. SVMs can build models that predict disease treatment outcome based on genetic profiles.

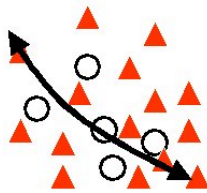
Attribute Importance

Oracle Data Mining's Attribute Importance algorithm helps to identify the attributes that have the greatest influence on a target attribute. Knowing which attributes are most influential helps you to better understand and manage your business and can help simplify modeling activities. Additionally, these attributes can indicate the types of data that you may wish to add to your data to augment your models.

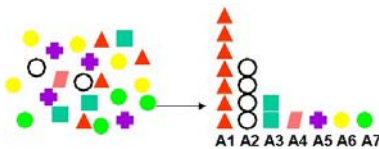
Attribute Importance can be used to find the process attributes most relevant to predicting the quality of a manufactured part, the factors associated with churn, or the genes most likely related to being involved in the treatment of a particular disease.

Unsupervised Learning Algorithms

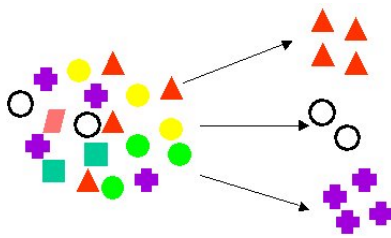
In "unsupervised learning," the user does not specify a target attribute for the data mining algorithm. Unsupervised learning techniques such as associations and clustering algorithms make no assumptions about a target field. Instead, they allow the data mining algorithm to find associations and clusters in the data independent of any a priori defined business objective.



ODM's Support Vector Machines support binary, multi-class, and regression models and excel at handling shallow, yet wide, data problems.



ODM's Attribute Importance algorithm finds the attributes that have the most influence on a target attribute.



ODM's Clustering techniques find natural groupings within the data.

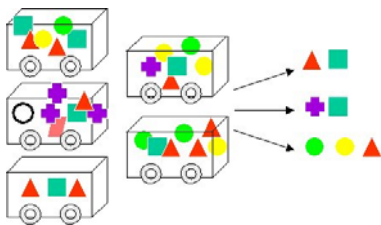
Clustering

Oracle Data Mining provides two algorithms, Enhanced k-Means and Orthogonal Partitioning Clustering (O-Cluster), for identifying naturally occurring groupings within a data population.

ODM's Enhanced k-Means and O-Cluster algorithms support identifying naturally occurring groupings within the data population. ODM's K-means algorithm supports hierarchical clusters, handles numeric and categorical attributes and will cut the population into the user specified number of clusters.

ODM's O-cluster algorithm handles both numeric and categorical attributes and will automatically select the best cluster definitions. In both cases, ODM provides cluster detail information, cluster rules, cluster centroid values, and can be used to "score" a population on their cluster membership.

For example, enhanced k-Means Clustering can be used to reveal subgroups within a diseased population or to find new customer segments.

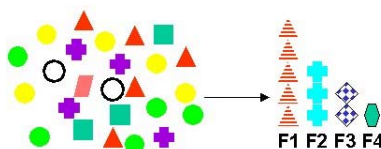


ODM's Association Rules performs "market basket analysis" to find commonly co-occurring items or events.

Association Rules

ODM's Association Rules (AR) finds co-occurring items or events within the data. Often called "market basket analysis", AR counts the number of combinations of every possible pair, triplet, quadruplet, etc., of items to find patterns. Association Rules represent the findings in the form of antecedents and consequents. An AR rule, among many rules found, might be "Given Antecedents Milk, Bread, and Jelly, then Consequent Butter is also expected with Confidence 78% and Support 12%." Translated in simpler English, this means that if you find a market basket having the first three items, there is a strong chance (78% confidence) that you will also find the fourth item and this combination is found in 12% of all the market baskets studied. The associations or "rules" thus discovered are useful in designing special promotions, product bundles, and store displays.

AR can be used to find which manufactured parts and equipment settings are associated with failure events, what patient and drug attributes are associated with which outcomes or which items or products is a person who has purchased item A most likely to buy?



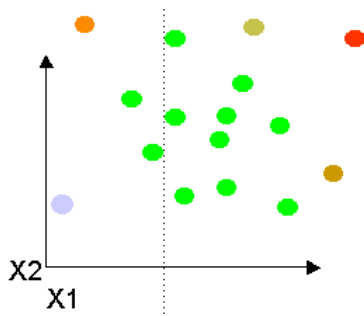
ODM's Nonnegative Matrix Factorization creates new attributes that can represent the same information using fewer attributes.

Feature Selection

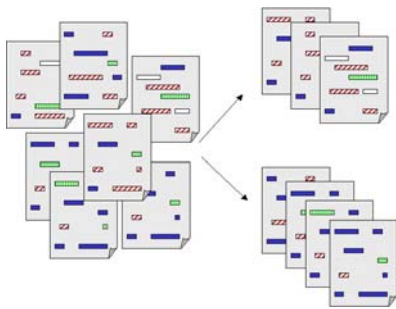
ODM's Nonnegative Matrix Factorization (NMF) is useful for reducing a large dataset into representative attributes. Similar to Principal Components Analysis (PCA) in concept, but able to handle larger amounts of attributes and in an additive representation model, NMF is a powerful, state-of-the-art data mining algorithm that can be used for a variety of use cases.

NMF can be used to reduce large amounts of data, e.g. text data, into smaller, more sparse representations that reduce the dimensionality of the data, i.e., the same information can be preserved using far fewer variables. The output of NMF models can be analyzed using supervised learning techniques such as SVMs or

unsupervised learning techniques such as clustering techniques. Oracle Data Mining uses NMF and SVM algorithms to mine unstructured text data.



Oracle Data Mining can learn what is considered “normal” and then flag unusual items, suspicious activity or rare events.



ODM can mine both structured and unstructured data, i.e., text, to build classification and clustering models.



Integrating SVM scoring with spatial data to find objects near stone towers:

```
SELECT DISTINCT
a.pixel.sdo_point.x x,
a.pixel.sdo_point.y y
FROM quadrant a,
stone_steel b
WHERE
SDO_WITHIN_DISTANCE
(a.pixel, b.pixel,
'distance=150') =
'TRUE');
```

Anomaly Detection

Release 2 of Oracle Data Mining 10g introduced support for a new mining application—anomaly detection, that is, the detection of “rare cases” when very few or even no examples of the rare case are available. Oracle Data Mining can “classify” data into “normal” and “abnormal” even if only one class is known. ODM uses a version of the Support Vector Machines algorithm to create a profile of a known class. When the model is applied to the general population, cases that don’t fit the profile are flagged as anomalies (that is, abnormal or suspicious). Oracle Data Mining’s anomaly detection algorithm can flag unusual accounting transactions as is required by Sarbanes-Oxley.

Text Mining

Oracle Data Mining provides a single unified analytic server capability as part of the Oracle Database to mine both structured, that is, columnar data organized in rows and columns, and unstructured data. ODM can mine unstructured data, such as, “text” combined with other structured data, for example, age, height, and weight to build classification, prediction, and clustering models. ODM could add, for example, a physician’s notes to the structured “clinical” data to extract more information and build better data mining models.

This ability to combine structured data with unstructured data opens new opportunities for mining data. For example, law enforcement personnel can build models that predict criminal behavior based on age, spatial location, number of previous offenses, income, and so forth, and combine a police officer’s notes about the person to build more accurate models that take advantage of all available information.

Additionally, ODM’s ability to mine unstructured data is used within Oracle Text to classify and cluster text documents stored on the Database, e.g. Medline. Oracle Data Mining’s NMF and SVM models can be used with Oracle Text to build advanced document classification and clustering models.

Spatial Data Mining

Effective with Oracle Database 10g, spatial data can be materialized for inclusion in data mining applications. Thus, ODM might enable you to discover that sales prospects with addresses located in specific areas (neighborhoods, cities, or regions) are more likely to watch a particular television program or to respond favorably to a particular advertising solicitation. (The addresses are geocoded into longitude/latitude points and stored in an Oracle Spatial geometry object.)

In many applications, data at a specific location is influenced by data in the neighborhood. For example, the value of a house is largely determined by the value of other houses in the neighborhood. This phenomenon is called spatial correlation (or, neighborhood influence). The spatial analysis and mining features in Oracle Spatial let you exploit spatial correlation by using the location attributes of data items in several ways: for binning (discretizing) data into regions (such as categorizing data into northern, southern, eastern, and western regions), for materializing the influence of neighborhood (such as number of customers within a two-mile radius of each store), and for identifying co-located data items (such as video rental stores and pizza restaurants).

Predictive Analytics

Predictive Analytics (PA) are totally automated data mining PL/SQL packages that are designed to provide the maximum data mining results by requiring the minimum information from the user. ODM provides three Predictive Analytic packages:

- **DBMS_PREDICTIVE_ANALYTICS.PREDICT:** User specifies a target attribute, a table and an output table name and PA_PREDICT automatically builds a predictive model and scores every record with a prediction.
- **DBMS_PREDICTIVE_ANALYTICS.EXPLAIN:** User specifies a target attribute, a table and an output table name and PA_EXPLAIN builds an attribute importance model and ranks attributes in their strength of relationship to the target attribute.
- **DBMS_PREDICTIVE_ANALYTICS.PROFILE:** User specifies a target attribute, a table and an output table name and PA_PROFILE builds a decision tree model with associated leaf node “segments” and information about each segment.

Each Predictive Analytic PL/SQL package may be accessed directly using the PL/SQL API, the Spreadsheet Add-in for Predictive Analytics, or the Oracle Data Miner graphical user interface.

DO MORE

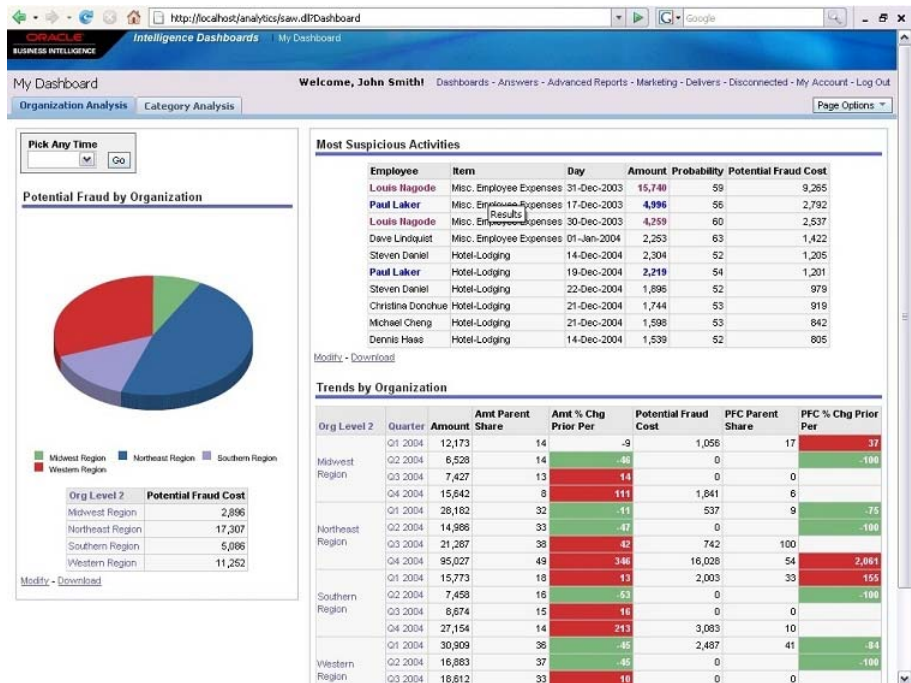
Moving from a “Tool” to an “Enterprise Application”

Successful data mining occurs when it can become hidden, integrated, and seamless—in essence, when data mining can be “operationalized” as a part of everyday activities. Because all of Oracle Data Mining’s functionality is embedded in the Oracle11g Database, you can easily implement automated data mining applications.

Oracle Data Mining makes it easy to quickly develop applications that sift through your corporate data and distribute ODM's newly extracted information and insights to those who need it—managers, technical analysts, and customer facing personnel. ODM eliminates the extraction of data from the database for data mining, thus significantly reducing total cost of ownership. With ODM, there is no need for multiple data storage hardware and software environments, multiple data analysis tools, and multiple support resources. With ODM, there are fewer “moving parts” resulting in a simpler, more reliable, and more efficient data management and data analysis environment.

ODM users can automate the extraction and integration of new insight and predictions into a variety of business applications, including call centers, web sites, campaign management systems, automatic teller machines (ATMs), enterprise resource management (ERM), and other operational and business planning applications. Rather than having special departments of advanced data analysts who work on ad hoc data mining projects, the true value of data mining is realized when the new insights and predictions are integrated and operationalized into existing business applications.

Oracle Data Miner enables users to gain greater insights from Query and Reporting tools and Dashboards such as this Oracle BI EE Executive Dashboard that tracks potential fraudulent activities



Automation of data mining tasks is facilitated by Oracle Data Mining's PL/SQL and Java-based APIs. Application programmers can control all aspects of data mining — they can expose complex settings for advanced users or completely automate the process for business users. Programmatic control extends from data

preparation (including the DBMS_DATA_MINING_TRANSFORM PL/SQL package and the Java APIs) and model building to on-demand scoring of single records and batch scoring of large data sets. Batch scores may be stored in relational tables for access by other business applications, for example, call centers or marketing campaign systems or called “on-demand” in interactive applications where new information is collected that must be factored into the predictive model.

Oracle Data Mining’s PL/SQL or Java-based APIs provide direct, asynchronous access to the ODM’s functionality. Oracle Data Mining’s PL/SQL and Java-based APIs enable application developers to enhance, for example, a call center application to highlight a customer’s likelihood to churn or to become a profitable customer. The probability that the customer will accept the special offers can be displayed for the customer service representative as a window pop up to provide better service the customer.

Because all results are created and stored in an open relational database, users have access to data mining results using a wide variety of business intelligence tools including Oracle BI EE, Oracle BI SE, Oracle OLAP, Oracle Reports, Oracle Portal, and Oracle Marketing Online.

Oracle Data Mining provides four interfaces to its data mining functionality in the Oracle Database:

ODM JAVA API—ODM’s Java API provides programmatic control of data mining functions to enable the automation of data preparation, model building, and model scoring operations to build advanced BI applications.

ODM DBMS_DATA_MINING PL/SQL API—Application Developers can use ODM’s PL/SQL API and DBMS_DATA_MINING_TRANSFORM PL/SQL package to build advanced BI applications using a set of SQL primitives that can be invoked in PL/SQL program block(s). ODM’s PL/SQL API provides a language and development methodology familiar to the majority of Oracle server developers and Database Administrators (DBAs).

Oracle Data Miner—Oracle Data Miner provides a graphical user interface for data analysts to build, evaluate, and apply data mining models. Oracle Data Miner guides the data analyst through the data mining process with complete flexibility and presents results in graphical and tabular formats. Oracle Data Miner can generate the PL/SQL code associated with a Mining Activity.

Oracle Spreadsheet Add-In for Predictive Analytics—This Add-In to Microsoft Excel allows users to access the fully automated PREDICT, PROFILE and EXPLAIN PL/SQL packages. Data may be in either Excel or the Database.

Oracle Data Mining is part of Oracle’s family of business intelligence products and features that include Oracle Discoverer, Oracle OLAP, Oracle Reports, Oracle Portal, and Oracle Marketing Online. With Oracle, the data can come from the same “single source of truth” and employ enterprise user access and security

schemes. By using Oracle Data Mining, building enterprise applications that automate data mining and distribute new insights within the organization is easy.

Fewer Moving Parts

Because Oracle Data Mining performs all phases of data mining within the database, each data mining phase results in order of magnitude improvements in productivity, automation, and integration. Significant productivity enhancements are achieved by eliminating the extraction of data from the database to special-purpose data mining tools and followed by the importing of the data mining results back into the database. This integrated architecture is most notable in the data preparation process which often can constitute as much as eighty percent of the data mining process. With Oracle Data Mining, data preparation is performed using standard Oracle SQL, PL/SQL and Java programs for data selection, access, transformation, manipulation and data mining. Because Oracle Data Mining leverages your investment in Oracle database technology and personnel, you can standardize on a consistent Information Technology (IT) platform and readily available technical skills to reduce total costs of ownership and to improve overall productivity.

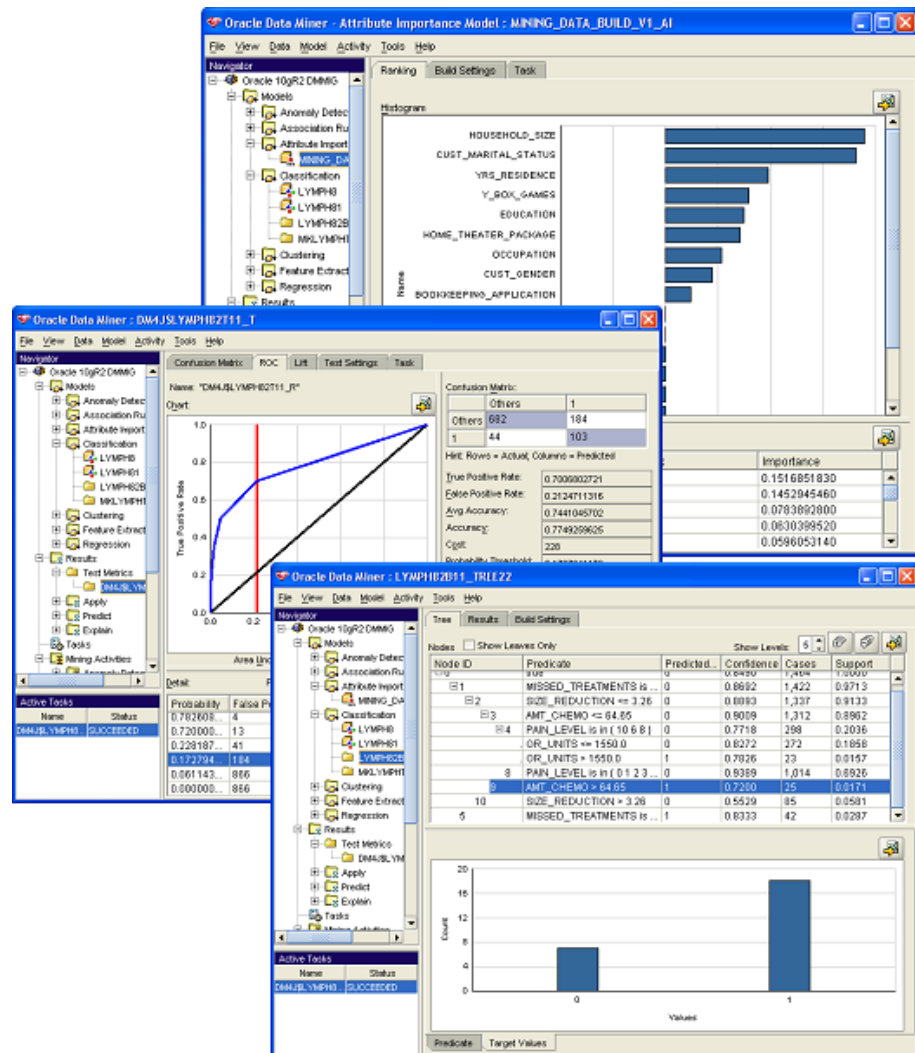
Benefits of in-database mining become apparent in the first steps of a data mining project. Data preparation includes the analysis and assembly of relevant data, definition of a target field, and possibly creation of new summary or calculated fields. Since mining will occur in the database, any tools or applications can contribute to this process. It is simple to use lists and measures from existing business applications as the starting point for data mining.

“Scoring” to make predictions is greatly simplified now, as the data and the model are in the same location — the database. There is no need to move the data to the model nor export the model as code and then move the model back to the data for scoring in the database. With Oracle Data Mining, everything occurs within the database.

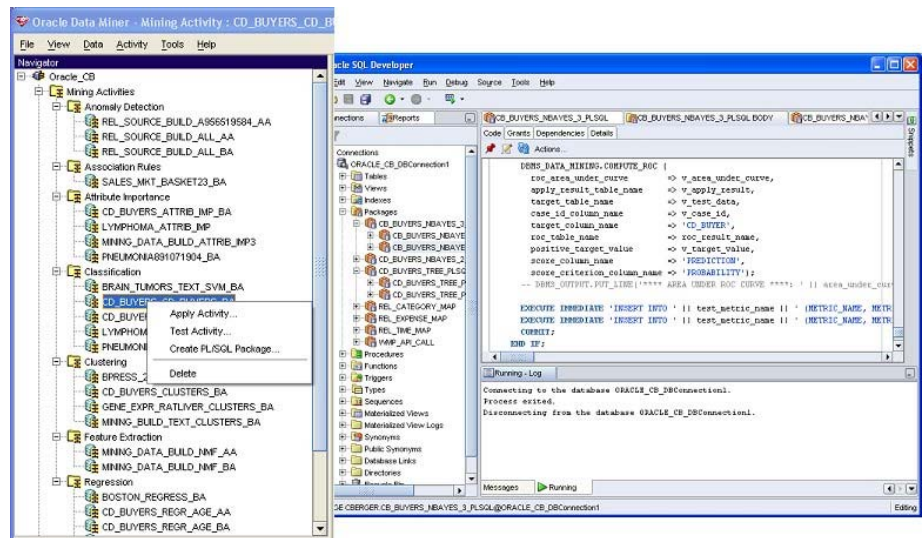
Oracle Data Miner

Starting with Oracle 10g, Oracle provides a graphical user interface “client” to Oracle Data Mining. Oracle Data Miner provides a wizards approach that supports all mining activities. Mining Activity Guides provide structured templates for all users to explore and mine their data. The expert analyst can adjust some or all of the parameters manually, but the option is available to allow the algorithms to optimize the parameters intelligently with no intervention. Oracle Data Miner supports the simultaneous building of models and applications because it can automatically generate the (Java and/or SQL code) associated with the data mining activities. This tight coupling of the data mining process and the applications development process uniquely positions ODM as the platform of choice for building automated and integrated data mining applications.

Oracle Data Miner provides an intuitive easy to use graphical user interface to Oracle Data Mining’s algorithms.



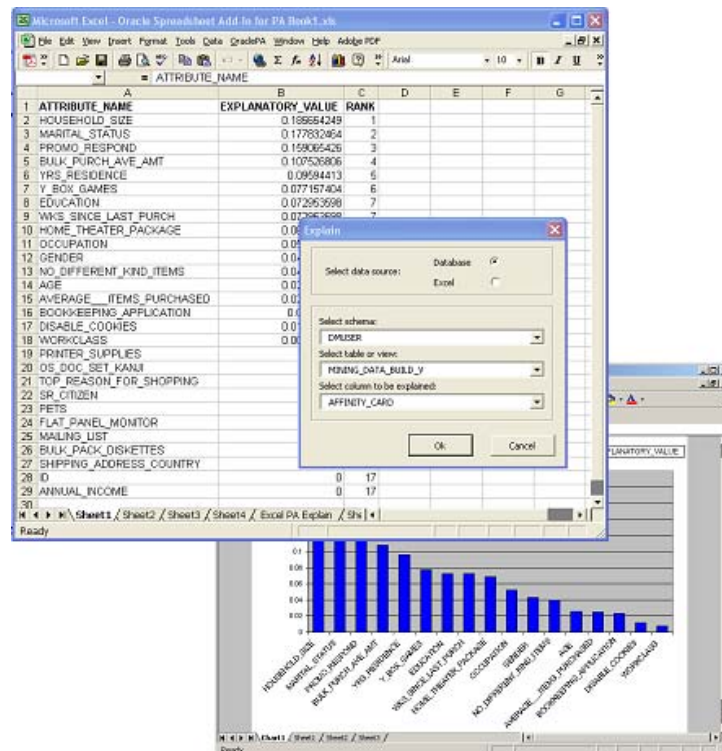
Oracle Data Miner works with SQLDeveloper and JDeveloper and automatically generates PL/SQL code for a Mining Activity. This significantly accelerates the development of advanced analytical applications.



Oracle Spreadsheet Add-in for Predictive Analytics

Oracle Data Mining can embed within an Excel spreadsheet Explain and Predict functions that act on a database table or view, or on data contained in the spreadsheet. The user need only identify the data and the target attribute; the data is prepared and the algorithm parameters are optimized before the function is executed, and the results are returned to the spreadsheet for display and storage.

The “one-click” PREDICT, PROFILE and EXPLAIN PL/SQL package automates data mining. Users can access data in either Microsoft Excel or the Database using this simple Excel Add-in.



Model Apply and Model Scoring

Once an Oracle Data Mining data mining model has been built, applying the data mining model to new data can score new data records. Scoring is the process of predicting outcomes. In Oracle Data Mining, the data mining algorithm which carries out model build and model apply, occurs in the database, rather than requiring the data to be extracted for the algorithm as in all other data mining tools. This is a more efficient architectural strategy, especially when dealing with large amounts of data.

Oracle Data Mining can apply a model in batch mode or on demand. Batch “model apply” takes a table of records as input and stores the predictions in another table. On-demand predictions takes a single record and returns a prediction that can be fed directly into the “calling” application.

Build Models and Applications Simultaneously

Oracle Data Miner is the graphical user interface for data analysts to view data tables and data views accessible by the Oracle Database, view the data in histograms and perform data transformations, build and evaluate data mining models and apply the models to “score” new data with model predictions. Oracle Data Miner facilitates interactive data preparation, data mining model creation, evaluation, refinement and model scoring. Interactive wizards and Mining Activity Guides guide the user through the data mining process. When a useful model has been identified, Oracle Data Miner and the JDeveloper Code Generator add-in automatically creates the Java and code necessary to automate the data mining process. This synergy between model creation and model deployment provides significant cost and productivity advantages.

Oracle CRM 11i

Customer Relationship Management Application That Integrates Data Mining

Oracle CRM 11i application provides integrated data mining to automate model-building and model-scoring activities for a set of predefined business problems. By constraining the problem and having a known schema, data mining can make certain modeling assumptions to automate the model-building and model-scoring process. CRM 11i's Marketing Online application allows business analysts and marketing professionals to take advantage of the power of data mining without the need to learn complex analytical techniques or have to perform extensive data preparation. CRM 11i now provides easy and accurate answers to business questions:

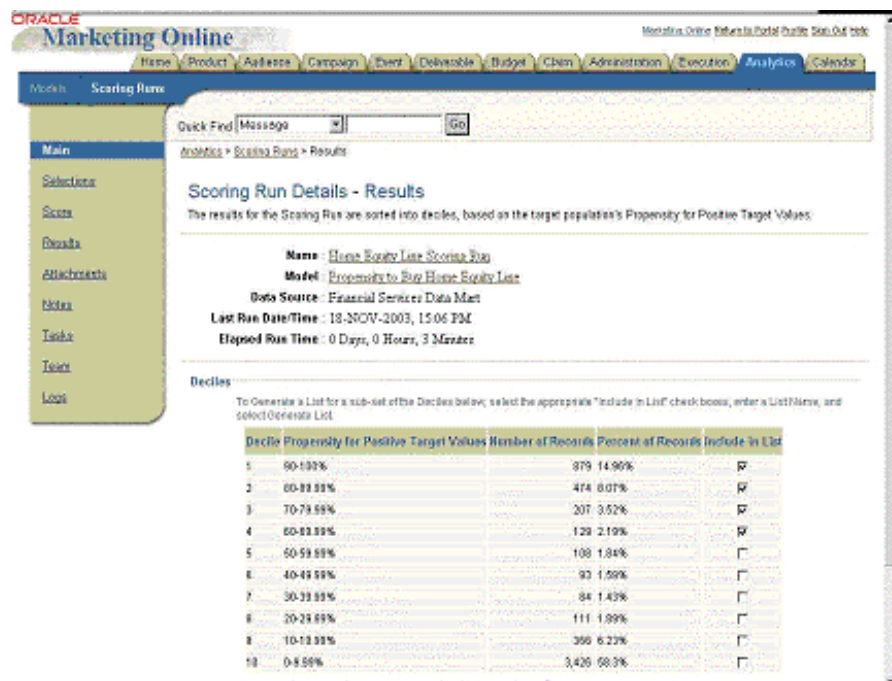
- Which customers are likely to respond to an email offer
- Which customers are likely to respond to a direct mail offer?
- Which customers are likely to respond to a telesales offer?

- Which customers are likely to remain loyal?

Oracle CRM 11i simplifies and integrates data mining. Marketing managers select past “good customers” and then ODM builds predictive models based on a standard schema and reasonable defaults. Marketing managers can then select the level of uncertainty about customer behavior that they want to accept.

The results of CRM 11i's Marketing Online application are available in the database for other applications, such as call centers, to display valuable customer insights. CRM users can design targeted campaigns by selecting “Audience” and answering a few simple questions. All of the data preparation, data mining, and data “scoring” occur automatically for the users. Market managers can simply select the top deciles of customers identified to be most likely to respond or behave in a particular way.

Using Oracle CRM11i, marketing managers can automatically select the level of uncertainty about customer behavior that they want to accept and implement marketing campaign that yield higher returns.



SPEND LESS

**Eliminate Redundant Data, Servers, Software and Personnel
 —And Pay Less for Data Mining Software**

Oracle Data Mining significantly reduces the cost of data mining. Oracle Data Mining eliminates the need for separate, dedicated analytical, statistical or mining servers and the need to extract data outside of the Oracle Database.

Savings are realized in avoiding additional hardware purchases for computing and storage environments, redundant copies of the data and multiple versions of the

data, duplication of personnel who perform similar functions but unnecessarily use different software packages. Additionally, by utilizing the same data and a “single source of truth”, problems resulting in working with the wrong, or out of date data, can be reduced.

Oracle Data Mining is a priced option to the Oracle Database Enterprise Edition (EE). Oracle Data Mining leverages your existing investment and relationship with Oracle. See your Oracle Sales Representative for pricing details.

CONCLUSION

Oracle Data Mining provides a powerful, scalable infrastructure for building applications that automate the discovery of valuable new business intelligence and insights and its integration into other applications. Oracle Data Mining’s wide range of “in-database” data mining algorithms can solve a wide variety of business problems and provides a powerful infrastructure for building advanced enterprise business intelligence applications.

By automating, integrating, and “operationalizing” the discovery and distribution of new business intelligence, companies can leverage their investment in data, find valuable new insights, operate more effectively, and obtain greater competitive advantage.



White Paper Oracle Data Mining

June 2007

Author: Charlie Berger

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:

Phone: +1.650.506.7000

Fax: +1.650.506.7200

www.oracle.com

Copyright © 2007, Oracle. 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.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective owners.