An Oracle White Paper
June 2009

Oracle Data Mining 11g
Competing on In-Database Analytics

ORACLE®

Disclaimer

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.

"Some companies have built their very businesses on their ability to collect, analyze, and act on data. …Although numerous organizations are embracing analytics, only a handful have achieved this level of proficiency. But analytics competitors are the leaders in their varied fields—consumer products finance, retail, and travel and entertainment among them."

—Tom Davenport, author, *Competing on Analytics*

Executive Overview

The Oracle Data Mining Option is powerful data mining functionality embedded in the Oracle Database that enables you to discover new insights hidden in your data and leverage your investment in Oracle Database technology. With Oracle Data Mining, businesses can build and apply predictive models that help them target their best customers, develop detailed customer profiles, find and prevent fraud, discover the attributes that most influence others, and find valuable new information hidden in their data. Oracle Data Mining helps companies find important patterns in their data and uncover valuable insights—that help them better *compete on analytics.*
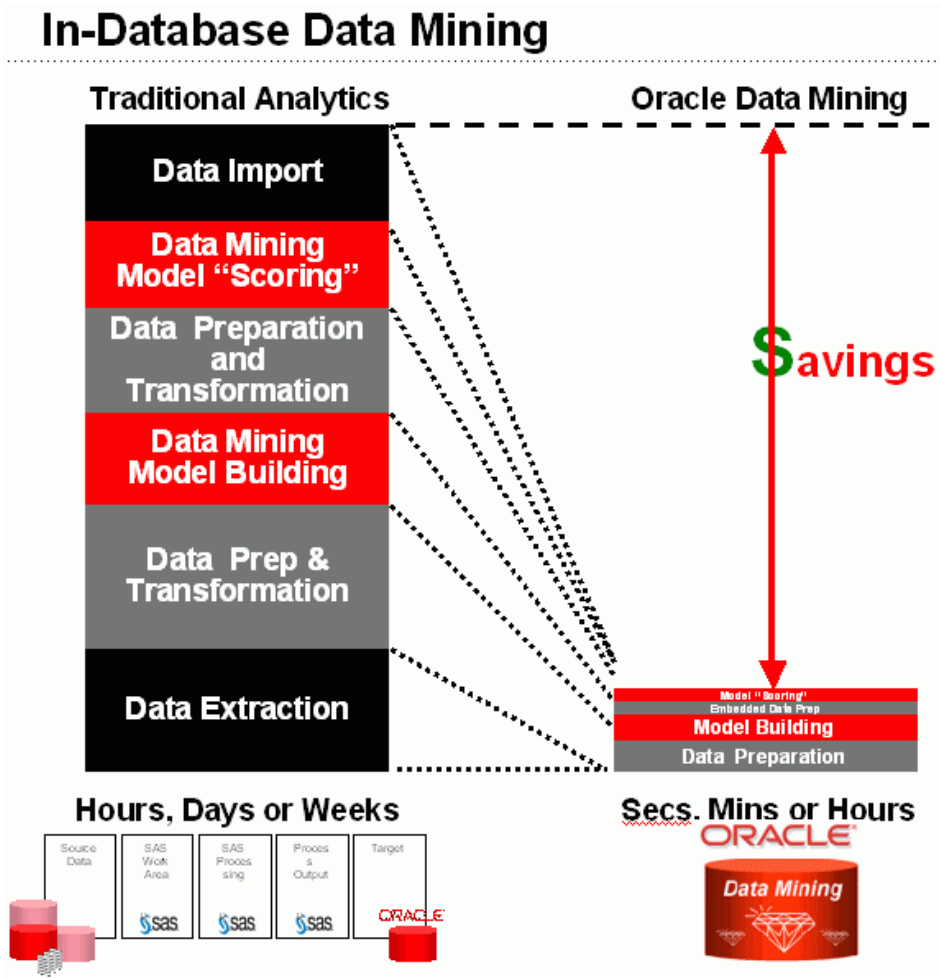
In-Database Data Mining

With Oracle Data Mining, everything occurs in the Oracle Database—in a single, secure, scalable platform for advanced business intelligence. Oracle Data Mining represents a breakthrough in business intelligence. In contrast to traditional statistical software that requires users to extract data to separate, unsecured, and costly dedicated statistical and data mining servers for analysis, Oracle Data Mining embeds a wide-range of mining functions inside the database—where the data is stored. Coupled with the power of SQL, Oracle Data Mining eliminates data movement and duplication, maintains security and minimizes latency time from raw data to valuable information.

Oracle Data Mining enables companies to:

- Leverage your data to discover patterns and valuable new insights

- Build and apply predictive models and embed them into dashboards and applications

- Save money. Oracle Data Mining costs significantly less than traditional statistical software. As an integrated component of your Oracle IT platform, Oracle Data Mining

significantly reduces your total cost of ownership.

In-Database Data Mining

Traditional Analytics

- Data Import
- Data Mining Model "Scoring"
- Data Preparation and Transformation
- Data Mining Model Building
- Data Prep & Transformation
- Data Extraction

Hours, Days or Weeks

Source Data	SAS Work Area	SAS Process sing	Process Output	Target
§sas	§sas	§sas	ORACLE	

Oracle Data Mining

$avings

- Model "Scoring"
- Embedded Data Prep
- Model Building
- Data Preparation

Secs. Mins or Hours

ORACLE

Data Mining

Oracle Data Mining eliminates data movement, data duplication and security exposures.

Key Benefits

- Oracle Data Mining mines data inside the database

- Oracle Data Mining eliminates data extraction and movement

- Provides a platform for analytics-driven database applications

- Increased security—leverages database security options

- Industry leading performance and reliability

- Delivers lowest total cost of ownership (TCO) compared to traditional data mining vendors

- Leverages 30+ years of experience of ever advancing Oracle Database technology

Oracle Data Mining enables you to go beyond standard BI and OLAP tools that tell you "Who are my top customers?", "What products have sold the most?", and "Where are costs the highest?". Data mining automatically sifts through data and reveals patterns and insights that help you run your business better. In today's competitive marketplace, companies must manage their most valuable assets — their data. Moreover, companies must exploit their data for competitive advantage. If they don't, their competitors will. With Oracle Data Mining, you can implement strategies to:

- Understand and target select customer groups

- Develop detailed customer profiles for building marketing campaigns

- Anticipate and prevent customer churn and attrition

- Identify promising cross-sell and up-sell opportunities

- Detect noncompliance and potential fraud

- Discover new clusters or customer segments

- Perform market-basket analysis to find frequently co-occurring items

"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

Introduction

Traditional business intelligence (BI) reporting tools report on what has happened in the past. OLAP provides rapid drill-through for more detailed information; roll ups and sometimes aggregates level forecasting. With a good tool, a good analyst and enough time, you *could eventually find* the information for which you are looking—like the infinite monkey theorem that states that a monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type a given text, such as the complete works of William Shakespeare.

Data Mining is now possible due to advances in computer sciences and machine learning that deliver new algorithms that can *automatically* sift deep into your data at the individual record level to discover patterns, relationships, factors, clusters, associations, profiles, and predictions—that were previously "hidden".

Oracle Data Mining, a collection of machine learning algorithms embedded in the Oracle Database, allows you to discover new insights, segments and associations, make more accurate predictions, find key variables, detect anomalies, and extract more information from your data. For example, by analyzing your best customers, ODM can enable discover profiles and embed predictive analytics in applications that identify customers who are likely to be your best customers. These customers may not represent your most valuable customers today, but they match profiles of your current best customers. With ODM you can apply predictive models to generate reports and dashboards of your most promising customers to your marketing department, sales force, or real-time predictions within call center applications. Knowing the "strategic value" of your customers — which customers 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 at just the right time into your operations is the key to successfully competing on analytics.

Oracle Data Mining

Oracle Data Mining, an Option to the Oracle Database Enterprise Edition, delivers a wide range of cutting edge machine learning algorithms inside the Oracle Database. With ODM, data mining functions reside natively in the SQL kernel of the Oracle Database to deliver unparrelled performance, scalability and security—the data and data mining functions never leave the database to deliver a comprehensive in-database processing solution.

Oracle Data Mining Release 11*g* supports twelve in-database mining algorithms that address classification, regression, association rules, clustering, attribute importance, and feature selection problems. Working with Oracle Text (which uses standard SQL to index, search, and analyze text and documents stored in the Oracle database, in files, and on the web), many ODM mining functions can mine both structured and unstructured (text) data.

ODM provides PL/SQL and Java application programming interfaces (APIs) for model building and model scoring functions. An optional Oracle Data Miner graphical user interface (GUI) that is available for download from OTN is available for data analysts who want to use a point and click GUI.   The Spreadsheet Add-In for Predictive Analytics provides a PL/SQL package that exposes automated predictive analytics through Microsoft Excel. The ODM graphical user interface(s) and APIs provide an analytical platform for data analysts and application developers to deliver data mining's results to BI dashboards and enterprise applications.

Ranges of Business Intelligence

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

BI, OLAP and Data Mining

BI	OLAP	Data Mining
Extraction of detailed and roll up data	Summaries, trends and forecasts	Knowledge discovery of hidden patterns
"Information"	*"Analysis"*	*"Insight & 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?**

Figure 1. BI, OLAP, Statistics and Data Mining. Oracle Data Mining differs from query and reporting (BI), and OLAP tools by automatically discovering new information that was previously hidden in the data and the ability to make predictions.

BI and 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 basic BI and allow users to rapidly and interactively drill-down for more detail, comparisons, summaries 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?"

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 when the amount of data begins to overwhelm the tool, 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. However, eliminating input variables and using small samples of data, makes you throw away valuable information.

Oracle Data Mining runs in the SQL kernel of the Oracle Database and doesn't suffer from the same limitations. Oracle Data Mining uses machine-learning techniques developed in the last decade to automatically find patterns and relationships hidden in the data. Oracle Data mining goes deep into the data and finds patterns from the data. Oracle 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?"

Data Mining Deep Dive

Oracle Data Mining provides a collection of data mining algorithms that are designed to address a wide-range of business and technical problems. Different algorithms are good at different types of analysis. Oracle Data Mining supports classification, regression, and prediction, clustering, associations, attribute importance and feature extraction problems. (For a full listing of algorithms, see Oracle Data Mining 11g Algorithms table on page 9.)

Sequence for Determining Necessary Data. *Wrong:* Catalog everything you have, and decide what data is important. *Right:* Work backward from the solution, define the problem explicitly, and map out the data needed to populate the investigation and models.

—James Taylor with Neil Raden, authors*, Smart (Enough) Systems*

The Data Mining Process

Before we start mining any data, we need to first define our problem to be addressed and most, importantly, have the right data gathered to help us answer that question. If we don't have the right data, we need to get it. With data mining, if not properly approached, there is the

possibility of "garbage in—garbage out". To be effective in data mining, successful data analysts typically following a four step process:

Defining the Business Problem

This is the most important step and is where a 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 e.g. "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.

Gathering and Preparing the Data

Now we take a closer look at our data and determine what additional data may be necessary to properly 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) and explore the data by looking at statistical summaries and histograms. Some processing of the data to transform for example a "Date_of_Birth" field into "AGE" and to derive fields such as "No_Times_Amt_Exceeds_N" is performed to attempt the tease the hidden information closer to the surface of the data for easier mining. The power of SQL makes this process simpler.

Model Building and Evaluation

Now we are ready to for the mining algorithms to sift through the data, discover patterns and build predictive models. Generally, a data analyst will build several models and change mining parameters in an attempt to build and evaluate many models to find the best or most useful model(s).

Knowledge Deployment

Once ODM has built a model that models relationships found in the data, you want to apply it to deploy 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 Mining provides the ideal platform for building and deploying advanced business intelligence applications.
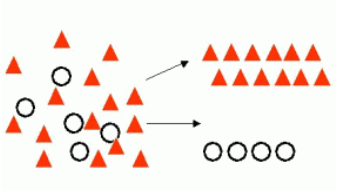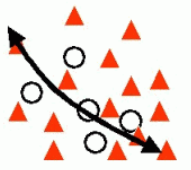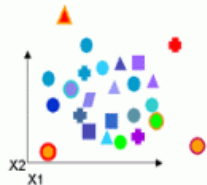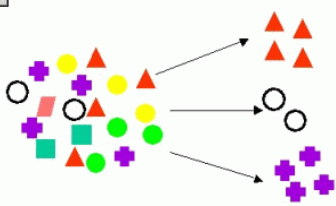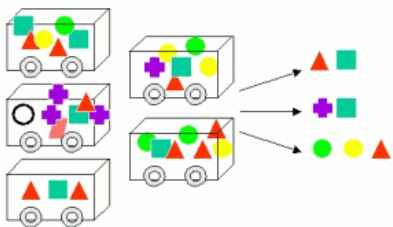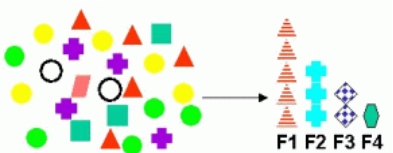
The Data Mining Process
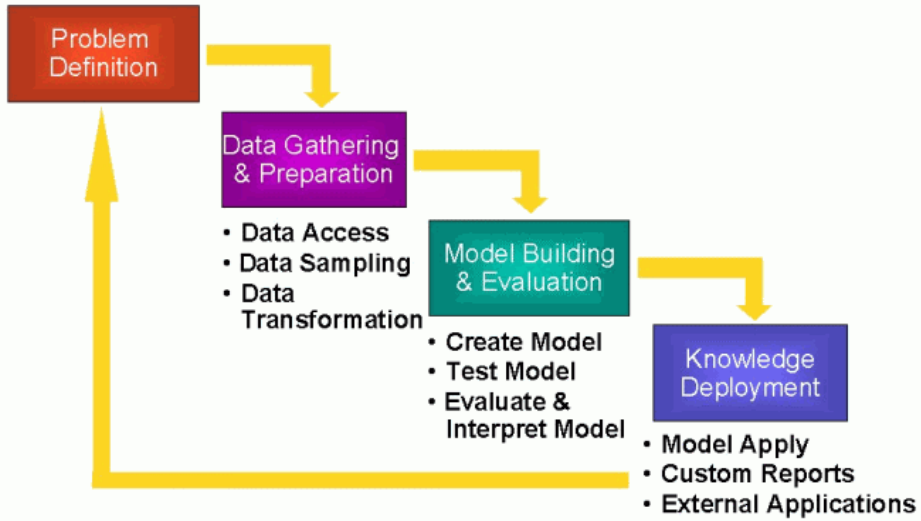
The data mining process involves a series of steps to define a business problem, gather and prepare the data, build and evaluate mining models, and apply the models and disseminate the new information.

A Wide Range of Cutting Edge Algorithms

Oracle Data Mining provides a broad suite of data mining techniques and algorithms to solve many types of business and technical problems:

ORACLE DATA MINING 11G ALGORITHMS

TECHNIQUE	APPLICABILITY	ALGORITHMS
Classification	Classification 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).	Logistic Regression (GLM)— classic statistical technique inside the Oracle Database. Supports thousands of input variables, text and transactional data Naive Bayes—Fast, simple, commonly applicable Support Vector Machine—Cutting edge algorithm that supports binary and multi-class problems. SVMs also excel at handling shallow, yet wide, and nested data problems e.g. transaction data or gene expression data analysis. Supports text mining use cases. Decision Tree—Popular algorithm useful for many classification problems that that can help explain the model's logic using human-readable "If…. Then…" rules.
Regression	Technique for predicting a continuous numerical outcome such as customer lifetime value, house value, process yield rates.	Multiple Regression (GLM)— classic statistical technique available inside the Oracle Database Supports thousands of input variables, text and transactional data Support Vector Machine — Cutting edge algorithm that supports regression problems. Supports text mining and transactional data use cases.
Attribute Importance	Ranks attributes according to strength of relationship with target attribute. Examples include finding factors associated with people who respond to an offer, factors most associated with healthy patients.	Minimum Description Length— Considers each attribute as a simple predictive model of the target class. Attribute Importance algorithm finds the attributes that have the most influence on a target attribute.

Anomaly Detection		Identifies unusual or suspicious cases based on deviation from the norm. Common examples include health care fraud, expense report fraud, and tax compliance.	One-Class Support Vector Machine —Unsupervised learning technique that trains on "normal cases" to build a model. Then when applied, it flags unusual cases with the probability that they are not "normal".
Clustering		Useful for exploring data and finding natural groupings within the data. Members of a cluster are more like each other than they are like members of a different cluster. Common examples include finding new customer segments, and life sciences discovery.	Enhanced K-Means—Supports text mining, hierarchical clustering, distance based Orthogonal Partitioning Clustering—Hierarchical clustering, density based
Association		Finds rules associated with frequently co-occurring items, used for market basket analysis, cross-sell, root cause analysis. Useful for product bundling, in-store placement, and defect analysis.	Apriori—Industry standard for market basket analysis.
Feature Extraction		Produces new attributes as linear combination of existing attributes. Applicable for text data, latent semantic analysis, data compression, data decomposition and projection, and pattern recognition.	Non-negative Matrix Factorization (NMF)— Creates new attributes that represent the same information using fewer attributes

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 with examples of the possible classes (e.g., 0/1, Yes/No, High/Med, Low, etc.). The supervised-learning technique then sifts through data trying to find patterns and relationships among the independent attributes (predictors) that can help separate the different classes of the dependent attribute.

In this example supervised learning data mining problem, let's say that we want to build a predictive model that can help our Marketing and Sales departments focus on people who are most likely interested in purchasing a new car.   The target attribute will be a column that designates whether each customer has purchased a car—for example, a "1" for yes and a "0" for no. The supervised data mining algorithm sifts through the data searching for patterns and builds a data mining model that captures the relationships found in the data. Typically, for the supervised learning, the data is separated into two parts — one for model training and another hold out sample for model testing and model evaluation. Because we already know the outcome — who purchased a car and who hasn't — we can apply our ODM predictive model to our hold out sample to evaluate the model's accuracy and make decisions about the usefulness of the model. ODM models with acceptable prediction capability can have high economic value. Binary and multi-class classification problems represent a majority of common business challenges addressed through Oracle Data Mining, including database marketing, response and 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.

Naïve Bayes

Naïve Bayes (NB) is a supervised-learning technique for classification and prediction supported by Oracle Data Mining. The Naive Bayes algorithm is based on conditional probabilities. It uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data. Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. If B represents the dependent event and A represents the prior event, Bayes' theorem can be stated as follows.

Bayes' Theorem: Prob(B given A) = Prob(A and B)/Prob(A)

To calculate the probability of B given A, the algorithm counts the number of cases where A and B occur together and divides it by the number of cases where A occurs alone.

After ODM builds a NB model, the model can be used to make predictions. Application developers can integrate ODM models to classify and predict for a variety of purposes, such as:

- Identify customers likely to purchase a certain product or to respond to a marketing campaign

- Identify which customers most likely to spend greater than $3,000

- Identify customers likely to churn

NB affords fast model building and scoring and can be used for both binary and multi-class classification problems. NB cross-validation, supported as an optional way to run NB, 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. Not having to

hold aside a portion of the data for testing is especially useful if the amount of build data is relatively small.

Decision Trees

Oracle Data Mining supports 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.
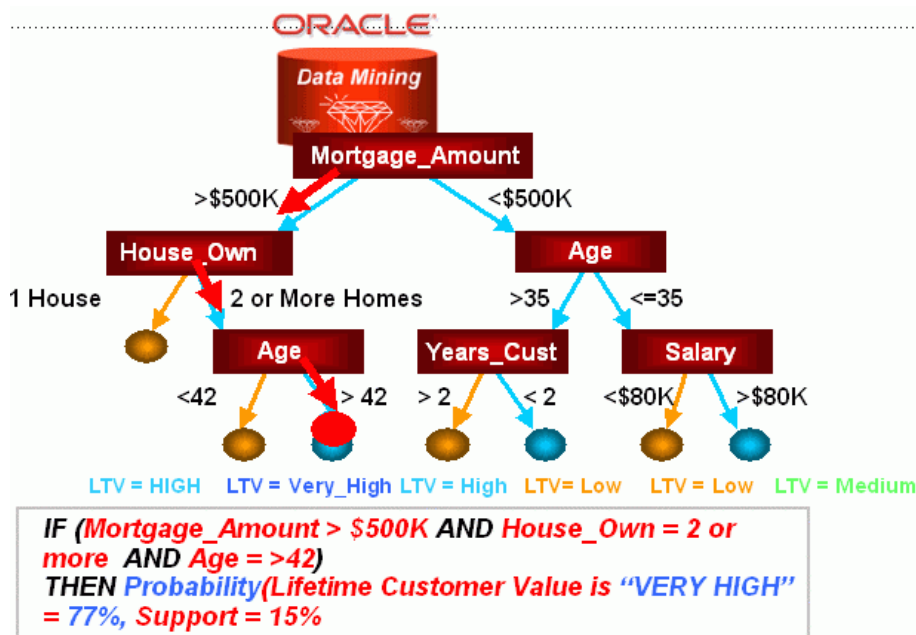
Support Vector Machines

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.

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.

Generalized Linear Models (Logistic and Multiple Regression)

ODM Release 11*g* adds support for the multipurpose classical statistical algorithm, Generalized Linear Models (GLM). ODM supports, as two mining functions: classification (binary Logistic Regression) and regression (Multivariate Linear Regression). GLM is a parametric modeling technique. Parametric models make assumptions about the distribution of the data. When the assumptions are met, parametric models can be more efficient than non-parametric models. Oracle Data Mining's GLM implementation provides extensive model quality diagnostics and predictions with confidence bounds.

Oracle Data Mining supports ridge regression for both regression and classification mining functions. ODM's GLM automatically uses ridge if it detects singularity (exact multicollinearity) in the data. 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 10-30 input attributes.

Attribute Importance

Oracle Data Mining's Attribute Importance algorithm helps to identify the attributes that have the greatest influence on a target attribute. Often, 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 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 finds associations and clusters in the data independent of any a priori defined business objective.

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 (EKM) and O-Cluster algorithms support identifying naturally occurring groupings within the data population. ODM's EKM 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 find new customer segments or to reveal subgroups within a diseased population.

Assocation Rules (Market Basket Analysis)

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?

Anomaly Detection

Release 2 of Oracle Data Mining 10*g* 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 special case of the Support Vector Machines algorithm to create a model of known cases. 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). ODM's anomaly detection algorithm is extremely powerful in finding truly rare occurrences when you have a lot of data but need to find needles in the haystacks.

Feature Extraction

ODM's Nonnegative Matrix Factorization (NMF) is useful for reducing a large dataset into representative attributes. Similar in high level concept to Principal Components Analysis (PCA), but able to handle much larger amounts of attributes and create new features in an additive nature, NMF is a powerful, cutting-edge 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.

Text Mining and Unstructured Data

Oracle Data Mining provides a single unified analytic server platform capable of mining both structured, that is, data organized in rows and columns, and unstructured data. ODM can mine unstructured data, that is, "text" as a text attribute that can be 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, 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.

Oracle Data Mining for Data Analysts

Oracle Data Mining empowers you to analyze your data like never before. With Oracle Data Mining you can build more models faster, and deploy those models more quickly and easily. With Oracle Data Mining, you can mine more sources and types of data, scale to high dimensional and large volumes of data, use state of the art algorithms, and build, evaluate and deploy models immediately throughout the enterprise. Oracle Data Mining provides disruptive technology that changes the way you extract information from data and reduces your dependence on traditional statistical environments. With Oracle Data Mining, you can now mine the data you want to mine.

Oracle Data Mining provides a graphical user interface, Oracle Data Miner, with easy-to-use wizards that guide data analysts through the data exploration, data preparation, data mining, model evaluation, and model scoring process. Intelligent defaults are provided to aide the data analyst to successfully mine their data. Power users may optionally change advanced settings or use the APIs.
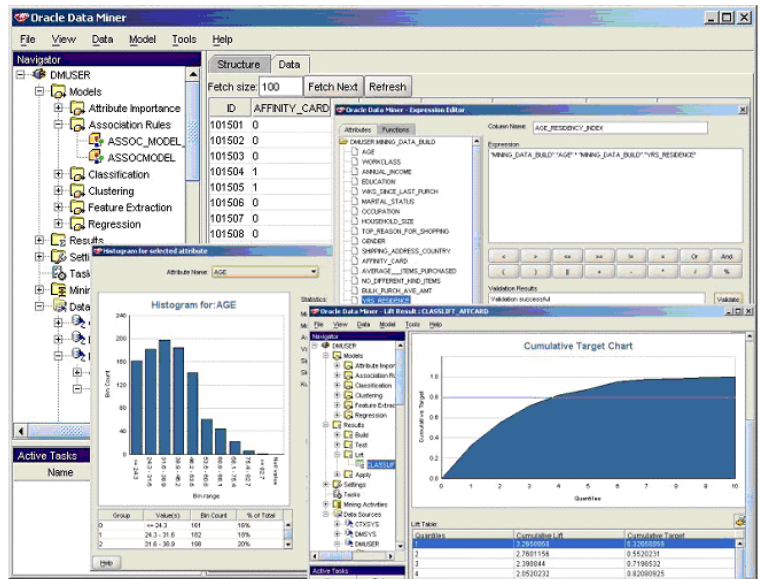
Oracle Data Miner, available for download from the Oracle Technology Network (OTN) guides the data miner through the mining process using wizards.

Oracle Data Miner can automatically generate PL/SQL code for mining activities to transform the data mining steps into an integrated data mining/BI application. SQL Developer and JDeveloper also provide Oracle Data Mining extensions for automatic model code generation.

Oracle Data Mining for Applications Developers

Oracle Data Mining empowers IT departments, and application vendors by making data mining a natural extension of the Oracle Database. Many companies already rely on Oracle Database to store and manage their data. Now, companies can leverage in-database analytics and perform data mining right where the data exists: in the database. In-database mining, as provided by Oracle Data Mining, keeps not only the data in the database where it is safe and secure, but also keeps the data mining models and results there as well, where they are immediately usable in SQL queries and applications. The Oracle Database with the ODM Option, Oracle Business

Intelligence and Oracle Applications Oracle Data Mining provide a complete platform for data management, data analysis and end user applications.

With ODM's SQL and Java APIs, you can integrate data mining into existing business processes, workflows, and applications. Oracle Data Mining provides numerous sample code examples that can be used as starting points for quickly building data mining applications. Additionally, the Oracle Data Miner graphical user interface generates PL/SQL code that helps you to operationalize your mining activities.

```
drop table CLAIMS_SET;
exec dbms_data_mining.drop_model('CLAIMSMODEL');
create table CLAIMS_SET (setting_name varchar2(30), setting_value varchar2(4000));
insert into CLAIMS_SET values ('ALGO_NAME','ALGO_SUPPORT_VECTOR_MACHINES');
insert into CLAIMS_SET values ('PREP_AUTO','ON');
commit;

begin
dbms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION',
 'CLAIMS', 'POLICYNUMBER', null, 'CLAIMS_SET');
end;
/
-- Top 5 most suspicious fraud policy holder claims
select * from
(select POLICYNUMBER, round(prob_fraud*100,2) percent_fraud,
    rank() over (order by prob_fraud desc) rnk from
(select POLICYNUMBER, prediction_probability(CLAIMSMODEL, '0' using *) prob_fraud
from CLAIMS
where PASTNUMBEROFCLAIMS in ('2 to 4', 'more than 4')))
where rnk <= 5
order by percent_fraud desc;
```

Competing on In-Database Analytics

 "Any company can generate simple descriptive statistics about aspects of its business—average revenue per employee, for example, or average order size. But analytics competitors look well beyond basic statistics. These companies use predictive modeling to identify the most profitable customers—plus those with the greatest profit potential and the ones most likely to cancel their accounts"

Havard Business Review, *"Competing on Analytics"*, **by Thomas Davenport.**

Successful companies must go beyond basic BI dashboards and reporting. They must harvest maximal information from their data and leverage it for competitive advantage. Traditionally, this process of extracting information from data has been left to the purview of highly technical data analyst specialists using specialized data analysis tools and extracted copies of data.

Oracle Data Mining removes this barrier, allows data analysts direct access to the data and enables them to more rapidly build, evaluate and deploy predictive analytics throughout the enterprise in dashboards and next-generation applications "*powered* by Oracle Data Mining". With Oracle Data Mining companies can:

- Eliminate data movement and collapse information latency

- Transformation their database repository into an "analytical database"

- Deliver new insights and predictive analytics throughout the enterprise
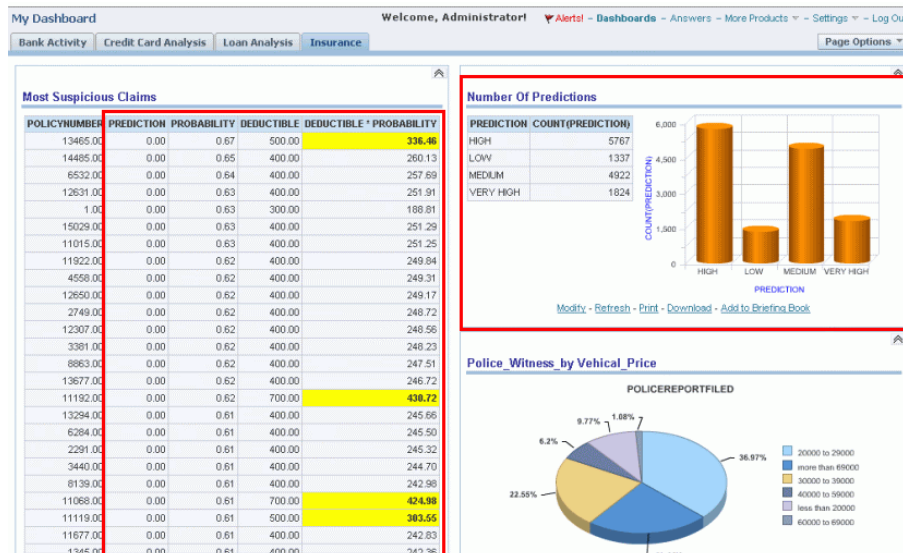
Oracle Data Mining's Predictions & probabilities available in Database for Oracle BI EE and other reporting tools

Beyond a Tool; Enabling Predictive Applications

The true value of data mining is best realized when the new insights and predictions are integrated and operationalized into existing business applications. Oracle Data Mining provides an ideal data and business intelligence IT platform to enable companies to be successful competing on analytics. With Oracle Data Mining users can deploy predictive analytics into business applications, 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 in silos far away from the database on ad hoc data mining projects, with Oracle Data Mining, the Database becomes more than just a data repository—it becomes an *analytical* database and can undergird many new advanced BI use cases.

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.

Automation of data mining tasks is facilitated by Oracle Data Mining's application programming interface (APIs). Oracle Data Mining supports both 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 through model build and to model apply for scoring of single records (on-demand) and batch scoring of large data sets. ODM model predictions may be called on demand as in the following example SQL query and/or stored in relational tables for access by other business applications:

```
SELECT * from(
SELECT A.CUSTOMER_ID, A.AGE,
          MORTGAGE_AMOUNT, PREDICTION_PROBABILITY (LIKELY_RESPOND, 'YES'
          USING A.*) prob
FROM CUSTOMER_DATA.INSUR_CUST_LTV A)
WHERE prob > 0.85;
```

Oracle Data Mining's 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 Business Intelligence EE, Oracle OLAP, Oracle Reports, Oracle Portal, and Oracle Applications.
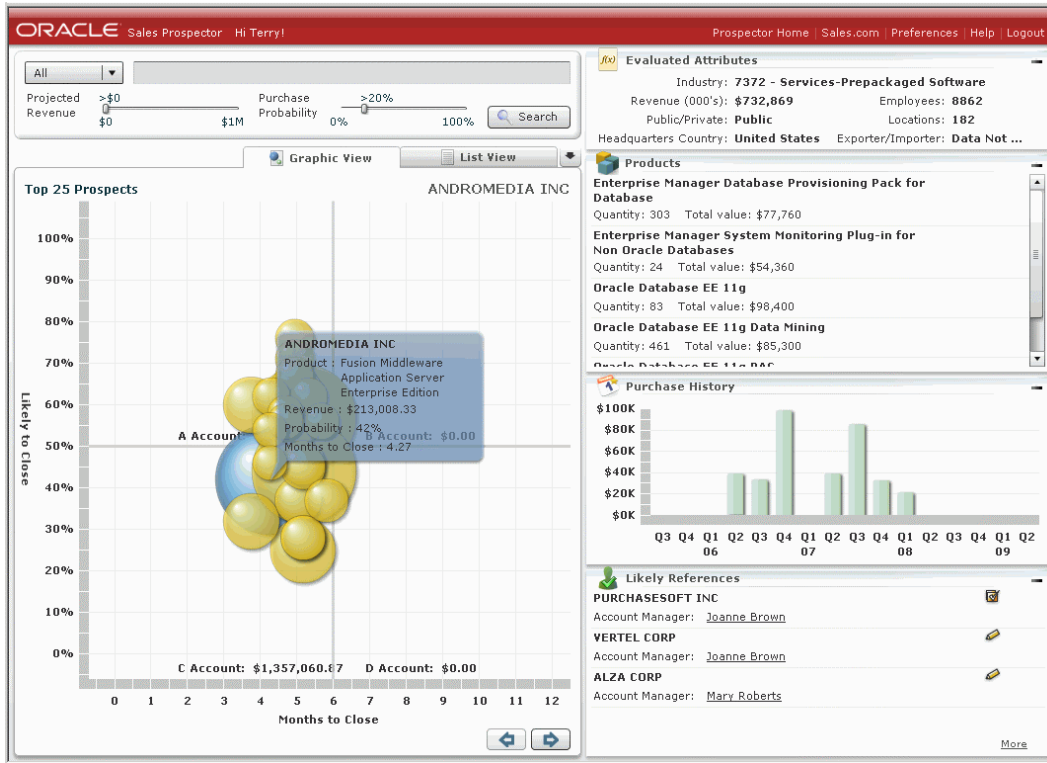
Oracle Data Mining powering the Oracle CRM OnDemand Sales Prospector with embedded and automated predictive analytics.

Oracle now delivers applications and industry data models that have integrated data mining. Oracle's CRM OnDemand Sales Prospector application mines the purchases and demographics of previous customers and then make suggestions to the Sales Rep about which customers in their territory are most likely to purchase what next , the estimated dollar amount, and which customer references are most likely to work best.

Oracle's Industry Data Models delivers a standards-based data model, designed and pre-tuned for Oracle data warehouses. Oracle Retail Data Model combines market-leading retail application knowledge with the power of Oracle's Data Warehouse and Business Intelligence platforms. With pre-built Oracle Data Mining, Oracle OLAP and dimensional models, it delivers industry-specific metrics and insights you can act on immediately. With Oracle Retail Data Model, you can jump-start the design and implementation of a retail data warehouse to quickly achieve a positive ROI for your data warehousing and business intelligence project with a predictable implementation effort.

Oracle Retail Data Models delivers pre-built integrated data mining and standard dashboards and reports from automated data mining.

Oracle Data Mining is part of Oracle's family of business intelligence products and features. 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.

Spend Less

Oracle Data Mining, an Option to the Oracle Database Enterprise Edition (EE) provides a cost effective alternative to expensive tranditional statistical analysis software. 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. Whereas traditional statistical software is rented under an annual usage fee (AUF) pricing scheme, companies can reduce their overall data analysis costs by making the specialized statistical software available for only those individuals who really need it. Production analytical applications can be implemented now inside the Oracle Database for in-database analytics.

Eliminate Redundant Data and Traditional Analytical Servers

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

Additionally, 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.
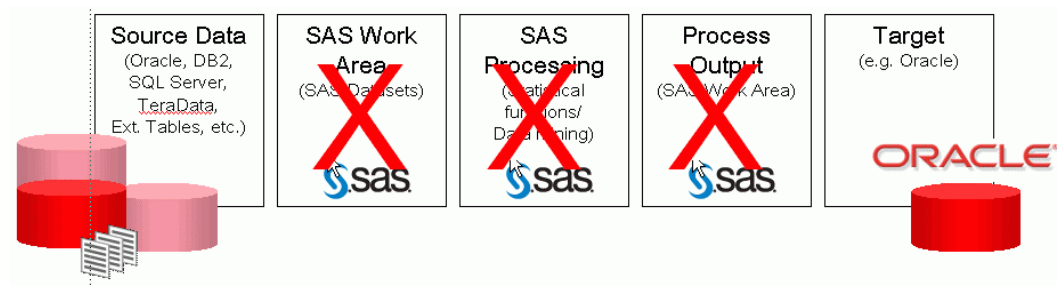
Oracle Data Mining eliminates data movement, data duplication and security exposures.

Conclusion

Oracle Data Mining provides a powerful, scalable in-database data mining engine for data analysts seeking to harvest valuable new information and an industry-standard infrastructure for application developers looking to build applications that automate the discovery and deployment of predictive analytics throughout the enterprise.

Oracle Data Mining's wide range of data mining algorithms, completely embedded in the Oracle Database, solve a wide variety of business problems and provide a powerful infrastructure for building, automating and deploying advanced enterprise business intelligence applications.

By automating, integrating, and operationalizing the discovery and distribution of predictive analytics and new business insights, companies can leverage their Oracle Database technology investment, operate more intelligently, and most importantly, gain competitive advantage.

ORACLE®

Oracle White Paper— Oracle Data Mining 11g:
Competing on In-Database Analytics
June 2009
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
oracle.com

Oracle is committed to developing practices and products that help protect the environment