

### DATA MINING FOR MICROSOFT COMMERCE SERVER

#### The Case for Data Mining

The capacity of digital data storage worldwide has doubled every nine months for at least a decade, twice the rate predicted by *Moore's Law* for the growth of computing power during the same period. This growth disparity is called the *Storage Law* and is a motivating reason for the increasing importance of data mining.

As a result of the Storage Law, our ability to capture and store business data has far outpaced our ability to process and utilize it. This growing challenge has produced massive, ever growing data tombs, or data warehouses that are effectively write-only; data is deposited to merely rest in peace, since in all likelihood it will never be accessed again. This situation is a significant waste of potential business intelligence (BI). Data warehouses represent the data assets of a business enterprise, and to leave this information untapped is a serious lapse in productivity.

Deployments of Microsoft Commerce Server serve as a good example of how companies overlook the valuable nature of data assets. An integral part of Commerce Server is the data warehouse that parallels the transactional database upon which all operational use of the e-commerce site is based. Microsoft provides basic OLAP functionality as part of the base product, but true data mining capabilities are not fully exploited.



The lack of data mining means that valuable business intelligence is not available for the typical Commerce Server implementation. Furthermore, significant knowledge about customer buying patterns, association rules for shopping cart analysis, and classification rules remain unexplored and unutilized.

Recognizing these deficiencies, AMULET Development Corp. offers specialized data mining services for new and existing Microsoft Commerce Server deployments.

- ♦ Fully utilize existing Microsoft Decision Tree and Clustering algorithms.
- ♦ Develop and integrate custom data mining algorithms.
- ♦ Provide professional data mining consulting services.

#### Data Mining

Data mining is defined as the identification of interesting structure in data. Structure designates patterns, statistical or predictive models of the data, and relationships among parts of the data. It is the predictive characteristic of data mining that separates it from OLAP. OLAP works views the data warehouse to yield analysis of historical data. Data mining takes an important step further by providing predictive capabilities.

AMULET fully deploys the out-of-the-box data mining functionality included with Commerce Server, specifically the Predictor Resource. The Predictor includes two data mining models: the predictive model, and the segment model. Here are some examples of the business intelligence you can achieve:

- What movies do people like to purchase together?
- What are the movies young female customers like to buy?
- What other movies should be recommended to a customer that bought "Star Wars?"

The most important facet of AMULET's data mining services is our ability to design, develop, and deploy custom data mining algorithms to suit your specific BI needs. Through years of research experience, AMULET staff is well versed in the cross discipline nature of data mining in fields such as mathematical statistics, computer science, and visualization. Here is a list of some of our technologies:

- **Decision Trees** – the Decision Trees algorithm is a classification type of algorithm that works well for predictive modeling. Using the algorithm, we can predict both discrete and continuous attributes.

Example: Decision Trees can predict whether a consumer will purchase a computer. Given data attributes *age*, *income*, *student*, and *credit rating*, the algorithm uses probability estimation to classify consumers in the data sample as likely to either buy a computer or not buy a computer.

- Clustering** – The Clustering algorithm uses iterative techniques to group records from a dataset into clusters containing similar characteristics. Using these clusters, you can explore the data, learning more about the relationships that exist, which may not be easy to derive logically through casual observation. Additionally, you can create predictions from the clustering model created by the algorithm.

Example: Consider a group of people who live in the same neighborhood, drive the same kind of car, eat the same kind of food, and buy a similar version of a product. This is a cluster of data. Another cluster may include people who go to the same restaurants, have similar salaries, and vacation twice a year outside the country. Observing how these clusters are distributed, we can better understand how the records in a dataset interact, as well as how that interaction affects the outcome of a predicted attribute.
- Naïve Bayes Classification** – The Naïve Bayes Classification algorithm quickly builds mining models that can be used for classification and prediction. It calculates probabilities for each possible state of the input attribute, given each state of the predictable attribute, which can later be used to predict an outcome of the predicted attribute based on the known input attributes.

Example: The Naïve Bayes algorithm is similar to the Decision Tree classification algorithm in that it can predict whether a consumer will purchase a computer. Naïve Bayes Classification is characterized with high accuracy and speed when applied to large databases.
- Sequence Clustering** – The Sequence Clustering algorithm analyzes sequence-oriented data that contains discrete-valued series. Usually the sequence attribute in the series holds a set of events with a specific order (such as a click path). By analyzing the transition between states of the sequence, the algorithm can predict future states in related sequences. The Sequence Clustering algorithm is a hybrid of sequence and clustering algorithms.

Example: A typical usage scenario for this algorithm is Web customer analysis for a portal site. A portal Web site has a set of affiliated domains such as News, Weather, Money, Mail, and Sports. Each Web customer is associated with a sequence of Web clicks on these domains. The Sequence Clustering algorithm can group these Web customers into more-or-less homogenous groups based on their navigational patterns. These groups can then be visualized, providing a detailed understanding of how customers are using the site.
- Time Series** – The Time Series algorithm creates models that can be used to predict continuous variables over time from both OLAP and relational data sources. Using the algorithm, we can choose one or more variables to predict, but they must be continuous. The case series identifies the location in a series, such as the date when looking at sales over a length of several months or years.

Example: We can use the Time Series algorithm to predict sales and profits based on the historical data in a cube. In addition, using cross-variable correlations in the model's predictions, prior sales at one store may be useful in predicting current sales at another store.
- Association** – The Association algorithm is specifically designed for use in market basket analyses. The algorithm considers each attribute/value pair (such as product/computer monitor) as an item. An itemset is a combination of items in a single transaction. The algorithm scans through the dataset trying to find itemsets that tend to appear in many transactions.

Example: The Association algorithm is useful for cross sell or collaborative filtering. For example, we can use an association model to predict items a customer may want to purchase based on other items in their basket.

## Who is AMULET Development Corp?

AMULET Development Corp. is a Web database and e-commerce integration firm founded in 1984 to provide quality technology solutions for businesses in a broad range of industries. Specializing in Microsoft server, Web, and database technologies, we've built many high profile e-business websites. Our current focus is business intelligence, analytics, and data mining using contemporary technology to help enterprises better utilize valuable data assets.

## For information about AMULET's Data Mining for Commerce Server services contact:

Sonya Franklin 1-877-722-7393  
 info@amuletc.com  
 www.amuletc.com

AMULET maintains a Microsoft  
 Competency in *Data Management*  
 Solutions with a *Business Intelligence*  
 (BI) specialization.

