The adversarial nature of the American legal system creates a natural conflict between how facts may be viewed by the various litigating parties. Data and information that the plaintiff claims is completely reliable is resoundingly criticized by the defendant as misleading. Mark Twain wrote in his autobiography:

>“Figures often beguile me, particularly when I have the arranging of them myself; in which case the remark attributed to Disraeli would often apply with justice and force: There are three kinds of lies: lies, damned lies, and statistics.”

In today’s business world, the software systems and platforms that drive commerce are heavily reliant on vast amounts of data and databases summarizing historical transactions and events. Depending upon the case facts, the amount of data can be staggeringly large and voluminous. The facts, the story and the evolution of events get lost in the details of the vast forest of data. In fact, the flow of transactions can be increasingly difficult to trace as certain databases only contain specific information or only a partial view of the entire transaction at large. As a general rule, when large amounts of data are involved, it is never simple and straightforward.

Understanding several characteristics about data mining can be helpful in identifying those opportunities where it can best be used in a litigation matter involving economic damages and/or identifying whether a company has complied with polices and procedures. What makes data mining useful is in those situations is where the “Four Vs” exist:

- The sheer volume of data that is available, produced, aggregated or requires analysis is large and covers an extended period of time (e.g., daily, monthly, quarterly or multiple years).
The velocity of data that is created and available to analyze (i.e., where data is available in real time).

The variety of data has become quite broad. (e.g., where data is available on customer purchase histories, unique impressions, locations bought and bundled purchase metrics)

The veracity of data can be matched. (e.g., a retailer can match and observe what customers really buy compared to what they respond to as buying in a consumer survey.

Where electronic processing of transactions takes place, information in databases are generally encrypted, making identification of payment information by a specific individual challenging. The resolution is often the process of cross matching information from separate databases through a common point of origin to fully pull the complete transaction together for further analysis or summary. This “Data Mining” process is not often obvious due to the nature of the databases and the vast volume of data.

The use of data mining techniques can be extremely powerful in proving, disproving or bringing to light the true facts. In many litigation cases, there is no option other than to apply various levels of data mining to sift through the large volumes of information. At Barnes & Company, we have applied data mining techniques to the following types of legal claims:

**Types of Damage Claims Where Data Mining Can Be Applied**

- Lost Profits
- Business Interruption
- Compliance with Policies & Procedures
- Breach of Contract
- Antitrust
- Bankruptcy
- Patent Infringement
- Fraud
- Lost Value

Certain industries are prone to high volume of transactions that allow for the application of data mining techniques, including consumer banking, insurance, brokerage & securities, telecommunications, retail, healthcare, oil and gas, manufacturing, utilities and transportation.
The Data Mining Process & Analytical Flow

Data mining involves collecting, processing, storing and analyzing data in order to identify and extract new information from it. From a commercial perspective, a large number of industries and companies use data mining internally to improve their knowledge of their customers and to improve their approach to marketing and delivering services to the end customers. In each of these instances a company is seeking to collect data to identify patterns and trends that then can be applied to improving profitability and specifically targeting customers. Data mining models can be applied to specific scenarios, such as:

- **Forecasting**: Estimating sales and estimating seasonality of sales
- **Risk and probability**: Identifying the best customers to target for mailings, assigning probabilities to specific outcomes
- **Recommendations**: Identifying which products are likely to be sold together
- **Finding sequences**: Analyzing customer purchasing habits and trends, predicting other customer behavioral events
- **Grouping**: Separating customers or events into clusters of related items

Internal data mining models used by a company to project and budget sales and operating levels may be highly useful to analyze lost profits or a business interruption loss. One of the benefits of these internal data mining models is they have been used specifically by company management to make business decisions and to project economic outcomes. This approach can be more reliable than creating damage models specifically for litigation, and for which may not incorporate historical data patterns and experience.
Data mining, like any analytical process, has a common methodology of application.

1. Defining the Problem
2. Preparing Data
3. Exploring Data
4. Building Models
5. Exploring and Validating Models
6. Updating Models

**Defining the Problem**

The first step in any analytical process is to identify the problem or questions to be answered. This step includes analyzing business requirements, defining the scope of the problem, defining the metrics by which the model will be evaluated, and defining specific objectives for the data-mining scope of work. This process may translate into identifying the following types of questions:

- What data or information are you looking for? What types of relationships or correlations are to be identified?
- Does the problem you are trying to solve reflect the policies or operational performance of the business?
- Do you want to make predictions of data, or just identify patterns and associations?
- Which outcome or attribute do you want to try to predict?
- What kind of data is available? Does the process of identifying data require any cleansing, aggregation, or multi-database processing to make the data usable?
- Is the data seasonal? Does the data accurately represent the processes of the business?

Addressing these questions may require the making inquiry or specific document discover of the type of data available.
Preparing Data

Data can be dispersed across a company and stored in different formats, or may contain inconsistencies such as incorrect or missing entries. Data may also be encrypted, thereby limiting the type of analysis or extraction that can be performed. Accordingly, a data cleansing process may be required as the first step to accumulating the data to be analyzed. Data cleaning is not just about removing bad data or inserting missing values, but about finding hidden correlations in the data, identifying sources of data that are the most accurate. Incomplete data, wrong data, and inputs that appear separate, but in fact, are strongly correlated all can influence the results of the analytical model being developed.

In most data-mining projects, the amount of data and size of the databases are extremely large and may not allow the examination of every transaction for data quality. In this case, the use of data profiling and automated data cleansing and filtering tools may be required, such as applying, Microsoft SQL Server Master Data Services or SQL Server Data Quality Services to explore the data and find the inconsistencies.

Exploring Data

Understanding the data, any cross relationships between the data and the type of data that can be extracted allows for proper decisions to be made when creating the data-mining models. Certain exploration techniques, such as, identifying if the data sets contain the specific customers, transactions and information required for answering the questions and queries to be performed. For example, is the data representative of the customers or business processes to be analyzed? Applying certain statistical analyses such as assessing the standard deviations of data samples and other distribution values can provide useful information about the stability and accuracy of the results. Data that strongly deviates from a standard distribution might be skewed, or may not represent an accurate picture of a problem.

By exploring the data in conjunction with the overall understanding of the business problem, can identify if the dataset contains flawed data and the proper strategy for fixing the problem.
Building Models

The actual data-mining model is linked to the source of data, but does not actually contain any data until it is processed and the application of data extraction begins. Depending upon the complexity of the data extracting and data mining, the results can be analyzed and filtered and cross checked with the source data to ensure that formulas and algorithms are properly pulling the correct data.

Exploring and Validating Models

Any data-mining model must be tested to identify how well the model performs. The first step is identifying if the results provide answers to the questions originally developed during the establishment of the scope of the project. Are the patterns and trends originally expected matching the underlying data? If the data-mining model developed is not providing clarity, or is inconsistent with expectations, then redefining the problem or reinvestigating the data in the original datasets is required. The answer may also indicate that the legal claim being alleged may not be accurate.

Deploying and Updating Models

Once the data-mining models are deployed and extracting the desired underlying data, they may require constant updating as additional data is made available.
Final Considerations in the Possible Application of Data Mining in Litigation Matters

The potential application of data mining techniques in litigation is as vast as the volumes of data that may be available. Its real value is the ability to sift through large volumes of information and extract the most useful parts related to the damage issues and causes of action being alleged. In a recent case, the results of analyzing over two million lines of data was reduced to a single page of analysis to present the findings.

* * * * * * * *

Scott A. Barnes, CPA, CFF, CGMA is the Chief Executive Officer of Barnes & Company who during the past 25 years has assisted corporations, investors and attorneys in assessing the financial issues surrounding complex and sophisticated accounting issues, merger and acquisition transactions, business valuation and industry and competitor analyses. He has extensive experience providing sophisticated commercial damage analyses and expert testimony for clients in a diverse range of industries, especially with respect to forensic analyses, intellectual property infringement, antitrust, lost profits, business interruption, contract disputes, business valuation, fraud, accounting, accounting malpractice and other complex commercial damage issues.

He has served on two editorial boards of national publications specializing in economic damages, business valuation and forensic accounting issues and published numerous articles on the subject of economic damages, business valuation and professional standards in the performance of litigation services. He can be reached at sbarnes@barnesco.com