

Risk Insights

Data Mining



Data mining enables organizations to gather greater intelligence

DATA MINING UNCOVERS anomalies, exceptions, patterns, irregularities or trends that might otherwise remain undetected amidst immense volumes of data. That enhanced awareness of what is occurring within an organization provides managers with greater intelligence, intelligence needed to make better decisions.

Data mining techniques are used to extract specific, relevant bits of information from immense data sets, based on user-defined queries and parameters. These data definitions may encompass date ranges, alpha text ranges, dollar amounts, system master files, IT system logs or other criteria. Extracted data is then processed by running scripts to match or analyze the data and the results displayed in tables, graphs, charts and other report formats.

Those reports allow for deeper and broader assessments than what might be attained through other means of manual evaluation. A typical audit, for example, will not examine transactions below an immaterial dollar value due to volume of the data, time, cost and other concerns. Data mining, however, makes it feasible to examine these lower value transactions and a much larger volume.

Such examination might reveal anomalies in numerous low value transactions that have not been previously evaluated due to cost-benefit constraints. The examination might also uncover an unusually large number of transactions in a particular value range that is under an approval threshold. For example, when a supervisor signature is required for transactions over \$5,000, we may find large numbers of similar transactions with a value in the \$4,990 to \$4,999 range. This could indicate that someone is splitting transactions to avoid the approval process. It is common for fraud to transpire in small transaction values assuming that the disbursements will not attract an auditor's attention.

Data mining can be applied toward virtually any collection of information stored in electronic format, including:

- Master files
- ERP system files
- Specific application or module files
- Database files
- Email files

Those various data sets may be active, archived, structured, non-structured or classified in some other manner. When examined through data mining, such information provides insight for more fully understanding the activities and processes that those data entries represent. That enables managers to recognize the impact those daily events have on overall organizational performance.

Risks Insights: Data Mining

General Benefits Associated with Data Mining

FOR VARIOUS AUDITS AND EVALUATIONS, organizations typically base conclusions upon findings derived from a data sample taken from a large data set. Findings from that sample are then extrapolated to represent the full data population. Selecting a representative sample requires a reliable methodology; however, anomalies and exceptions not present in that sample may still go unrepresented. Such transactions may not be detected elsewhere.

Data mining mitigates those risks by examining an entire data population. The resource requirements and potential for inconsistent evaluations that accompany manual data reviews are eliminated.

Improvements in data mining functionality make it easier to access and read targeted information. Data mining users benefit, too, from advances that make technology in general more powerful and affordable.

Various Internal Uses for Data Mining

DATA MINING SUPPORTS a variety of internal evaluation functions, including financial auditing, internal auditing, data integrity monitoring, IT control and system log reviews, and fraud detection. Data mining findings also illustrate opportunities for process improvement and greater efficiency.

Internal or external financial audits require tracing financial totals back to original transactions, verifying those totals through various forms of reconciliation and identifying any adjustments or exceptions requiring further investigation.

Data mining makes it much easier to follow audit trails back to supporting transaction details. It also enables auditors to locate and compare corresponding forms of documentation, such as purchase orders and invoices.

Exceptions or adjustments require scrutiny because they indicate that improper activity may have occurred. Data mining techniques are used to identify transactions that require further examination. It is intended that each item that is identified as an exception would be evaluated for appropriateness.

For example, a receivable balance that was written off may include multiple transactions that sum to an amount that exceeds a dollar threshold for approval. Data mining identifies these transactions and groups them as exceptions requiring further examination. Some records may have been modified during non-work hours or close to a reporting period closing. Data mining calls out those and other adjustments that may merit additional evaluation.

It may be determined that these items are appropriate. Such items are often referred to as false positive exceptions. The expectation is that the data mining approach would allow for a large amount of data to be analyzed and the data set be narrowed to only those transactions that would require further examination.

Data Mining Monitors Data Integrity

DATA MINING ALSO makes it easier to evaluate data integrity throughout an organization. Data mining reports based upon date ranges may display outdated records containing private information. Proper disposal of those records then lowers data storage costs and the potential liability associated with retaining such information.

At the data entry level, customer address fields may contain five-digit zip codes, nine-digit zip codes and zip codes with transposed digits. Data mining identifies such inconsistencies and inaccuracies, enabling managers to take corrective measures.

Two records may exist for the same employee – one with her married name and one with her maiden name. Multiple vendor records may appear for a business that relocated. Data mining uncovers such instances of duplicate records for the same activity, individual or entity.

Data Mining Allows More Complete, Regular Monitoring of IT System Activity

IT SYSTEM LOGS REQUIRE regular reviews to detect potential incidents and to monitor the effectiveness of preventative controls. A data mining query may quickly reveal instances where someone tried to access a secure server with a valid username but a series of incorrect passwords. That could indicate unauthorized access attempts by someone who obtained a valid username.

An individual may have needed access to crucial files only for a short-term project. Another employee may be on extended leave. By running scripts on an IT access log, data mining lets managers know whether those and other employees were properly denied or granted access, in accordance with established user provisioning policies.

Unauthorized changes can cause system crashes and increase risk levels for all IT-related threats. An administrator may easily view a critical file's most recent modification date, but evidence of previous alterations may be less visible. Data mining compares current file characteristics against baseline attributes defined when the last approved change was made and identifies all subsequent modifications.

Data Mining Detects Potential Incidents of Fraud

FRAUD IS A CONTINUAL THREAT, and data mining reveals incidents of possible improper activity throughout an organization. Multiple postings may exist for one transaction, indicating that someone may have split a large fraudulent transaction to elude detection.

An employee may have entered numerous disbursements for a fictitious vendor, disbursements whose check numbers are too close in sequence to be normal recurring monthly payments. The same user name may appear for incompatible IT functions, such as payroll and accounts payable processing. Information extracted via data mining calls out such suspicious circumstances.

Fraudulent schemes increasingly involve unauthorized uses of nonpublic information. Data mining uncovers clues that such schemes may be occurring. A data mining search may detail that an employee copied a database column containing customer credit card numbers.

Another search of logged user activity may indicate that an employee spent an inordinate amount of time viewing human resources files containing social security numbers. Both searches present findings requiring further investigation and illustrate preventative control vulnerabilities.

Data Mining Identifies Opportunities for Operational Improvement

WHILE DATA MINING uncovers potential fraud schemes, it also helps organizations identify inefficiencies and opportunities for process improvement. An organization may have multiple offices using the same office supply vendor. Data mining shows, however, that each office has its own vendor account, with separate monthly payments made for each account. Merging those accounts into a centralized account then simplifies record keeping and lowers disbursement costs.

Another organization may have two production facilities, one of which it purchased from a different company. Although production output measures for those two sites are similar, considerable disparities exist in the inventory data they provide. A legacy system's constraints at one facility may be prompting individuals to work around defined processes. Managers at the two sites may also be accustomed to using differing methodologies – rather than uniform procedures – to account for discounts, returns and unit costs.

Data mining provides the capability to thoroughly examine all related available information and then investigate the mostly likely cause of those disparities. When applied to functions throughout an organization, data mining provides similar knowledge needed to assess performance concerns and make necessary improvements.

Limitations of Data Mining

WHILE DATA MINING can be immensely useful, it carries limitations. Data mining only identifies anomalies, exceptions, irregularities, patterns or trends; it does not assign meaning or value to that information. That may result in false positive findings, findings that prompt initial concern, but for which there are sound explanations.

“Managers need to remember that **data mining** is only **an aid** for decision making and not a substitute for **sound judgment.**”

A vendor file review may highlight two different companies with the same street address and phone number. That could signify fraud. An ensuing examination, however, shows that the first company listed completed a rebranding effort during that reporting period and took on the name of that second company.

Data mining activities can also have an adverse impact on IT operations and data integrity. Data mining requires considerable IT processing power, especially if it involves complex queries that examine very large data sets. Individuals working on other tasks may encounter delays and other difficulties if data mining is being conducted on a production server or other system they are using. For regular data mining examinations, a separate, dedicated server should be used as a data warehouse or information repository.

Using live files for data mining also subjects the underlying data to data manipulation. If that information must be mined, only copies of live files should be used.

In addition to recognizing the potential for false positive findings and the need to segregate data mining from live IT activities, managers also need to remember that data mining is only an aid for decision making and not a substitute for sound judgment.

Data Mining Sustains Focus on Critical Information

EFFECTIVELY MANAGING EVER-INCREASING volumes of data is a critical concern. An even greater concern is extracting meaning and insight from all of that information.

Data mining provides the capability to do that. That enables organizations to more efficiently assess the completeness and accuracy of data, detect anomalies or exceptions, recognize irregularities, patterns or trends, and identify needs or opportunities for improvement.

By giving managers a powerful tool for making better decisions, data mining enables an organization to more fully capitalize on the benefits afforded by information technology.

CONTACT US

Alyssa G. Martin, CPA, MBA
Partner-in-Charge
Risk Advisory Services
alyssa.martin@weaver.com

Weaver's risk advisory services are strategic, executable and measurable—and our nimble process is designed to help companies remain optimally functional as they identify and manage risk. We work closely with our clients to customize services that fit their existing staff structure and operations. Integral to this sensitive work, we believe our communication skills are as valuable as our technical knowledge and professional insight. You will experience thoughtful, purposeful communication throughout the process. Specific services we provide include:

- Business continuity planning
- Business process improvement
- Contract monitoring and compliance
- Enterprise risk management
- Internal audit
- Internal control evaluation
- Integrated financial and IT audit
- Performance audit and measurement
- Regulatory compliance
- Risk assessment
- Sarbanes-Oxley compliance

Disclaimer: This content is general in nature and is not intended to serve as accounting, legal or other professional services advice. Weaver assumes no responsibility for the reader's reliance on this information. Before implementing any of the ideas contained in this publication, readers should consult with a professional advisor to determine whether the ideas apply to their unique circumstances.

© Copyright 2014, Weaver and Tidwell, L.L.P.

