

Democratizing Data Science and Predictive Analytics

New research identifies business initiatives to extend the benefits of data mining, sciences, machine learning, and predictive analytics, and the challenges that companies are facing with these plans.

BUSINESS INSIGHTS AND DECISION-MAKING have been supercharged in recent years by the combination of huge data collections and sophisticated “big data” analytics tools. All too often, however, cutting-edge data mining and business intelligence (BI) capabilities remain sequestered within organizations.

Many companies have found these data analysis tools too costly, too complex, and/or too specialized to distribute to all of the potential users throughout their various business units. Instead, data mining and BI have been restricted to small teams of business analysts and finance professionals, who use the tools for broad strategic purposes that often overlook more tactical uses and needs.

There are signs that this segregation is changing. A new survey into data mining practices and requirements has found that 92% of respondents want to deploy advanced analytics more broadly across their organizations.

Along with this desire to “democratize” data mining and BI, the IDG Research survey of nearly 70 IT and business executives and managers paints a picture of today’s data analytics landscape. The survey documents the diversity and complexity of this environment, as well as the many benefits that companies are realizing through the use of powerful data mining solutions.

Data mining drivers and objectives

Digital data is the lifeblood of modern organizations, so there’s little surprise that companies want to leverage this asset to the maximum degree possible. Data mining, which analyzes large amounts of data in an automated fashion to discover hidden patterns and predictive information, is a critical tool for supporting this effort.

Most of the survey respondents place a high value on data mining and its benefits (see Figure 1).

Yet, despite the many positive results they’ve experienced, many organizations have unresolved data mining needs. Top among them is the need to deploy advanced analytics capabilities cross-company, an objective cited by 92% of the respondents.

This near-universal goal is broadly unmet,

in part, because 78% of respondents say they are struggling to find the right data mining strategy or solution. Much of the difficulty in doing so stems from the wide range of data mining and BI tools available, and the many use cases to which they can be applied.

The complex data mining landscape

The data mining field is nothing if not diverse. Survey respondents identify nearly 30 different data mining use cases, with the average respondent selecting about 10 that were relevant to them. In fact, about half of the respondents identified five use cases:

- Analyze risk/risk management (cited by 51%)
- Cybersecurity (event detection, analysis, remediation) (49%)
- Forecast demand for products or services (49%)
- Predict customer behavior (49%)
- Optimize software (47%)

To tackle this range of use cases, organizations are leveraging (or planning to leverage) a wide variety—and many combinations—of data mining tools and algorithms.

Figure 1. Top 10 Benefits of Data Mining



Source: IDG Research Services, August 2017

A new survey into data mining practices and requirements has found that 92% of respondents want to deploy advanced analytics more broadly across their organizations.

Fortunately, new-generation tools designed to be affordable and easy to learn and use are addressing many of the complexities and costs long associated with data mining solutions.

The average company expects to leverage three different tools in the coming year. Among those tools are:

- Cloud-based tools such as AWS and Azure machine learning services (79%)
- Open source tools such as R and Python (65%)
- Productivity applications such as Microsoft Excel (63%)
- Proprietary tools such as those from SAS and Matlab (62%)
- BI/data visualization tools (60%)

Within these tools, organizations will be running more than 20 data mining algorithms. Half or more of the survey respondents say they would use four of the many available algorithms in the coming year: cluster analysis, decision trees, anomaly detection, and association rules.

Data mining challenges and needs

The diversity and complexity of data mining tools and algorithms are the root causes of some key challenges organizations face. For example, 38% of the respondents say data mining tools are not intuitive or conducive to self-service, and 31% say they lack the skill sets needed to leverage the tools.

As shown in Figure 2, companies have to clear plenty of other data mining hurdles. These range from security concerns to scalability requirements to difficulty integrating tools with other applications.

Fortunately, new-generation tools designed to

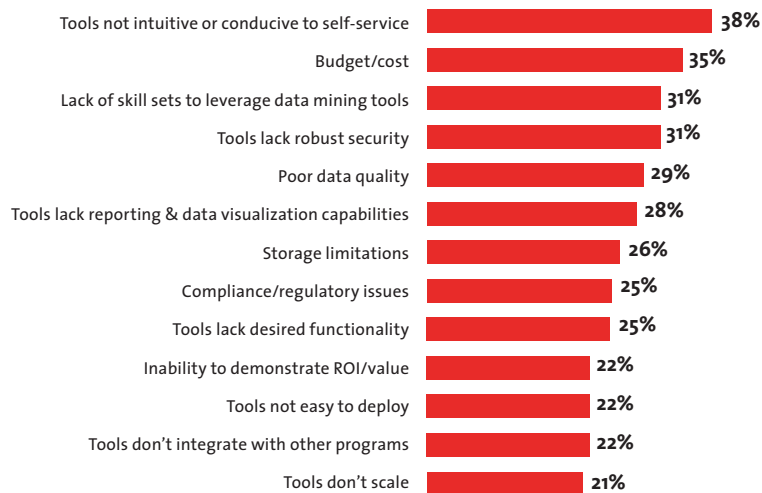
be affordable and easy to learn and use are addressing many of the complexities and costs long associated with data mining solutions.

Democratizing data mining throughout organizations

As companies seek to capitalize on the many benefits that broadly distributed data mining can deliver, they must follow several steps:

1. **Identify the core business needs** and use cases—tactical as well as strategic—that data mining can address.
2. **Identify and evaluate the sources of data** against which the data mining tools will execute, ensuring the data’s currency, accuracy, and relevance.
3. **Determine the applications**, including BI systems, with which the data mining tools must interoperate.
4. **Identify which of the available data mining solutions** address the full scope of their requirements, including their budget constraints and the breadth and technical savvy (or lack thereof) of their target user bases.
5. **Standardize a data mining tool** that can be deployed and used organization-wide. Ideally, this tool will meet the demands of the data scientists who build models and push information out to others, as well as the requirements of business users who can visualize and apply the information in ways specific to their needs.

Figure 2. Top Challenges to the Adoption of Data Mining Tools



Source: IDG Research Services, August 2017

Discover how **MicroStrategy** makes it easy to deploy data mining, predictive analytics and machine learning applications.

BI analytics and data mining vendor MicroStrategy has been at the forefront of modern analytics solutions that marry high performance and broad functionality with ease of use and cost efficiency. Data Mining Services are fully integrated into the MicroStrategy analytics platform, permitting users to access, design, visualize, and publish data-rich data mining reports through the web, email, portals, Excel, mobile, and other interfaces.

The MicroStrategy BI and data mining platform can support hundreds of thousands of users and handle databases of any size. Among other features, the platform supports more than 400 analytical functions out of the box and complies with Predictive Model Markup Language (PMML), the industry-standard language for data mining models. The platform also easily integrates with the most popular data mining solutions including R, SPSS, and others.