Planning Successful Data Mining Projects

A practical, three-step guide to planning your first data mining project and selling it internally
Table of contents

Executive summary............................................................................................................... 3
1. Start with a strategic end in mind .................................................................................... 4
2. Line up essential resources .............................................................................................. 7
3. Define an executable data mining strategy .................................................................... 11
Next steps .......................................................................................................................... 14
About Clementine® for data mining .................................................................................... 15
About SPSS Inc................................................................................................................... 15
Executive summary
Predictive analytics solutions deliver some of the most impressive and quantifiable returns of any technology investment. To realize these results, organizations often start by developing the business case required to get their first data mining project off the ground and prove the value of data mining to their organization. To ensure success, the initial planning phase is the most important phase of any data mining project—from understanding project objectives from a business perspective, to converting this knowledge into a data mining strategy and preliminary project plan.

Planning Successful Data Mining Projects is a practical, three-step guide for planning successful first data mining projects and selling their business value within organizations of any size. It’s designed to help project leaders work around common data mining obstacles to enable rapid, business-focused predictive modeling. The following steps are based on the best practices of leading organizations and lessons learned from applications of the de facto standard data mining methodology, CRISP-DM, for nearly a decade.

Plan for data mining success by following these three steps:
1. **Start with a strategic end in mind**—avoid the “ad hoc trap” by focusing data mining on a strategic objective
2. **Line up essential resources**—sell your business case internally to get the commitments you need
3. **Define an executable data mining strategy**—plan how to achieve your objective by capitalizing on resources

**Apply these steps to realize the promise of predictive analytics**
Data mining and predictive analytics are proven technologies that have become an integral part of the daily operations of leading organizations—from the FORTUNE 500 to government agencies and academic institutions. Data mining is the process of uncovering patterns in data using predictive techniques. These patterns improve decision making by providing the actionable insight needed to improve business processes—helping organizations achieve critical business goals. Predictive analytics combines data mining with decision optimization technologies—further automating and applying the results of this advanced analysis to everyday decisions.

Several years ago, MIT Technology Review hailed data mining as one of ten “emerging technologies that will change the world.” Today, it is changing the world—organizations in nearly every industry have built data mining into their day-to-day operations to address strategic objectives.

Data mining is improving the decisions related to key initiatives such as:
- Customer relationship management
- Fraud detection and prevention
- Risk management
- Manufacturing quality improvement
- Healthcare quality improvement
- Homeland security

Follow the three steps above to plan a successful first data mining project and realize the benefits of predictive analytics in your organization.
1. Start with a strategic end in mind

Leadership guru Stephen Covey’s maxim, “begin with the end in mind,” is directly applicable to leading change with data mining in your organization. Successful data mining is a business process focused on business goals. But Covey’s maxim should be applied with one caveat—the end must be strategic.

Strategic context is critical to maximizing the value of data mining and avoiding the “ad hoc trap”—resources and time are wasted when data mining is executed with no clear business focus. Before beginning a data mining project, define the business goal that your project will address, and the benefits you hope to gain. Beginning with a strategic end in mind helps to ensure that your first data mining project starts off in the right direction—and with broad support.

**Step one goal-focusing activities:**
- Strategic objective definition
- Related business goal identification
- Clear success metric description
- Cost-benefit analysis

Define your strategic focus

At the start of your project, review the basic information that is known about your organization’s business situation and strategic issues. These details help identify the business goals to be achieved, key project stakeholders, and solutions currently in place. For example, many companies already have some form of basic reporting in place to measure current progress against their strategy, but may not have the analytical sophistication needed to identify ways to improve future progress.

The most successful data mining initiatives are driven by strategic business goals which, by definition, already have executive commitment. Organizations committed to strategic business goals are the most open to the change required for data mining initiatives to be successful—and the most ready to allocate essential resources. Most often, in organizations where data mining has really taken root, the first data mining project informed decision makers on an important strategic issue. Decision makers tend to view data mining most favorably when they can take immediate action based on the results. Throughout the planning phase, keep your emphasis on delivering actionable results that apply to a strategic objective, not on technical issues such as predictive accuracy or technique selection.

Successful data mining initiatives often start with a narrow focus, addressing the most critical organizational issue, such as retaining customers longer, and then address other issues once they’ve proven the value of data mining. If you don’t already have a strategic end in mind, review your organization’s strategic plan. Is there an area in which you aren’t making the progress you hoped for? What insight is needed to improve decision making and get things back on track?

Choose a strategic goal that is keeping people up at night. Who has a stake in achieving the goal? Which executive is responsible for it, and is probably losing sleep? Perhaps this executive is your manager. That individual should be your executive sponsor—someone who will be more than happy to support your project. For example, the goal may be “improve cross-selling to increase revenue” and the owner of that goal might be your VP of sales.
Customer relationship management (CRM) analysis projects are the most popular data mining applications within commercial enterprises, as customers drive thriving businesses. Customer-focused businesses often report their earnings in a manner that is tied directly to customer value improvements—how they created and captured value throughout the customer lifecycle.

For example, a major wireless telecommunications company recently exceeded quarterly earnings expectations, attributing much of its success in the quarter to improved customer retention and one of the lowest churn rates in its business. These improvements were driven by a best-practice analytical infrastructure fueled by data mining. By demonstrating the value of data mining in helping to achieve a critical strategic objective through incremental project successes, this telecom was able to integrate predictive analytics into its core business processes.

**List business questions related to your strategic objective**
Understanding your true goal is critical to managing your data mining project successfully, and ensuring that the project does not result in producing the right answers to the wrong questions. Identify your primary business objective as well as related questions that will need to be addressed.

Let’s say you are the marketing VP for a bank and that your primary business objective is to retain current customers who are at risk of moving to a competitor. Related business issues might include:

- **Channel influence on churn**—How does the interaction channel (for example, ATM, branch, or Web) affect whether customers stay or go?
- **ATM pricing association**—Will lower ATM fees significantly reduce the number of high-value customers who leave?
- **ATM pricing association with customer segments**—Will lower fees affect only one particular customer segment?

Try to specify all of the known business issues related to addressing your strategic objective to ensure that your data mining project is as business-focused as possible.

**Describe success metrics**
Data mining costs and resource requests must be clearly justified by outlining quantifiable benefits. Decide the best way to quantify project results and prove value to your stakeholders. Your executive sponsor and project stakeholders must be able to see how investments in data mining are worthwhile. Plan for the evaluation of your data mining results—which is critical to selling your project internally and generating the positive “data mining buzz” necessary to secure resources for additional data mining projects.

For example, success may be measured by reducing customer churn by 10 percent or simply achieving a better understanding of customer segments and their value. Beware of setting unattainable goals and make sure that the success metrics relate to at least one of your specified business objectives. There are a number of ways to quantify how data mining results achieve your business objective—from measuring the accuracy of predictive models on historical data using graphs such as gains charts (Figure 1) to testing deployment in a direct mail campaign.
Testing new business strategies on a small but representative sample using “champion-challenger” programs proves the benefits of data mining while limiting the risks. Champion-challenger testing provides you with a scientific, empirical method of comparing existing business processes against changes proposed by data mining, in order to verify that they will meet your success criteria.

Standard Life Bank successfully followed the champion-challenger approach. A remortgage mailing campaign was planned and executed using predictive response models. The mailing included a randomly selected control group. The models created scores for each customer, allowing the bank to focus its efforts on the best prospects for the remortgage product. This enabled them to more effectively target direct mail—producing a nine times greater response than that achieved by the control group.

Begin cost-benefit analysis with success benefits

Quantify the benefits of achieving your business goal as the foundation for a cost-benefit analysis. Decide how you will evaluate the success of your data mining project and measure the benefits it delivers.

Analyze the benefits of addressing your strategic objective with your initial data mining project:

- **Define each success metric**—for example, “improve direct mail campaign response rate by 25 percent”
- **Specify monetary measures whenever possible**—for example, determine the revenue that a 25 percent improvement in response rate is likely to produce
- **Total the benefits of your success criteria**—quantify the total projected monetary benefits of achieving your success metrics
- **Identify the “goal owner” who will assess the success metrics**—this will most likely be your executive sponsor

<table>
<thead>
<tr>
<th>Step one project readiness checklist:</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔ Strategic objective identified</td>
</tr>
<tr>
<td>✔ Related business goals listed</td>
</tr>
<tr>
<td>✔ Success criteria defined</td>
</tr>
<tr>
<td>✔ Success benefits quantified</td>
</tr>
</tbody>
</table>

After completing activities, assess your readiness using traffic lights—red, yellow, or green. These readiness status lights from each step will be assembled at the end of this white paper to create a readiness scorecard that provides a snapshot of areas you need to address to ensure the success of your first data mining project.
2. Line up essential resources

According to Alan Abrahams, professor of operations and information management at The Wharton School of Business, there are four critical data mining success factors—the right application, the right people, the right data, and the right tools. Once you’ve identified the right application in step one—an achievable business goal with justifiable solution costs and executive commitment—securing the necessary resource commitments is usually much easier. In step two, outline the resources available to accomplish your data mining project, and identify current resource gaps.

<table>
<thead>
<tr>
<th>Step two resource commitment activities:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The right people</td>
</tr>
<tr>
<td>2. The right data</td>
</tr>
<tr>
<td>3. The right tools</td>
</tr>
</tbody>
</table>

Line up the right people

People are the most cooperative and pre-disposed to collaborate when they share your goal, so sell your data mining project’s value to each person you hope to get on board. One of the most common mistakes made on data mining projects is to simply “throw a project over the wall” to the data miner for execution. Involving the right people up front in the right roles helps to ensure that your first project is a success.

Consider the roles below. Who is available in your organization to own the responsibilities for each role? Keep in mind that it is not uncommon for one person to serve multiple roles. In some situations, two people may cover all of these roles. Your specific resource requirements will depend on the size of your organization and the scope of your project.

Essential data mining project roles and responsibilities:

- **Executive sponsor**—guides the project from a strategic business perspective and ensures that data mining results are applied to improve business outcomes
- **Project leader**—manages the project and ensures team member accountability with the support of the executive sponsor
- **Data miner**—understands the questions that can be answered using data mining and how to answer them
- **Data expert**—knows the data available and how to access it
- **IT sponsor**—understands existing IT capabilities and signs off on data mining integration with those systems
The data miner is the most critical role needed to successfully complete a data mining project. Staffing this role appropriately, and performing data mining collaboratively, greatly increases your chances for success. It is occasionally thought that the project leader or executive sponsor can reasonably tell the data miner, “Here is the data, go do your data mining, and come back with the answers.” However, even in a best case scenario, this approach would take numerous costly iterations to produce useful results.

Plan for collaboration—at every stage of the data mining process—between your data miner, the project leader, and someone with a detailed knowledge of the business. Ideally, the data miner would possess the business knowledge. If your data miner lacks this domain expertise, he or she should literally sit next to someone with the required business knowledge. This person will most likely be the project leader. When this is not possible, regular “working sessions” should be scheduled to keep your project goal-focused. For this to work effectively, highly interactive data mining tools with good response time are also required.

Three options generally exist for staffing the data miner role:

- **On staff**—if you already have a data miner on staff
- **Train staff**—if you have a potential data miner on staff, such as a technically-savvy person with knowledge of the business problem, that simply needs training to come up to speed and prepare for the project
- **Outsource**—if no existing, internal resource exists and it makes more sense to outsource this role

**Line up the right data and secure buy-in from the data expert**

Discover what relevant data is available to meet your strategic business goal. In order to perform data mining, you need a data expert to help you identify the right data by answering questions such as: “What do the codes in this field mean?” and “Can there be more than one record per customer in this table?” Consider how much data knowledge is available and evaluate any risks caused by its absence or scarcity.

Get the right data lined up as easily as possible:

- **Start with readily accessible data.** While there are many large datasets that organizations can benefit from mining, it would be a mistake to believe that these should be the sole focus of data mining. Many successful data mining projects are performed on small or medium-sized data sets.
- **Validate data understanding statements.** Another pitfall to avoid is erroneous assumptions on the part of project leaders and data experts. Data miners should verify data understanding statements by examining the data.
- **Choose a data mining tool with data-understanding functionality.** This helps your data miner to spot mistakes in assumptions about the data before they cause quality problems in the data mining process.
Line up the right tools and secure IT buy-in
To achieve your business goal, models must be developed quickly and deployed cost-effectively for use within current operational systems and business processes. Successful data mining requires data mining tools that meet three requirements: an open architecture, rapid model development capabilities, and flexible model deployment options. This section helps you create a data mining tool evaluation scorecard with specific questions based on time-tested best practices.

Open architecture to capitalize on existing systems
Data mining is not performed in isolation. To achieve your business goal, data mining will need to be integrated with your data infrastructure and more broadly with your business processes. To gain buy-in from your IT department, data mining tools should integrate with your existing data infrastructure—relevant databases, data warehouses, and data marts—and should provide open access to data and the capability to enhance existing databases with scores generated by data mining.

This requires a data mining toolset with an open architecture that is designed to capitalize on your existing IT investments and scale to enterprise demands. Open, standards-based data mining tools should provide integrated capabilities for the entire data mining process, including flexible model deployment options for virtually every operating environment, and auditable model management.

Evaluate tools for “openness” by asking the following questions:
- Does the tool leverage your existing IT investments?
- Does the tool work well with your existing data stores?
- Does the tool work with open data standards or does data need to be migrated to a proprietary data format?
- Does the tool preserve your existing investment in technology such as other analytical tools?
- Does the tool require additional expenses for software, hardware, and services that reduce your chances of realizing a significant return?

Flexible deployment options to minimize costs
Different operating environments have different data mining integration requirements. Your data mining tool should provide you with the ability to integrate data mining within your current operating environment cost effectively—as well as the flexibility to deploy models into future environments. This includes the ability to export a complete data mining process—rather than just scoring code—which helps you realize significant cost savings over the life of the deployment by eliminating programming requirements. The programming costs related to data mining deployment options provided by some vendors, which deploy scoring code that must be integrated within existing systems by hand, can quickly add up to make your business case less compelling.
Evaluate model deployment flexibility by asking the following questions:
- How can I deploy my finalized predictive models (now and in the future)?
- How can models be integrated into operational systems?
- Can integration be done cost effectively or will it require a substantial programming investment? How easily can my solutions be updated?

**Rapid modeling capabilities accelerate time-to-value**
Data mining tools that enable rapid predictive modeling generate cost savings by minimizing data mining time-to-value—the time required to develop and deploy models that deliver value within business operations. Soft costs, such as the time cost of a delay in deploying data mining results, should be factored into data mining cost-benefit analyses if possible.

For example, what is the time cost of a 30-day delay in deploying a customer retention model that is predicted to save 2,000 customers per month? Lifetime value calculations provide a concrete time cost for such a scenario. Other significant time cost scenarios include businesses that operate in volatile markets—where delays in model deployment may mean that the model is outdated shortly after it is deployed.

Evaluate modeling productivity by asking the following questions:
- How does the tool prepare data? Can the tool automatically extract data for preparation?
- How does the tool enable interactive exploration and visualization of data?
- How does the tool boost analyst productivity? Does the tool enable data miners to develop effective models quickly? How easily can data preparation be done to meet specific model needs?
- Does the tool offer a wide range of techniques? Does the tool enable the combination of techniques?
- Does the tool achieve consistently high results?

**Step two project readiness checklist:**
- [x] Available people inventoried
- [x] Available data inventoried
- [x] Available data mining tools inventoried
3. Define an executable data mining strategy

How can data mining help solve the strategic business objective? This is the focus of step three activities—planning how to apply the three resources identified in step two to achieve your business goal. Collaborate with your data miner and other key stakeholders to define an executable data mining strategy.

**Step three data mining strategy creation activities:**
- Data mining methodology commitment
- Data mining goal definition
- Project plan creation
- Clear data mining success metric description

**Commit to a standard data mining process**

Data mining strategy was a hot topic at the Gartner Business Intelligence Summit 2005. During the session *CRM Analytics—Realizing your Potential*, Gartner Research Director Gareth Herschel pointed out that successful data mining requires a strategic focus on decision improvement, rather than a top-down, tactical focus on analysis. As he explained, “What is needed is a bottom-up approach to analytics, which begins by identifying the desired decision and then works back to the analysis that will drive the decision, and the data that will drive the analysis.” Herschel elaborated on bottom-up data mining during the session *Taking BI to the Next Level: Using Data Mining and Predictive Analytics*. He recommended that data mining projects start with both a well-defined problem and process. “There are several industry process models that can be employed in data mining to leverage the different technologies available to data miners (a popular example is CRISP-DM).”

Proving the value of data mining for improving decisions directly related to your business goal will greatly increase your chances of gaining the internal support needed. To produce useful data mining results that improve decision making, clearly define business and data mining goals—as well as model deployment plans. A simple way of ensuring this is to use a standard data mining methodology such as the CRoss-Industry Standard Process for Data Mining (CRISP-DM). Such a process helps to enable business-focused data mining project planning and provides the common language necessary for collaboration among “the right people” identified in step two.
CRISP-DM is the de facto standard for implementing data mining as a strategic business process to optimize decision making (Figure 2). An open standard developed by an international consortium, including DaimlerChrysler, NCR, and SPSS, CRISP-DM applies data mining to a wide range of business objectives. The process begins with an understanding of your business goals and available data, and ends with the deployment of data mining results into business operations to optimize critical decisions and deliver measurable ROI.

Other methodologies can also be used with the planning steps in this paper. However, this paper focuses on CRISP-DM, as it is the most widely used methodology.

**Define achievable data mining goals**

Your data mining goal must state project objectives in business terms such as, “Predict how many widgets a customer will buy, given their purchases in the past three years, their demographic information, and the item price.”

If your business goal cannot be effectively addressed by data mining capabilities and translated into a data mining goal—and executable data mining strategy—it may be wise to consider redefining the business problem at this point.

**Create a project plan for achieving your goals**

The project plan describes the intended plan for achieving the data mining goals, including outlining specific steps and a proposed timeline. Create a project plan based on the process steps of the data mining methodology you have chosen. Make sure to confirm resource commitments as you collaboratively plan with project stakeholders.

In our experience, projects with a narrow focus are usually the most successful vehicles for introducing data mining to your organization. Your first project should be a focused, well-planned effort to sell the value of data mining. Focus model development and deployment on a specific, small project that can deliver rapid results for your business goal. Narrow the scope of the project as much as possible—30 days is a reasonable time frame for a focused project.

---

**Table: Initial project plan**

<table>
<thead>
<tr>
<th>Task Name</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set selection</td>
<td>9/5/00</td>
<td>9/6/00</td>
</tr>
<tr>
<td>Model development</td>
<td>9/6/00</td>
<td>10/30/00</td>
</tr>
<tr>
<td>Model deployment</td>
<td>11/1/00</td>
<td>12/31/00</td>
</tr>
<tr>
<td>Model evaluation</td>
<td>1/1/01</td>
<td>2/28/01</td>
</tr>
<tr>
<td>Deployment</td>
<td>3/1/01</td>
<td>4/30/01</td>
</tr>
</tbody>
</table>

---

**Figure 3: CRISP-DM focuses data mining on rapid model development and deployment to optimize decisions**

**Figure 4: Initial project plan**
Apply CRISP-DM, or a similar process that focuses on improving goal-focused decision making, to plan how you will achieve your business goal. Make sure to map the process steps from your business goal to your data mining goals, all the way through to how data mining results will be deployed into the business process you aim to improve.

**Define data mining success**

The most common data mining success criterion is the predictive accuracy of the model—but models only need to be accurate to a certain level to achieve business objectives. Plan to make tough trade-offs to determine when predictive accuracy is high enough to achieve your business goal.

It is true that a predictive model should have some degree of accuracy, because this demonstrates that it has truly discovered patterns in the data. However, the usefulness of an algorithm or model is also determined by a number of other properties, one of which is whether understanding the resulting model requires deep technical knowledge or is something that can be understood by a typical analyst.

For example, British Telecom (BT) used data mining to improve customer acquisition campaigns, and reached their goal even before finalizing their models. Stephen O’Brien, senior consultant at BT, commented, “Even before completing our final models, we were able to surpass our original target—and increase the campaign response rate by 100 percent.”

---

**Step three project readiness checklist:**

- ☑ Commitment to a data mining methodology
- ☑ Achievable data mining goal defined
- ☑ Initial project plan complete
- ☑ Clear data mining success metric chosen
Next steps
Each of the three planning steps in this white paper ended with a project readiness checklist. Assess your estimated readiness for each of the steps by circling your current status using the traffic lights—red, yellow, or green—in the data mining project readiness scorecard below. These status lights provide a clear picture of your first data mining project’s readiness.

<table>
<thead>
<tr>
<th>Data mining project readiness scorecard</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Strategic focus:</strong></td>
</tr>
<tr>
<td>Strategic objective identified</td>
</tr>
<tr>
<td>Related business goals listed</td>
</tr>
<tr>
<td>Success criteria defined</td>
</tr>
<tr>
<td>Success benefits quantified</td>
</tr>
<tr>
<td><strong>2. Resource commitment:</strong></td>
</tr>
<tr>
<td>Available people inventoried</td>
</tr>
<tr>
<td>Available data inventoried</td>
</tr>
<tr>
<td>Available data mining tools inventoried</td>
</tr>
<tr>
<td><strong>3. Data mining strategy:</strong></td>
</tr>
<tr>
<td>Commitment to a data mining methodology</td>
</tr>
<tr>
<td>Achievable data mining goal defined</td>
</tr>
<tr>
<td>Initial project plan complete</td>
</tr>
<tr>
<td>Clear data mining success metric chosen</td>
</tr>
<tr>
<td>r= Red  y=Yellow  g=green</td>
</tr>
</tbody>
</table>

Complete cost-benefit analysis with resource costs
Now that you have a general idea of your data mining project readiness, finalize the cost-benefit analysis you started in step one by quantifying the costs of addressing your business goal. What resource gaps exist and what are their approximate costs? Use your data mining project plan created in step three to illustrate the effect resource gaps may have on the schedule.

Analyze the costs of addressing your strategic objective with data mining, factoring in costs such as:

- Initial software purchase costs
- Services/outsourcing costs
- Training costs
- On going software maintenance costs

Justify your data mining project’s costs with the benefits you’ve quantified—subtracting your total costs from your total benefits. Assess whether the required investment in data mining justifies the benefits you hope to achieve. Note that this cost-benefit analysis—and the evaluation of your data mining project’s results—is based only on the results of your first project. With those results in hand, we typically find that you will have an easier time justifying and allocating resources for subsequent projects.
About Clementine® for data mining

Clementine is an enterprise data mining workbench that enables you to quickly develop predictive models using business expertise and deploy them into business operations to improve decision making. Designed to support CRISP-DM, Clementine’s open architecture utilizes your existing IT investments to enable rapid predictive modeling and high-ROI model deployment. Other data mining workbenches fail to support data mining business processes, focusing instead on modeling performance within closed systems that often require extensive additional investments before deploying models that deliver business value.

Clementine addresses the three data mining tool requirements needed to improve project success:

- **Rapid predictive modeling.** Clementine’s interactive visual workflow interface enables you to deploy valuable models into your organization’s operations in less time. Support and scaling for every data mining process phase makes Clementine the most productive workbench available.

- **Flexible model deployment.** Achieving a high ROI using data mining requires the flexibility to deploy models within current operational systems and processes. Clementine provides deployment options for nearly every operating environment—from basic batch scoring to model deployment into real-time predictive applications such as PredictiveCallCenter™.

- **Open architecture.** Clementine’s open design makes it possible to add a complete data mining platform to your existing systems. Integrated capabilities for the entire data mining process—including database access, text and Web data preparation, and model management—enable you to maximize value from your existing IT assets. Other vendors’ closed systems require additional, third-tier database technology and complex annual “rental” pricing.

Contact SPSS Sales for help in reviewing Clementine’s capabilities and for recommendations to ensure that your first data mining project is a success.

About SPSS Inc.

SPSS Inc. (NASDAQ: SPSS) is the world’s leading provider of predictive analytics software and solutions. The company’s predictive analytics technology improves business processes by giving organizations consistent control over decisions made every day. By incorporating predictive analytics into their daily operations, organizations become Predictive Enterprises—able to direct and automate decisions to meet business goals and achieve measurable competitive advantage.

More than 250,000 public sector, academic, and commercial customers, including more than 95 percent of the FORTUNE 1000, rely on SPSS technology to help increase revenue, reduce costs, and detect and prevent fraud. Founded in 1968, SPSS is headquartered in Chicago, Illinois. For additional information, please visit [www.spss.com](http://www.spss.com).