become open-loop, with researchers chasing down rabbit-holes and not reporting back when necessary. Frequent (e.g., weekly) detailed meetings and intermediate reports are painful (from the point of view of a technical person eager to try the next research idea) but essential for the survival of the project.

At the end (either problem or budget runs out), submit findings in a complete report (a short paper, headed by an executive summary, with model and methodology details in appendices) and a presentation, deliver the results in the form of computer code or logical rules that can be converted to the best format for the client, and work closely with those responsible for implementation to ensure that the models integrate seamlessly into the existing systems environment. Always remember to suggest valuable next steps suggested by your findings; you’ll want to make it easy for the client to ask you to continue your valuable work!

Professional Development

Continual learning is vital. The latest software tools are powerful, able to tremendously augment the productivity of an analyst—but only one who knows in what direction to push. That experience is hard to accumulate in faster than real time, but a single trait—humility—helps greatly. Usually, the data mining expert is the person in the room who knows the least about the specific problem being faced by a client. No matter how expert you become in the mining craft, you must be open to learning the business constraints, metrics, and vocabulary of the client as quickly as possible. Clients have great tolerance for (in their mind) stupid questions in the first couple of weeks of an engagement, but they aren’t encouraged if they have to explain things more than once. Fortunately, many problems that seem completely unrelated become clean data mining problems with the right level of abstraction. We have used data mining to successfully

- Select profitable stocks;
- Detect fraudulent claims;
- Score an applicant’s credit;
- Discover cross-selling opportunities for products;
- Quantify drug efficacy;
- Forecast new product sales;
- Discover new customers;
- Recognize objects in images;
- Verify identify through biometrics; and
- Anticipate shifts in market sectors for hedge funds.

The data miner, with experience, gains a breadth of perspective that enables him or her to find creative and effective solutions to a great variety of challenges. It’s this variety, and the “detective” nature of the work, that can be most satisfying.

Though the data miner has the least domain-specific knowledge, he or she brings an analytic expertise that is rare and necessary to the team. Statistical thinking is hard—it’s been shown to be the “least caught” discipline taught at universities—and uncommon.