IT@Intel White Paper
Intel IT
Enterprise Advanced Analytics
April 2014

Broadening Access to Advanced Analytics in the Enterprise

With training and support, a broader segment of Intel employees can learn the fundamentals of advanced analytics and bring significant value to their teams and the enterprise.

Asoka Diggs
Enterprise Architect, Intel IT

Christy Foulger
Enterprise Architect, Intel IT

Executive Overview

Intel IT has laid the groundwork for mining big data using business intelligence and advanced analytics techniques. Traditionally, advanced analytics have been viewed as the domain of a few highly skilled data scientists. We believe that there is an opportunity to extend some of this knowledge to business teams throughout the enterprise. By learning the basic skills in advanced analytics, employees can frame their business problems as data mining problems, enabling them to work more effectively with data scientists and even perform intermediate data mining projects themselves. We have some initial results from our efforts so far and are still exploring how well this idea works in practice.

We are focusing on several key areas to broaden the access to advanced analytics:

- **Training.** We identified several levels of training appropriate for different skill sets and objectives and have developed and delivered an introductory level class. We have also certified trainers to meet the growing demand for this class.
- **Candidate teams.** We continually identify and encourage teams and individuals to pursue training in advanced analytics and to begin applying what they learn to their business data.
- **Mentoring.** We encourage trainers and others with the necessary background to serve as mentors who can help individuals and candidate teams apply what they have learned in class to their real-world business problems.
- **Community of practice.** We plan to establish a virtual community or workgroup where employees can support one another in their advanced analytics discovery process.
- **Analytics tool kits.** We are creating an analytics virtual machine that employees can use to practice their advanced analytics skills and perform proof of concepts or advanced analytics discovery.
- **Metrics.** We currently track participation metrics and are designing outcome metrics and other success indicators.
- **Business processes.** We plan to deploy a business process around analytics to improve the efficiency and consistency of the results.

When data science is limited to just a few experts, business teams and the enterprise do not realize the full benefit of advanced analytics. With training and support, a broader segment of Intel employees can learn the fundamentals of advanced analytics and bring significant value to their teams and the enterprise.

Contents

IT@INTEL

The IT@Intel program connects IT professionals around the world with their peers inside our organization—sharing lessons learned, methods, and strategies. Our goal is simple: share Intel IT best practices that create business value and make IT a competitive advantage. Visit us today at www.intel.com/IT or contact your local Intel representative if you'd like to learn more.

BACKGROUND

Intel IT is integrating advanced analytics with business intelligence (BI) as a method of extracting business value from big data. Advanced analytics skills, such as statistics, text mining, machine learning, and visual analytics, are essential, but only a few highly trained individuals currently have these skills.

At the same time, Intel's business units are increasingly requesting big data solutions without providing detailed requirements, assuming that big data is a black box that will solve any business problem. The conversations about what is required are further complicated by the lack of even basic knowledge in the field of advanced analytics.

This knowledge gap can slow the progress of the various teams who are trying to solve business problems using big data. These teams may include the following:

- **Business analysts.** Although business analysts want to accomplish self-service BI and analytics, they do not have the necessary training and guidance and must rely on IT BI professionals to provide reports. We believe that advanced analytics will provide these analysts with an additional way to approach their data, business problems, and opportunities.
- **IT BI professionals, such as report developers and systems analysts.** These BI professionals support business analysts and implement the analytics models. We believe that gaining the basic skills in advanced analytics will enable these professionals to work more effectively with analysts and provide better support for deploying analytics models.
- **Data scientists.** Data scientists can take much longer to produce analytics models when working with employees who lack knowledge in this area. Similarly, if the data scientists do not have expertise in the employees' specific business domain, the models they produce can miss important information.
- **Teams with data.** These teams do not have the necessary skills to frame business problems as data mining problems. This deficiency may lead to the inability to establish clear requirements, adding months to an analysis project. With moderate skill development in advanced analytics, some of these individuals and teams will be able to frame business opportunities as data mining problems and perform some of the data preparation required at the beginning of an analysis.

We wanted to explore ways to make advanced analytics accessible to more employees in the enterprise. We theorize that employees who understand the basics of advanced analytics will be better able to contribute to projects and analyses.

SOLUTION

Intel is embarking on a strategy to extend the knowledge of advanced analytics to a broader portion of the enterprise. This strategy is an important component of our larger strategy to use big data as a source for competitive advantage. Our approach moves from the idea of data as a source of power in the hands of a small group and instead aims to put data and analysis of the data into the hands of as many employees as possible, while maintaining privacy and security.

Central to our strategy is revising our assumption that only data scientists with deep analytics skills can provide business insight. While this group still retains the critical role of analyzing data, they are augmented by a broader group of employees with less extensive analytics skills, significantly expanding the amount of advanced analytics-based insights that the company can pursue. This broader group might not be able to create complex analytics models, but they can better participate in advanced analytics work.

We anticipate that with more employees and job roles involved in advanced analytics, we will be able to take advantage of more opportunities through our data science teams and projects. Ultimately, we hope to initiate a cultural change where advanced analytics are used not only for large projects but also as a routine part of how we approach our daily business problems and opportunities at all levels of the enterprise.

Getting business results through advanced analytics involves people, technology, and processes. We are focusing on the people aspect first—in particular, growing their skills in advanced analytics.

Bringing Advanced Analytics to the Enterprise

Our task was to identify the challenges involved with bringing knowledge of advanced analytics to the enterprise and then develop some working ideas about how to meet these challenges.

We divided the challenges into several categories:

- **Candidate teams.** Which teams are good candidates for learning advanced analytics skills?
- **Training.** How can we develop training appropriate for different skill sets and objectives?
- **Mentoring.** Who can help employees apply what they have learned in class to their real-world business problems, data, and teams?
- **Community of practice.** How can employees support one another in their advanced analytics learning and discovery process?
- **Analytics tool kit.** What tools can we provide to facilitate the learning and initial practice of advanced analytics skills?
- **Metrics.** How do we measure success?
- **Business processes.** How can we integrate the data mining process into our existing business processes?

With these challenges identified, we launched our efforts to broaden access to advanced analytics with a program we call Advanced Analytics for Everyone. This program addresses two important points:

- **Skills first, then tools.** A common misconception is that given the right tools one can perform advanced analytics. However, we believe that having data mining skills is more important than choosing a particular tool. We also believe that establishing a mental model of data mining is the key to widespread adoption of advanced analytics into business decision making. This mental model is the internalized thought process of how data mining works in the real world.
- **The "everyone" aspect.** Our goal is to bring this program to every knowledge worker at Intel. In reality, with an organization of around 100,000 employees, not every employee will receive training immediately. However, as our program name implies, the data mining framework is a basic skill that all employees can acquire and apply in an information economy (see the "What Does a Company with Advanced Analytics-Capable Employees Look Like?" sidebar).

We plan to initially target training to project and functional teams to drive early business value. These early adopters will help build momentum and support for our program.

The Relationship between Business Intelligence and Advanced Analytics

Big data generally refers to the large, unstructured data sets that constitute up to 90 percent of enterprise data. This data includes text, audio, video, click streams, log files, and more. *Business intelligence* (BI) refers to the methodologies, tools, and technologies for mining data (big and small) to gain richer and deeper insights into business patterns and trends. BI focuses largely on historical data that answers questions such as What happened? Where is the problem? and What actions are needed? BI is also known as *descriptive analytics*.

Analytics is the scientific process that transforms data into business insights. The emphasis is on gaining insights that enable the organization to make better decisions about the future. Intel is particularly interested in *advanced analytics,* which includes a variety of analytics approaches, such as the following:

- **Predictive analytics.** Focuses on predicting future probabilities and trends. Predictive analytics attempts to answer the question, What will happen next?
- **Prescriptive analytics.** Focuses on finding the best course of action for a given situation. Prescriptive analytics answers the question, What should I do?
- **Text analytics.** Focuses on preparing and interpreting unstructured text, such as email and social media streams, to make it accessible for other types of analysis, such as descriptive, predictive, or prescriptive analytics. Text analytics can be used to answer questions such as, How do my customers feel about my products?

What Does a Company with Advanced Analytics-Capable Employees Look Like?

We believe that there is the potential for a profound business transformation as we implement the Advanced Analytics for Everyone program. As more employees become fluent in analytics, the entire enterprise increasingly becomes a learning organization. A learning organization approaches opportunities and challenges differently than a traditional organization. Because learning organizations can learn faster than their competitors, they are naturally more agile, more adaptable, and more competitive.

For example, the positive effect advanced analytics has had in our company has led the manufacturing and purchasing teams to work more closely together. Using detailed analytics data, the teams are able to discuss defects and parameters to improve materials being purchased for high-volume manufacturing.

Another example is the use of Human Resources, IT, and Corporate Services data to better identify and serve employee needs. Each team has different employee data sets, such as roles, applications, physical environments, and devices, which together form a comprehensive view. With input from the security and legal teams, we can use this data to provide better services to employees.

As more teams embrace advanced analytics skill sets, we will be able to provide even better cross-organizational communication.

CANDIDATE TEAMS

We want to encourage learning as something a team does together rather than something that individuals do. Therefore, identifying and encouraging teams to take training—and getting their buy-in—continues to be important. Our early experience is that employees and teams throughout the company are eager for this knowledge.

We consider teams to be groups of employees, not necessarily reporting to the same manager, who work together on projects or in particular business domains. We believe that our training and mentoring will be most successful with teams comprising three to seven people.

Some examples of teams that we have targeted for early adoption of training include the following:

- **BI teams.** These small groups already build BI applications and want to extend their team's skill set to include more analytics in their solutions. BI teams tasked with supporting big data initiatives are particularly interested in the training.
- **Business data analysis teams.** These teams are looking for new ways to think about their data and improve their business decisions. Often they benefit most rapidly from the training as they integrate data cleaning and preparation into their current data analysis to improve and expedite their work.
- **Functional teams.** These teams include business analysis teams with their BI team counterparts. By learning about and applying advanced analytics together, they gain a shared vocabulary and mental model that improves the quality and efficiency of the solutions they build. Whenever possible, we prefer working with these types of partnerships.
- **Individuals with a passion for the topic, regardless of their specific job role.** These individuals already have an interest in and motivation for learning advanced analytics, which draws them to the training. Some of these individuals will continue their education and become the practitioners and data scientists that the enterprise needs. We also believe that many of these individuals will become the advocates that help promote the Advanced Analytics for Everyone program.

Our primary goal is to identify teams that would like after-training mentoring and assistance to help them apply advanced analytics on the job. As shown in Figure 1, we developed a workflow model that illustrates how we envision teams applying their training to their real-world business problems and opportunities.

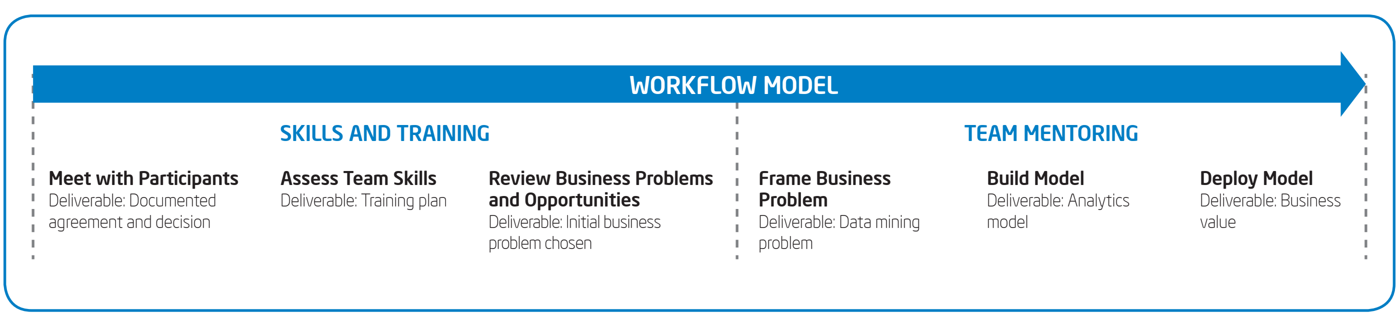

Figure 1. Our workflow model provides a conceptual view of the tasks and deliverables that teams will generate as they learn and apply advanced analytics. The first three tasks focus on the skills and training that teams will need to get started. The remaining three tasks focus on team mentoring, where teams learn how to apply what they learned to their business.

TRAINING

Another significant challenge was addressing the limited training available for different skill levels and levels of effort. Traditional thinking has been that acquiring the skills for data mining requires a graduate degree, which is not practical for most employees. Our solution was to develop or provide training courses for various skill levels and objectives, as shown in Figure 2.

Data Mining Survey

This 30- to 120-minute demonstration highlights the importance of data mining and the need for employees to learn more about this subject. The main message is that employees can obtain an appropriate, useful level of training with the minimal 20-hour introductory class.

Introduction to Data Mining

This two-and-a-half-day class gives employees a common vocabulary and a consistent way to frame business problems as data mining problems. This class also helps employees recognize opportunities to initiate data mining projects.

The audience is IT BI and software engineering teams, business analysis teams, and any team that can use advanced analytics in their business. Management, existing statisticians, and existing data scientists also gain value from this introductory class, primarily from its focus on learning how to frame business problems as data mining problems.

After taking the training, employees will have tangible skills that when combined with their existing business knowledge, skills, and expertise will be of increased value to the enterprise.

We developed this class and currently deliver it internally. We chose this approach because we have not found external training that focuses on the mental model of data mining or how to frame business problems as data mining problems. We consider these two objectives important for shaping the scope of the class and enabling it to be effective in a short period of time. For further details on this class, see "The Introduction to Data Mining Class: A Deeper Look."

Practitioner Training

From our perspective, the practitioner skill level can be achieved through a 200-hour graduate-level class. At this level of skills development, employees are beginning to create and validate analytics models. This level of training will also further develop their skills for framing business problems as data mining problems.

Practitioner training falls into two groups:

- For some employees, the practitioner training represents an aspirational level of achievement that provides depth in analytics to augment their domain expertise or other skills and job experience. Within this group, we currently envision two levels, with training tailored appropriately to each:
  - The first level is a business-focused class that explores in depth the framing of business problems as data mining problems.
  - The second level is a data preparation class for extract, transform, and load (ETL) developers and others with an interest in exploring the intricacies of preparing data for analysis.

  These additional levels of training have not been created internally.
- For other employees, practitioner training is the beginning of training in formal analytics and possibly the start of a multiyear effort to transition into a data science role. For this group, the class would be structured similarly to a graduate-level data mining class. The class will have a greater emphasis on data mining techniques and methods than the previous two practitioner classes mentioned. One common source of candidates with this focus is employees with a math, statistics, science, engineering, or similar background involving the study of numerical analysis but who haven't formally studied the application of data mining to business.

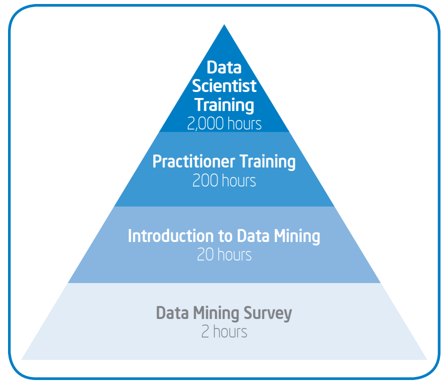

Figure 2. We provide training courses for various skill levels and objectives. A smaller number of employees may choose to pursue higher levels of training and travel farther up the pyramid. Our training efforts focus on the middle tiers of the pyramid, at the introductory and practitioner skill levels.

We plan to use training from a massive open online course (MOOC) and other third-party trainers. We believe that these structures will scale to our level of demand. More importantly, the topics and style of the teaching will be consistent with industry views for this level of training and should also achieve our objectives.

We plan to develop and deliver the class internally in cases where we are unable to identify external training, and, more particularly, whenever we believe we can add significant value beyond the training we can receive from the industry.

Data Scientist Training

This level of training requires an employee to either earn a graduate degree or obtain approximately 2,000 hours of self-study and experience. Data scientist training is an important level of skill development for some employees of the company. However, this level of skill development is not addressed in this paper.

Completing the introductory and practitioner levels of training can help employees identify whether they want to pursue this more in-depth course of study.

MENTORING

Another challenge we identified was the lack of tangible guidance for teams and individuals. After undergoing training and self-study, employees have limited opportunities to get help for applying advanced analytics to their business. We envision a mentor who can bridge the gap between what the team learns in training and the team's real-world business problems. Ideally, the trainers of the Introduction to Data Mining class become the mentors for their teams, but that is not required.

The mentors provide guidance while the teams do as much of the work as possible. With repetition of this work, the team should be able to at least frame their business problems as data mining problems. Ideally, these teams are also able to initially prepare their data for analysis and with study and practice are able to identify a wider range of the analytics models needed.

Mentors help teams answer several key questions:

- Is the business problem explained in enough detail that a data mining problem can be framed?
- What are the data mining problems?
- Is the team's data ready for analysis? If not, what needs to be done?
- What additional skills do the team members need?

Together, mentors and teams can identify a roadmap for identifying and solving data mining problems. For the mentor, the focus is not on building analytics models; the focus is on the team and assisting the team members as they begin to apply their advanced analytics skills. Ultimately, the team objective is to continue using their analytics skills without needing the mentor's assistance.

COMMUNITY OF PRACTICE

Because the field of advanced analytics is vast, data scientists specialize even when taking graduate-level courses. Many business solutions may require a variety of algorithms, tools, or different levels of business knowledge or skills.

With that in mind, we recognized the need for a larger community of employees to share their insights and seek help for applying advanced analytics to their data mining problems. Our idea is to establish a working group to support the growth of advanced analytics skills throughout the enterprise. Social and collaboration software can provide improved connection opportunities between the participating individuals. Within this community, employees can provide answers to common analytics questions, minimizing the learning curve of advanced analytics. We also envision that employees can share their data mining projects and analytics models, which may yield useful results for other business groups.

ANALYTICS TOOL KITS

To practice their advanced analytics skills, employees are requesting an environment with more parallel processing power and memory than their laptops provide.

Our proposed solution is to create an "analytics sandbox"—a virtual machine (VM) image that includes analytics tools and data sets to practice on. This analytics VM will enable employees who are developing their advanced analytics skill set to have access to more processing power than their individual laptops provide. The VM image will also provide better security and supportability. During training, the analytics VM provides a consistent learning environment for employees. After training, employees have a work environment that is capable of processing the low-end data sets that are common to many business problems and discovery situations. Data scientists are also requesting analytics VM images to perform discovery.

We are currently in the early phases of this project and are working with hosting, security, and license teams while establishing the images that will be the most useful for the majority of our users.

METRICS

As we implement solutions to meet our advanced analytics challenges, we need to identify and document success indicators. Our first attempt at monitoring progress is through training participation metrics. In 2013, two trainers delivered the Introduction to Data Mining class to over 190 employees. During that time, we added 15 more certified trainers. We have received positive feedback on the training (see the "Success Story: Applying Learnings to Business Problems" sidebar).

In 2014 we plan to continue to deliver the introductory class, mentor teams, generate candidate outcome or result metrics, and begin to evaluate results.

Some of the areas we have identified for future measurement include the following:

- Teams that have identified and are modeling solutions to problems (a qualitative, rather than quantitative, measure)
- Improved efficiencies associated with analytics efforts (cost improvement)
- Increased revenue associated with analytics efforts
- Time to obtain results for advanced analytics projects
- Longevity of models in production
- Collaboration metrics based on interactions in the community of practice, such as positive movement toward completion of a goal or the number of best practices shared
- Team deliverables completed at each stage of the analytics workflow

For example, we could measure success based on the number of teams that have reached each milestone in our engagement workflow model. As shown in Table 1, these deliverables are completed during various stages of the advanced analytics workflow (shown in Figure 1).

BUSINESS PROCESSES

Advanced analytics has a variety of defined processes. Currently our approach is to adopt the Cross Industry Standard Process for Data Mining (CRISP-DM), an iterative data mining process model that includes stages for moving from business problems to deploying analytics models. As shown in Figure 3, we use a simplified conceptual model with three stages:

1. **Define the business problem.** In this stage, teams frame the business problem as a data mining problem. This stage can comprise over 50 percent of a project, which is why we emphasize this activity in the Introduction to Data Mining class. We believe that business teams that can carry out this stage effectively can engage with a data science team more effectively and complete projects in less time.
2. **Solve the data mining problem.** In this stage teams prepare the data and build and evaluate the analytics model. For simpler problems and analyses, business teams can perform this work themselves. Otherwise, a data science team gets involved at this stage.
3. **Use the analytics model.** In this stage, teams can now apply the created model to their business problem. For example, the team might use the model to make a single decision. Another example might be that the team will use the model in a highly repetitive business process, and so the model needs to be delivered to production. The specific form that the model takes can vary widely, from a web service, to pre-calculated results that are viewed in typical BI-type reports, to automated business rules that use the model's output to implement decisions.

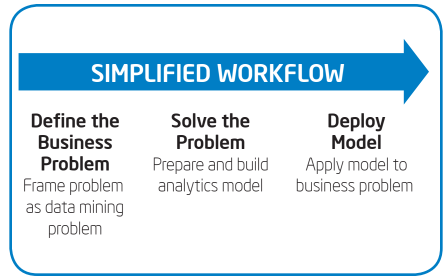

Figure 3. We use a simplified version of the Cross Industry Standard Process for Data Mining (CRISP-DM).

Table 1. Metrics Idea Based on the Number of Teams That Have Completed Deliverables during the Analytics Workflow

	2013		2014			
Deliverable	Q3	Q4	Q1	Q2	Q3	Q4
Scope Decision Complete	1	2	5	8	10	12
Training Plan Documented		1	2	4	5	8
Training 90-Percent Completed				2	3	3
Business Problem Chosen			1	1	3	5
Data Mining Problem Framed				1	2	3
Analytics Model Validated					1	2
Model in Production						1

Success Story: Applying Learnings to Business Problems

One participant who has successfully applied what he learned in class is Idris Motiwala, who works in the analytics group of Intel Sales and Marketing. He attended the Introduction to Data Mining class to understand how to relate business data to the business problems that needed to be solved. He also wanted to learn how to prepare and cleanse data and how to interpret statistical modeling techniques such as co-relation, attribution, regression, and decision tree analysis.

After attending the introductory class, Idris reviewed his sales-focused business environments to identify business problems that data mining could help solve. He identified two use cases where diagnostic and predictive analytics could be applied:

- Explain why sales volume actuals differed from the organization's goals for a popular PC client feature
- Predict which pending online sales opportunities will lead to actual sales, and by when

Idris and his team members are actively working to further proliferate data mining and the predictive analytics approach and principles when eliciting business needs and helping solve business problems for Sales and Marketing. When asked if he would recommend the class to others, Idris replied, "Yes. We at Intel are drowning in loads of data, and data mining helps us derive insights, patterns, and rich analysis that are sometimes challenging to uncover in the volume of data."

We believe that a natural extension to the existing BI workflow and processes is to include advanced analytics. Our intent is to incorporate the processes, roles, and methodologies in the training and mentoring we deliver as a baseline for advanced analytics teams.

A Roadmap for Implementing the Advanced Analytics for Everyone Program

We created a roadmap that highlights the major tasks involved in implementing advanced analytics for everyone across the enterprise (see Figure 4). These tasks are planned each quarter and are divided into three stages: design, deploy, and improve.

As mentioned previously, we focus on growing the necessary advanced analytics skills first. We start with the Introduction to Data Mining class, followed by the practitioner training, mentoring, and community of practice that augment the introductory training. Deploying an analytics VM is also an important part of supporting advanced analytics skill development. Metrics come next in our effort to measure progress and success.

As the advanced analytics skills are developed, we plan to develop a complete tool, technology, and platform specification (not discussed in this paper). We will then develop updates to BI and analytics business processes.

We believe that employees with advanced analytics skills will be able to build some solutions using whatever tools are available. As employees practice their skills, they will start to gain a better idea of their tool requirements. With a sufficiently robust community, we can make better quality tool and platform decisions without delaying or interrupting skill development. Similarly, we plan to postpone process development to avoid constraining skill development.

These ideas are demonstrated in the roadmap shown in Figure 4. The information for Q1 2014 and earlier represents our actual efforts and experience. For Q2 2014 and beyond, the information represents how we presently think the Advanced Analytics for Everyone program will develop.

The Introduction to Data Mining Class: A Deeper Look

As discussed previously, we identified three levels of training appropriate for different levels of skill and effort. We focused on developing and delivering the Introduction to Data Mining course, targeted at the base level of knowledge in the skill pyramid. We consider this class to be a significant way to bring the appropriate level of knowledge of advanced analytics to employees.

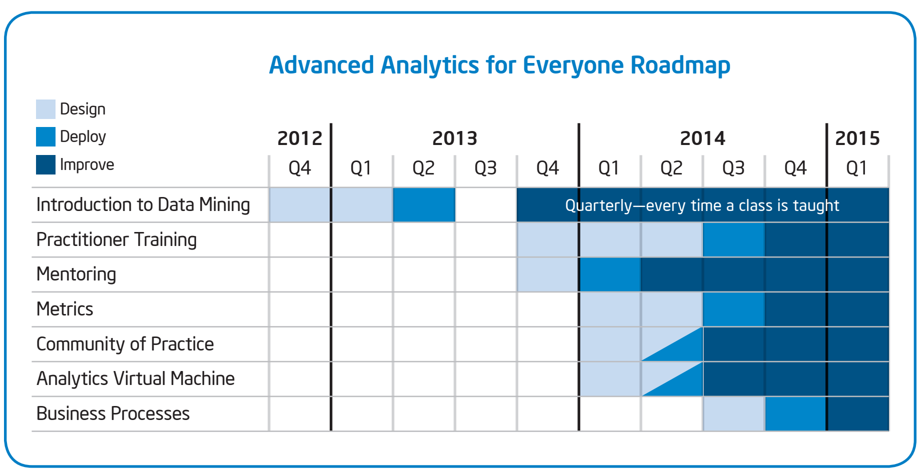

Figure 4. Our roadmap for implementing advanced analytics to the enterprise highlights major tasks that are divided into three stages: design, deploy, and improve.

OUTCOMES

The main objective of the Introduction to Data Mining class is to help employees learn how to frame business problems as data mining problems. Even highly skilled data scientists have benefited from this training, though this was not the intent of the design.

The course is designed for an employee who has no prior exposure to data mining. Through hands-on practice, they learn how to use a data mining tool to clean and prepare data, build and visualize basic statistical models, validate models, generate model predictions for data sets (score data), and complete an analysis of a variety of data sets.

Even if after taking this class an employee decides not to advance to practitioner training, he or she will still have acquired a mental model of what occurs in data mining. Employees will be able to apply this framework to the skills and experience they already possess, such as BI-specific technical skills or business domain expertise, and will find they are better able to participate in data science projects. For example, employees who have taken the class have reported that they can prepare data for analysis in less time.

OUTLINE OF COURSE CONTENTS

As shown in Table 2, the general outline that we developed for the class includes seven modules to be covered in two-and-one-half days. The outline can be tailored for a particular team's needs.

CLASS RESOURCES

We use two primary resources in the Introduction to Data Mining class: a data mining textbook[1] and an open source data mining tool.

The undergraduate-level textbook that we use provides an introduction to data mining and predictive modeling techniques. This book includes screenshots and step-by-step instructions for analyzing various data sets. To simplify the topic so that it can be addressed in a brief amount of time, the book makes two assumptions:

- **Exclude the math performed by the various algorithms.** We agree with this choice because we believe that when learning how to frame business problems as data mining problems, it is more important to learn the structure of the data used as input to each technique and the structure of the output.
- **Exclude the model validation until the end of the book.** We do not follow this choice. Instead, we include model validation with the first model and continue to include it as a component of every model built. It is our view that creating an analytics model and validating that model are really a single activity and that unpredictable business results can occur by separating the two.

A highlight of the book is that each business problem includes a data dictionary—an excellent example of role modeling good analytics practice by clarifying what each attribute actually means.

We think it is crucial to get employees actively engaged in building analytics models themselves. Building models enables employees to immediately try out and apply the concepts that they are learning. We chose to use the data mining tool that the textbook uses in its examples. This open source data mining tool is GUI-centric (as opposed to being a programming language) and user friendly. After providing a brief orientation, we have observed that employees do not struggle with learning and using the tool. Instead, they are able to focus on learning data mining concepts.

[1] North, Matthew. *Data Mining for the Masses*. Global Text Project, 2012.

Table 2. The Outline for the Introduction to Data Mining Class

	Module Title	Module Description
Day 1	Overview and Cross Industry Standard Process for Data Mining (CRISP-DM)	▪ Introduction to data mining concepts ▪ Demonstration of building an analytics model ▪ Overview of CRISP-DM, with an emphasis on formulating the data mining problem
	Classification: Decision Tree and Validation	▪ Introduction to decision trees, which can be used to classify data and as input to decision making ▪ Discussion of model validation and testing approaches
	Classification: Neural Network	▪ Introduction to neural networks and related inputs, outputs, and models
Day 2	Classification: Logistic Regression	▪ Introduction to logistic regression and related inputs, outputs, and models
	Linear Regression/Numeric Response	▪ Overview of linear regression, a supervised learning technique that provides a continuous variable as output
	Unsupervised Learning	▪ Overview of unsupervised learning, which is marked by the absence of a response or outcome variable ▪ Discussion of unsupervised learning techniques, including cluster analysis and association rules
	Text Mining	▪ Overview of preparing and structuring text to extract information from unstructured text ▪ Analysis of the resulting term-document matrix ▪ Discussion of compatible analytics models that can be applied on prepared text ▪ Hands-on practice with a text mining problem
Day 3	Analysis Practice	▪ Practice identifying business problems, data mining problems, and applicable modeling techniques with real-life scenarios

For the initial delivery of the class, we have been using the data sets provided with the textbook as our practice data. Our goal is to transition the class to internal data sets and business problems as we identify opportunities to do so, as we believe that internal examples will be more interesting and motivating for our employees.

LEARNING AS WE GO

We continue to refine the class based on the results and feedback. We discovered that hands-on practice, where employees are able to work on as many data sets and create as many analytics models as possible, yields better results than listening to hours of lectures.

We also discovered that the trainer and the class setting add value. Employees can embark on self-study by reading textbooks. However, self-study does not provide the explanations, interpretations, and interactions that the trainer and other employees provide.

We hope to reach a 100,000 advanced analytics-capable employee organization. To be successful, we believe that some form of web-based training that enables self-study will be necessary. We started developing that web-based training in the first quarter of 2014.

TRAINER BACKGROUND

For trainers delivering the Introduction to Data Mining class, we look for individuals with practitioner-level knowledge, which is roughly the equivalent of taking one graduate course on data mining. This knowledge can also be acquired through self-study. The study of other relevant subjects, such as statistics, machine learning, math, science, and engineering can also add to the instructor's background. We have found that though the material is presented at an undergraduate level of difficulty, it is helpful for the instructor's knowledge or experience to be more advanced in order to answer the questions that arise during class.

To identify a trainer, we interview candidates to assess their motivation and background before a class. The candidates then attend the class to see the class in action and participate. Afterward, the new instructors attend a two hour train-the-trainer class that covers the course pacing, learning objectives, and points to emphasize.

CONCLUSION

We believe that applying BI and advanced analytics to big data can be a source of competitive advantage for Intel. No longer viewing advanced analytics as the domain of a few data scientists, we want to expand that knowledge to other teams in the enterprise. In particular, business analysts, BI teams, and others can learn how to frame business problems as data mining problems, enabling them to be more effective contributors in data analysis projects.

Our approach focuses on advanced analytics skill development, which includes developing and delivering training classes, mentoring, building a community of practice, and identifying instructor candidates. To further support these skills, we are designing an analytics VM for learners to practice on and use for small-scale data sets. We have basic participation metrics and plans in place for developing success metrics. As we continue to build skills, we are also thinking about how to enhance the BI process with the advanced analytics process.

With Intel's substantial investment in the tools and technology for mining big data, we believe that investing in employees' analytics skills is equally important. These skills can empower employees to be better contributors and in turn can help the promise of big data be fully realized in the enterprise. We look forward to engaging in conversations with other leaders in the industry who are considering or implementing their own efforts to bring advanced analytics to a wider audience.

RELATED INFORMATION

Visit www.intel.com/IT to find content on related topics:

- "Managing Data for Intel's Enterprise Private Cloud Infrastructure"
- "Mining Big Data in the Enterprise for Better Business Intelligence"
- "Roadmap for Transforming Intel's Business with Advanced Analytics"

For more information on Intel IT best practices, visit www.intel.com/IT.

ACRONYMS

BI	business intelligence
CRISP-DM	Cross Industry Standard Process for Data Mining
ETL	extract, transform, and load
MOOC	massive open online course
VM	virtual machine