



THE
MODELING
AGENCY

Advancing the Analytics-Driven Organization

A Strategic Framework for Leading a Goal-Centered Analytics Practice

Presented by Keith McCormick

Content Sample | Please Do Not Distribute

One Oxford Centre
301 Grant St, Ste 4300
Pittsburgh, PA 15219

Phone: +1 (281) 667-4200, Ext 3
training@the-modeling-agency.com
www.the-modeling-agency.com



THE
MODELING
AGENCY

Advancing the Analytics-Driven Organization

A Strategic Framework for Leading a Goal-Centered Analytics Practice



Keith McCormick

Senior Consultant
keith@the-modeling-agency.com
(281) 667-4200 x233
@tma_analytics

NOTES



THE
MODELING
AGENCY

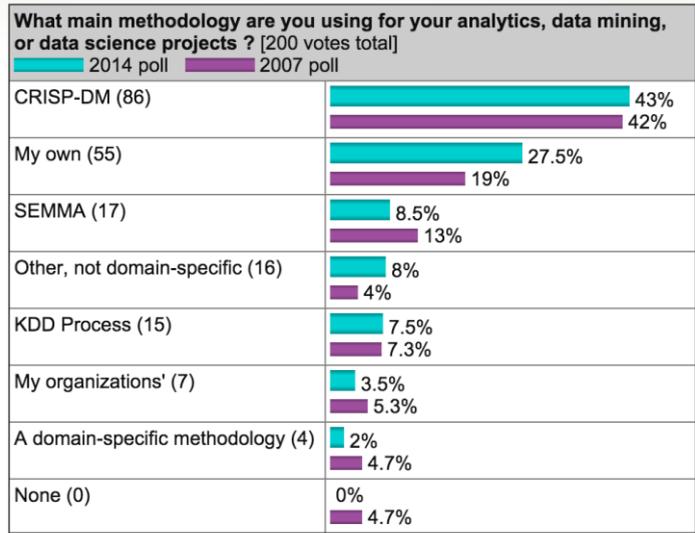
Predictive Analytics

Process Models

NOTES



Predictive Analytics is a Process



NOTES

This survey has been conducted a number of times. Even though it uses a “convenience sample” it is often cited because the web site is popular with the appropriate audience. The results over many years have been consistent. One of the reasons for its ongoing popularity is that CRISP-DM seems to have taken hold with the open source crowd.

Some processes have been formalized and are generalized. Others are specialized processes and some are home-grown.

Notice that three of the top four processes are formal processes (CRISP-DM, SEMMA and KDD) and that the CRISP-DM has a greater than 40% share.

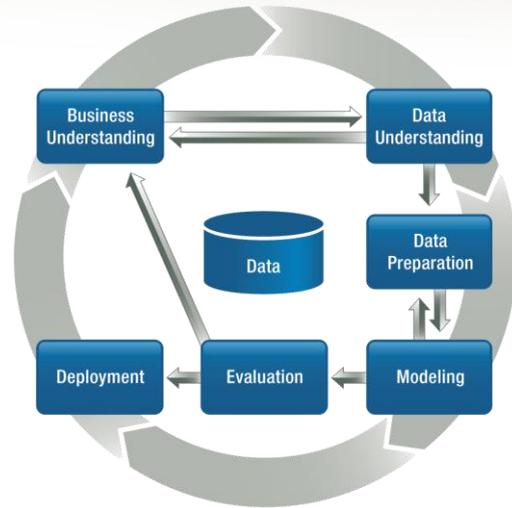
As we’ll see, the top three formal processes are similar at their cores. However, there’s some features of CRISP-DM that makes it more appealing. However, all of the process methodologies stop at the project-level and do not take into consideration many dynamics and aspects inherent in analytics at the larger organizational level.

Source: KDNuggets.com



Predictive Analytics is a Process

- **C**Ross-Industry **S**tandard **P**rocess for **D**ata **M**ining
- Developed by a consortium in 1999 (led by SPSS, NCR, and DaimlerChrysler)
- Great popularity due to:
 - Generic Adaptability
 - Comprehensive Scope
 - Greatest Depth



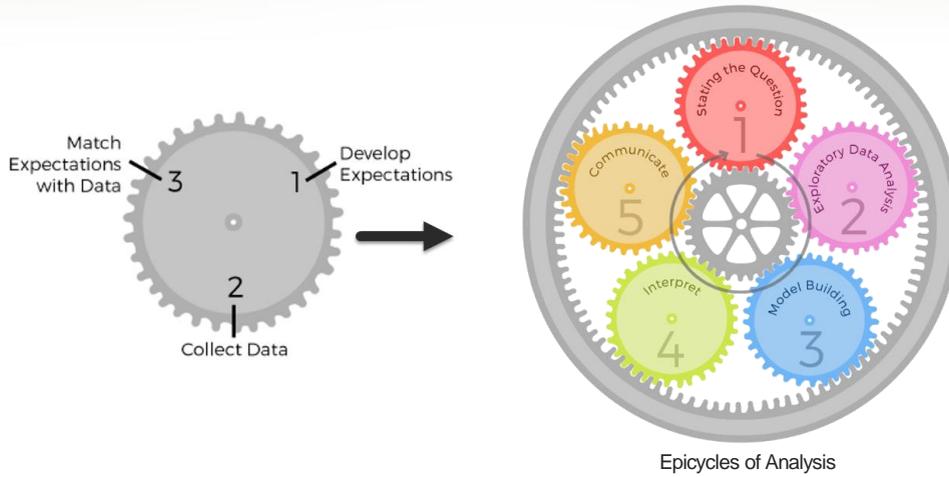
NOTES

The CRISP-DM process is the leading process because it's generic ... it can be adapted to any industry or application. But it's also the most highly-developed.

The diagram we see here defines an iterative process of six major steps at it's highest level.

CRISP-DM is definitely not a waterfall process. There is a presumption that although the sequence of the phases is meaningful that you will frequently go back to earlier phases. It is a powerful model but on your very first project you might get a little confused by a step in an earlier phase that depends on a task in a later phase. Veterans of project know the model well and incorporate this into their project plan.

Predictive Analytics is a Process



NOTES

In the Art of Data Science Roger Peng and Elizabeth Matsui address this iterative nature in a fascinating way. They propose a five step process that gets repeated three times. For a total of 15 steps.

The Art of Data Science

A Guide for Anyone Who Works with Data

Roger D. Peng and Elizabeth Matsui

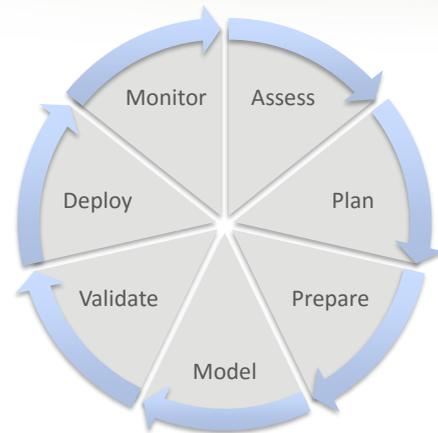
<http://leanpub.com/artofdatascience>

This version was published on 2016-04-08



MPF™ Motivations Over CRISP-DM

- TMA's expansion of the popular open-source CRISP-DM process model for strategic implementation
- Emphasis on goal-driven analytics at the enterprise / practice level
- Better delineation between assessment and planning – as well as data understanding and data preparation
- Improved structure for repeated tasks
- Addresses dynamics and complexities specific to larger organizations:
 - Organizational culture & attitudes
 - Full analytic team construct
 - Various role functions & collaboration
 - Mindset shift & change management



TMA Modeling Practice Framework™

NOTES

The Modeling Agency is an enthusiastic fan of the CRISP-DM process model, and had followed it for decades.

With advances in the organizational use of analytic modeling; the scale, variety and velocity of data; and the need for sustained modeling activity at a practice level, has motivated both the expansion and simplification of the overall process model design.

TMA's Modeling Practice Framework™ (MPF) accommodates recent environmental developments and streamlines overall project development with greater strategic emphasis at the practice level. TMA's motivation is to guide clients to establish sustainable predictive modeling factories and reinforce the skills with which to maintain a sustainable, measurable and actionable predictive modeling practice.

Remember, as with all conceptual frameworks: "the map is not the thing mapped." The framework is intended as a guide for action -- and not as a silver bullet, cookbook or rigid procedural guide to be blindly followed.

While most people (especially those new to the modeling process) will find the MPF to be useful as a comprehensive guide for establishing a residual modeling process or practice, every modeling project will have unique and highly situational aspects where many parts of the Framework will not apply or precisely align in sequence,

Experienced practitioners will develop the acumen and soft skills to understand how to apply and adapt any general process or framework to the realities and challenges of each situation. TMA's Modeling Practice Framework™, formal training and consultative oversight will guide in adapting the Framework into a tailored process to achieve sustainable and actionable outcomes for each specific modeling environment and organizational goal.



MPF is More Linear than CRISP-DM

CRISP-DM's highly iterative nature keeps it streamlined, simple, and elegant ...

But the "completion" of a phase only to return to it later can be confusing to the uninitiated and to management.



NOTES

- ✓ CRISP-DM's iterative nature keeps it streamlined, simple, and elegant.
- ✓ That same iteration prompts you to repeat phases multiple times which can be confusing to the uninitiated and to management.
- ✓ All predictive analytics projects are iterative and unpredictable. But a more linear approach helps in writing up formal project plans and managing teams.
- ✓ Also, explicitly listing seemingly similar phases that naturally repeat the MPF can clarify who is doing what and when.
- ✓ There is a "cost." Naturally it has more explicit tasks than CRISP-DM.
- ✓ The greater detail makes it easier to build a project road map from the MPF

For instance, CRISP-DM describes the Data Preparation Phase as:

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection as well as transformation and cleaning of data for modeling tools. (underline added)

In contrast, the MPF has the Explore Data and Verify Data Quality task in Plan, and then explicitly repeats a similar task in Prepare, Explore Integrated Data.

Why? The tasks are done at different times, frequently by different people in different roles, and the nature of the data at the different stages colors the task to the point that it is qualitatively different.



THE
MODELING
AGENCY

Modeling Practice Framework™

TMA's Practice-Level Expansion of CrISP-DM

NOTES



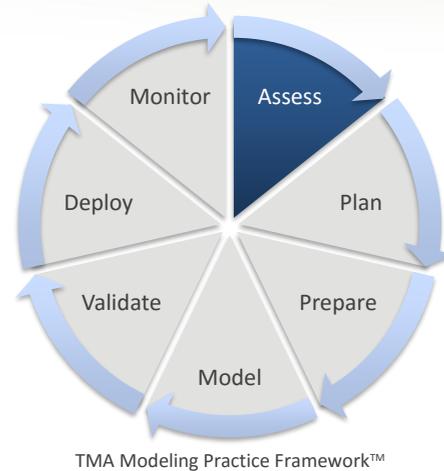
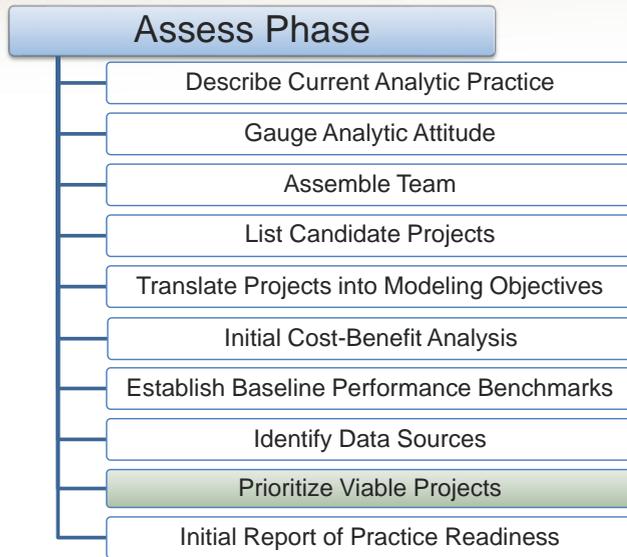
THE
MODELING
AGENCY

Assess Phase

NOTES



MPF™ Assess Phase



NOTES

ASSESS PHASE

- 1 Describe Current Analytic Practice** | What are we starting with? What's already in place? What's working and what has been tried?
- 2 Gauge Analytic Attitude** | How do the key players define and perceive the role of predictive analytics? Who can be counted on for support? Where are possible causes and sources of organizational resistance?
- 3 Assemble Team** | Our very first step is to understand all of the existing personnel who either contribute to- or benefit from- the analytic effort.
- 4 List Candidate Projects** | The initial list of projects is defined by pain points that need addressing and where it is perceived predictive analytics can help. They are initially listed in the language of the organization - not in modeling terms.
- 5 Translate Projects into Modeling Objectives** | Before potential projects can be vetted they have to be translated into modeling terms. Some problems will translate better than others which is part of the vetting process.
- 6 Initial Cost-Benefit Analysis** | While the specific benefit, in dollars, may not be clear or known at this stage, the size of the problem can be measured and weighed with other factors.
- 7 Establish Baseline Performance Metrics** | It is important to reduce current performance to writing, for all to see and agree to, so that you can determine if progress was made at the end of the project.
- 8 Identify Data Sources** | Don't rule out data, yet, on ease of access. List and describe everything that might possibly be available. If there are known logistical problems then identify them.
- 9 Prioritize Viable Projects** | Weigh initial cost benefit analysis, the current performance, and possible data sources against all candidate projects. It is important to vet several, and not simply coronate a single project proposal.
- 10 Initial Report of Practice Readiness** | This report and assessment is especially important if launching a new analytics practice, but is always wise. Take everything that has been learned thus far, especially the current state of the practice and analytic attitude. Are you ready to proceed? Reserve the right to pause, and address practice level issues before launching the project.

Assessment Qualifiers



- Brainstorm a few ‘interesting concepts’ into 12 to 20 potential projects
- Each project should run through all qualifiers
- The few that survive are the lowest-risk projects on which to engage
- The next step will move the low-risk viable projects to highest-impact for the organization

NOTES

During the Assess Phase, all desired projects should be considered. The Assessment Phase overlays numerous qualifiers against each potential project. Those that survive into the center of the overlay are viable. Those that fall short of qualifying should be placed into a holding tank where they may await certain criteria or steps to reach viability.

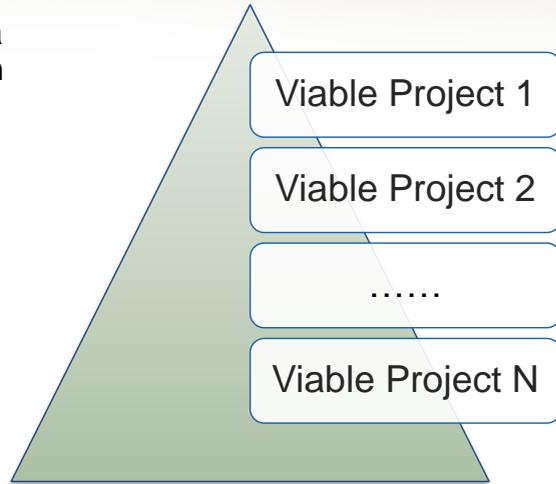
When TMA as consultants initially engage with an organization, we will typically find that there are one to three initial projects in mind. The Assess Phase encourages one or two dozen potential projects to be considered. So, we take the original few potential ideas, brainstorm them up to about 20, then disqualify the vast majority.

The result: we end up working on the few highly qualified survivors from a larger population rather than a few untested interesting concepts. Which would you rather proceed with?



Prioritized Viable Projects

- Prioritize viable projects by criteria most important to the organization
 - Proof of Value
 - ROI
 - Reduce Risk
 - Encourage mindset shift
 - ...
- Start with the lowest-risk / highest-impact viable projects



NOTES

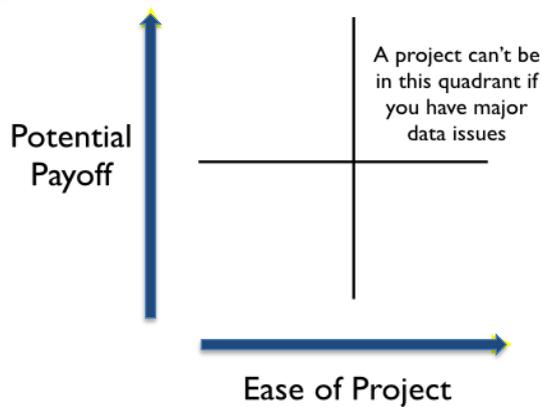
Projects that are deemed viable through the Assess qualification process may then be prioritized for the greatest overall impact – whether the success metrics be proof of value, ROI, reduced risk – or accelerating the organizations confidence in moving from gut-level to data-driven decisioning.

Very few organizations conduct this kind of qualifier at their own peril. They jump directly into data preparation and modeling to arrive at technically valid models – only to later discover that the project was strategically doomed before data was even selected.

Conversely, the Assessment Phase nearly cheats by eliminating failed projects in advance – then proceeding initially on the most meaningful or impactful projects. It’s like designing the answer key prior to taking the exam in the course that directly aligns with the end goal.

This approach pays untold dividends. Organizations who proactively and methodically qualify potential projects through a strategic Assessment exercise not only virtually guarantee success by eliminating challenges in advance, but concentrate the outcome by defining clear expectations and starting with the most potent project(s) up-front.

List Viable Projects



AADO Course

Copyright © 2000 – 2017 Keith McCormick and The Modeling Agency, LLC

NOTES

In many organizations, Predictive Analytics is launched with one very specific project in mind, usually with one specific metric or pain point in mind. In nearly every instance, there are multiple ways of framing projects around that organizational pain point. Also, there are always other viable projects. Considerable thought should be given to what these project ideas are. The perfect first project – which will set a precedence for future projects – is best chosen from a wider field. During the Assess phase, the job is simply to grow the list. Narrowing, ranking the list, choosing an initial project will occur in the Plan phase.

The team should not hold back in generating candidates in this phase. The Assess phase is about information gathering not about closing doors. The data experts in the meeting – usually IT – can describe data availability and give a rough description of what the data is like. Very possibly, they can project data dictionaries, or the like, on the screen for team members to see. However, it is important not to get bogged down and to stay at a high level.

There will be member of the meeting who will be difficult to assemble on a regular basis. This brief description should allow an experienced analyst to guesstimate 'ease of project'. Ease of project is 80-90% determined by data availability and data quality. The SMEs can also chime in as they have probably attempted some of these projects, and they can comment on potential road blocks.

The Sponsors – especially senior sponsors – are in a good position to perform 'back of the envelope' estimate of Potential Payoff. In fact, many sponsors can bucket projects into Low, Medium, and High payoff without scratching some numbers out at all.

It is critical to note that the best first projects is not necessarily the easiest first project. 'Low hanging fruit' that only yields less than a salary year of one of the team members is rarely the best first choice. It is understandable why many are tempted to go for the seemingly easy low yield project, but these projects are never as easy as they appear to be. You must take Potential Payoff into account.



Current Performance Benchmarks

- Unplanned manufacturing line stoppages due to equipment failures occur every 18 days, on average
- Hospital readmissions within 30 days of discharge for cardiac care patients is 32%
- Customer retention for loyalty card customers is 68% per year

NOTES



Define Critical Success Factors

- Decrease unplanned manufacturing line stoppages by 80% per year
- Decrease hospital readmissions with 30 days of discharge by 50%
- Increase customer retention by 10% and increase average monetary value of loyalty card customers by 15%

NOTES



Define Deployment Options

- What is the metric to be returned by the modeling effort?
 - A propensity score for each customer?
 - A probability of occurrence?
 - Identification of attributes of a customer/patient/piece of equipment?
- How will the metric be deployed?
 - In an existing database to be used in batch processing?
 - As an alert to an operator?

NOTES



Definition of Analytic Stages

Analytics Stage	What does it do?	Source of ROI
Stage 1 Automated ROI =188%	Automates reporting	Increased productivity for data analysts and reduced workloads for IT department
Stage 2 Tactical	Multiple deployments organized around business units	Uses analytics to answer business questions and improve decision-making
Stage 3 Strategic ROI =968%	Uses analytics to align daily operations with organizational goals	Pervasive use of analytics embedded into all business processes
Stage 4 Predictive ROI =1,209%	Intensive use of organizational data and data sources often external to the organization. Data is varied, often large, and may change rapidly. Frequently uses "Big Data" – massive volumes and varieties of data integrated for best predictive capability	Uses data to understand consumer sentiment and deploys personalized sales and marketing campaigns. Predicts individual consumer demand and creates varying products and services to precisely match demand. Often the platform for innovation and new product development.

NOTES

While a number of companies have defined their own version of an “analytics maturity model”, one study attached ROI calculations to each stage of their model.

A study titled, “The Stages of an Analytics Enterprise”, by Nucleus Research divided analytics into 4 stages – automated, tactical, strategic, and predictive. You can see the definitions of each, and associated ROI, in the summary slide here. Note Stage 4, use of predictive analytics returns a whopping 1,209% ROI.

This stage often uses so-called “big data” – data from a variety of external sources integrated with internal customer and product data to predict and deliver personalized products and services. We will talk about big data more in a few minutes.



THE
MODELING
AGENCY

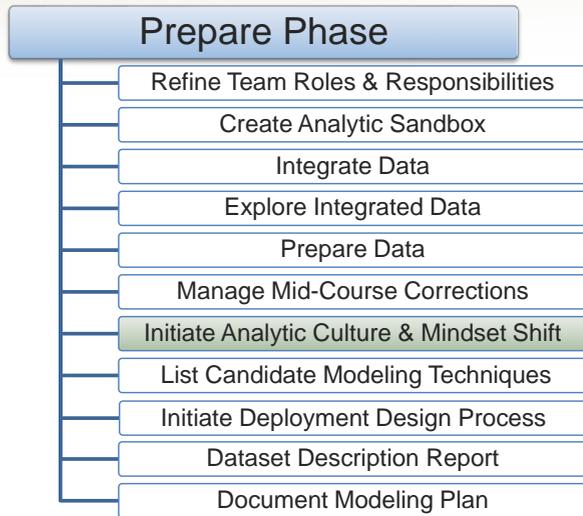
Prepare Phase

Modeling Practice Framework

NOTES



MPF™ Prepare Phase



TMA Modeling Practice Framework™

NOTES

PREPARE PHASE

1 Refine Team Roles & Responsibilities | Without a project plan, it is impossible to assign specific tasks. The team may change, expand, or even shrink, now that a concrete plan is in place. Team members whose availability, departmental affiliation, or skill set do not align to this particular project may work on parallel or future projects.

2 Create Analytic Sandbox | A sandbox, specific to the project, should be created. It is not refreshed. For bandwidth and security purposes a copy of the historical data is made. Within this sandbox the historical dataset is prepped into the form necessary for model building. IT and the modeling team must collaborate.

3 Integrate Data | Typically 6 to 20 sources or data, or more, are combined representing both internal and external sources, and multiple departments within the organization. The ETL requirements are often non-trivial and IT often gets involved.

4 Explore Integrated Data | Tasks similar to those done in Explore Data and Verify Quality, but different in character, are performed on the single combined data file.

5 Manage Mid-Course Corrections | While seemingly technical and tactical tasks are being performed time lines are being adjusted, management briefings are held, and goals are constantly realigned with reality. Poorly managed projects “go dark” during the very lengthy Prepare phase, but at their peril.

6 Prepare Data | Tasks include: Cleaning, Formatting, Constructing, Extracting and Enriching data. It is the most complex and labor intensive part of Prepare, which itself is the most labor intensive and complex phase of the entire process.

7 Initiate Analytic Culture & Mindset Shift | In Assess you identifies potential roadblocks to adoption present in the organizations culture. Now that a detailed plan is in place, it is time to prepare and educate those whose roles may be effected by the model. It is also a time to listen. A common strategic blunder is to what until the end to do this, and then it is too late.

8 List Candidate Modeling Techniques | Based upon the current state of the data and the nature of the problem to be addressed, what modeling techniques are both available and relevant?

9 Initiate Deployment Planning Process | Based upon the data, the analytic culture, technology issues, and the project’s mission what form of deployment seems to make the most sense? Do we have the necessary resources to do it?

10 Dataset Description Report | CRISP-DM describes this kind of report in some detail. It is the first of two reports that should be presented to the broader team before proceeding with the Model Phase.

11 Document Modeling Plan | This the last task of Prepare simply because this should get broad review and acceptance before actually proceeding with the Model Phase. It’s completion and even receipt of signature are often tied to contracts.



Analytic Culture / Mind Shift

- Understand that data and analysis can be a threat...
 - It is a shift from -
 - Expert driven, or
 - HIPPO (highest paid person's opinion) driven
 - To data-driven opinions and decisions
- Develop a change management plan
- The best change management is when “the boss” is seen making decisions based upon data rather than corporate politics, personal relationships, or other biases

NOTES



Data Preparation Law

Data Preparation is more than half of every analytic modeling process.

“The reason that data preparation is important, and forms such a large proportion of data mining effort, is that the data miner is deliberately manipulating the problem space to make it easier for their analytical techniques to find a solution.”

-Tom Khabaza

NOTES

The following is an excerpt from Tom Khabaza’s Nine Laws of Data Mining. A co-author of CRISP-DM, Khabza is an influential thought leader. Reading the entire Nine Laws is highly encouraged.

“While (data preparation) automation can be beneficial, there is a risk that proponents of this technology will believe that it can remove the large proportion of effort which goes into data preparation. This would be to misunderstand the reasons why data preparation is required in data mining.

The purpose of data preparation is to put the data into a form in which the data mining question can be asked, and to make it easier for the analytical techniques (such as data mining algorithms) to answer it. Every change to the data of any sort (including cleaning, large and small transformations, and augmentation) means a change to the problem space which the analysis must explore. The reason that data preparation is important, and forms such a large proportion of data mining effort, is that the data miner is deliberately manipulating the problem space to make it easier for their analytical techniques to find a solution.

There are two aspects to this “problem space shaping”. The first is putting the data into a form in which it can be analysed at all – for example, most data mining algorithms require data in a single table, with one record per example. The data miner knows this as a general parameter of what the algorithm can do, and therefore puts the data into a suitable format. The second aspect is making the data more informative with respect to the business problem – for example, certain derived fields or aggregates may be relevant to the data mining question; the data miner knows this through business knowledge and data knowledge. By including these fields in the data, the data miner manipulates the search space to make it possible or easier for their preferred techniques to find a solution.



The Data Preparation Phase

Key Considerations

- Where are all the data sources?
- How much cleaning does the data need?
- Do I need to create new data elements?
- How do I integrate data into one analytic file?
- How much data do I need to adequately support the modeling effort?
- Are there format issues to address?
- Can these processes be replicated in production?

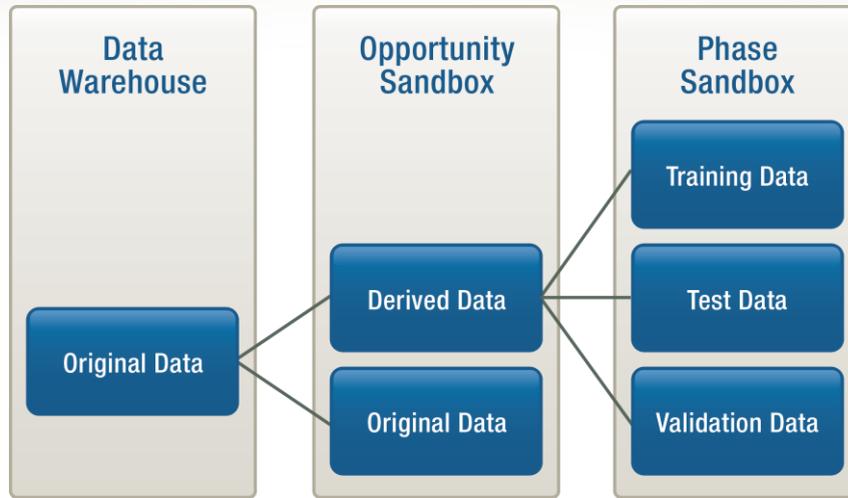
NOTES

Of all considerations, one that seems to be underestimated is the last point. Often a project is developed in an analytical environment and then put into production in another environment. So processes must be replicated in that environment.

One will be well-served to keep focus on the possibility of implementation issues.



Analytic Sandbox



NOTES

If your project plan is your blueprint, then data is your raw material. While you may have large quantities of data, you need to carefully consider whether you have the data necessary to support the decision process you are modeling, in a format that meets your needs, and in a way that can support not only your modeling effort, but also support your implementation plan.

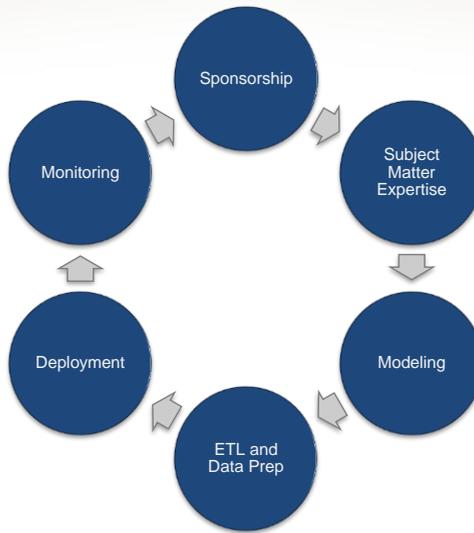


THE
MODELING
AGENCY

Building an Analytics Team

NOTES

Assemble Team - Project Roles



NOTES

These are the major functional role categories in a Predictive Analytics project. One person may take on multiple roles. More than one person may be involved in any role. Recognizing the complexity of both the business processes and the technology contributions is essential. A team approach that incorporates a complete perspective almost always provides a better solution than allocating the enhancement search to one individual. In a very real sense, it is virtually impossible to complete a project with one person.

Even if all technical roles are filled by one person (which is not rare), and even when that same person is the SME, the ‘customer’ of the project is rarely the same, and should probably not be the same. For the customer (internal or external) of the project to be the same as the PA expert is a bit like an MD treating themselves or a lawyer representing themselves. An arm’s length distance between analyst and the problem’s owner is wise.

The sponsor(s) bring the business problem to the team. The business problem is the whole reason that the project exists. It is important that the sponsor not merely delegate the problem to the team. A key points during the project the sponsor should participate in translating the business problem into a predictive analytics problem.

The Subject Matter Expert (also called the Domain expert) is critical to any Predictive Analytics problem. There is occasionally some debate as to whether, given the data, an SME is necessary. The argument goes that Data Mining competitions don’t give competitors access to an SME, so perhaps their role is not important. Dig deeper, and you’ll realize that those who attempt to argue this position are discussing only the Modeling phase. During the Assess, Plan, and Prepare phases and at other points throughout the process, SMEs are absolutely critical.

The Modeling phase is not the longest phase, but it does require someone with prior experience. Even if the other primary roles are filled by first timers, it would be wise to find a Modeler that has been through other projects. If they are the only one with experience, they may become a de facto project lead or they can be paired with a project manager.

Data Preparation dominates the workload of any Predictive Analytics team, but this is especially true of first projects. Predictive Analytics data simply isn’t in the same form as Data Warehouse data. They have different requirements and different constraints. It always requires modification. Someone on the team – sometimes the Modeler – must have the responsibility. In many cases, the roles are quite different, but when this is so, they must be highly collaborative.

It is tempting to postpone discussion of Deployment. After all, even after a model is built, it must be tested, and discussed pushing Deployment months into the future. However, some discussion of this topic must be made at the earliest stages even if this role is left vacant at the start of the project. Deployment is simply the integration of the Model results into existing business systems. It can range from very basic to extremely complex.

Monitoring is a role that is typically a low commitment part time role, checking to see if the model’s prediction begin to degrade. This will always happen eventually and someone has to be explicitly tasked with the monitoring. Very busy teams with many models may eventually make this a full time “model administrator” type role.

Project Team – First Meeting



- Sponsorship: who benefits from success?
- The ‘hands-on’ team
- Data management
- Subject Matter Expertise
- Who’s job may evolve when this is done?

NOTES

Who should attend a kick-off meeting?

Often the key Sponsor is quite senior in the organization – SVP, C-suite, etc. It is highly recommended that this key sponsor attend the first portion of a kick-off meeting. It is very helpful for them to describe, in their own words, the business problem so that the entire team can ensure that the necessary translation into a Predictive Analytics problem stay true to the original inspiration for the project. Their presence in the kickoff will often communicate very strongly the importance of collaboration between the sponsor department and other department because the senior sponsor is typically not at the departmental level. When the Sponsor is senior, their delegate – departmental level head, project lead, etc. - should also be present at the kickoff and for most of the project will represent the sponsor.

The ‘hands-on’ team members will be those tasked with using some form of Predictive Analytics software whether it be open-source, existing, or a recently acquired tool. If this is a first project, it may not be clear at first who these folks are going to be. It may not yet be known their comfort level or aptitude for what may be new skills. One should invite liberally at a kick-off meeting so that any potential members of the hands on team are present.

A good kick-off meeting will have data experts and their management. Typically this is an IT director that those members of the IT team that are most familiar with the relevant data. It is quite common that those that are most familiar are also least available – they are often the star member of the IT team. One client described a meeting attendee as being the one that “knew where the bodies were buried”. It is often very difficult for someone like this to be available to the team on a regular basis. If this is the case than the most likely IT collaborator with the team must also come to the meeting.

It is very common that you will not know which SME is ideal until after the first meeting. When this is in doubt, it is a good idea to invite someone at the middle management level that is familiar with several possible SMEs. A combination of expert resource may be necessary. As in the case of other “star” team members, the perfect SME is probably very busy. A good Predictive Analytics project manager stays highly aware of this and only asks them to attend meeting when it is useful and necessary the project.

The last category of attendees to a kickoff meeting



THE
MODELING
AGENCY

List Candidate Projects

Marketing

- Improve customer retention
- Decrease customer acquisition cost
- Increase promotion response rates

Operations

- Create 2-year demand forecast
- Decrease distribution costs
- Predict equipment failure

Customer Support

- Predict which customers will file complaints

Human Resources

- Decrease employee churn
- Improve employee productivity

Product Development

- Decrease time-to-market

NOTES

In many organizations, Predictive Analytics is launched with one very specific project in mind, usually with one specific metric or pain point in mind. In nearly every instance, there are multiple ways of framing projects around that organizational pain point. Also, there are always other viable projects. Considerable thought should be given to what these project ideas are. The perfect first project – which will set a precedence for future projects – is best chosen from a wider field. During the Assess phase, the job is simply to grow the list. Narrowing, ranking the list, choosing an initial project will occur in the Plan phase.

The team should not hold back in generating candidates in this phase. The Assess phase is about information gathering not about closing doors. The data experts in the meeting – usually IT – can describe data availability and give a rough description of what the data is like. Very possibly, they can project data dictionaries, or the like, on the screen for team members to see. However, it is important not to get bogged down and to stay at a high level.

There will be member of the meeting who will be difficult to assemble on a regular basis. This brief description should allow an experienced analyst to guesstimate 'ease of project'. Ease of project is 80-90% determined by data availability and data quality. The SMEs can also chime in as they have probably attempted some of these projects, and they can comment on potential road blocks.

The Sponsors – especially senior sponsors – are in a good position to perform 'back of the envelope' estimate of Potential Payoff. In fact, many sponsors can bucket projects into Low, Medium, and High payoff without scratching some numbers out at all.

It is critical to note that the best first projects is not necessarily the easiest first project. 'Low hanging fruit' that only yields less than a salary year of one of the team members is rarely the best first choice. It is understandable why many are tempted to go for the seemingly easy low yield project, but these projects are never as easy as they appear to be. You must take Potential Payoff into account.



Project Role “Hats”

	Support	Core Team	Management
Project Sponsorship		Internal Customer	Senior Sponsor
SME	End User		Process Management
			Process Expert
Data	Data Steward	Data Prep	IT Management
Modeling		Modeler	Architect
			Analytics Management
Deployment	Interface Support	Deployment Lead	Interface Management
Monitor	Model Manager		

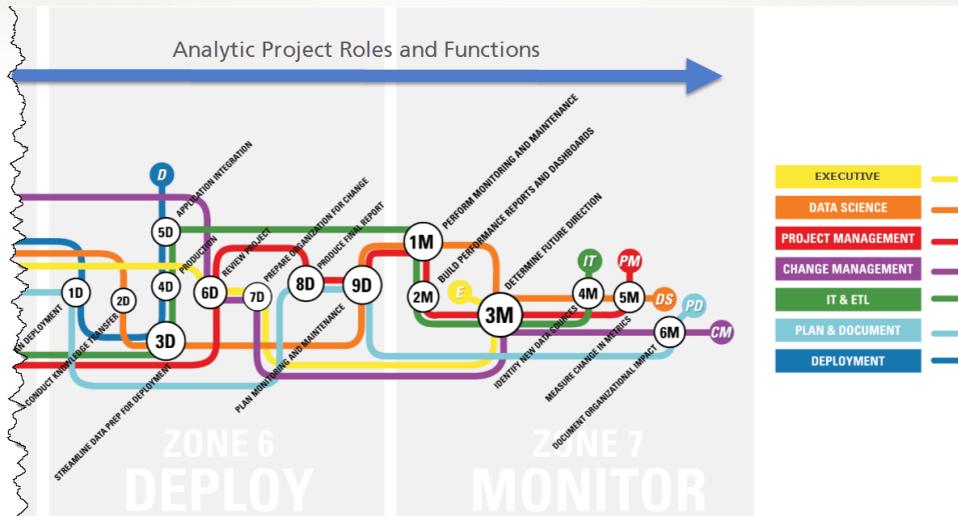
NOTES

There is a remarkable variety of roles on a project. While a project rarely has more than a ½ dozen active daily participants, with each wearing multiple hats, there are at least a dozen different roles, each requiring a different skill set. Some of these roles might require only attending a meeting or two. They do not all contribute equally. Some projects don't require some of these roles. This table identifies 15.

It is quite common for any one of these 15 project roles that they are filled by one person that is uniquely qualified to play that role either because of their prior experience, vertical knowledge, specialization in proprietary (or even confidential) processes, or highly technical expertise. However, it would be extremely rare to have all of these roles filled by a single individual.



Role Functions in Full Project Build



NOTES

All Phases methodically follow an application storyboard and implementation framework of an end-to-end project build.

Major project roles and functions run respective 'subway lines' through all Phases of a full implementation.



THE
MODELING
AGENCY

Additional Resources

NOTES



Resource Categories

Professional Organizations

Conferences

Data Mining Processes

Standards

Mega Sites

Course Supplements

- Online Video
- Techniques

Commercial Software

- Packages
- Database Integrated
- Specialty

Open Source Software

- Packages
- ETL / Data Integration
- Other / Related

Networking

- Social Networks
- Key LinkedIn Groups

Data

- Consumer
- Free Datasets

Recommended Reading

- Books
- News and Newsletters
- Blogs

Helpful Tools

NOTES