International Journal of Information Management 38 (2018) 45–51

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Research Note

A framework for evaluating the analytic maturity of an organization

Robert L. Grossman[a,b,*]

[a] University of Chicago, United States
[b] Open Data Group Inc., United States

ARTICLE INFO

Keywords:
Analytic maturity
Deploying analytic models
Analytic operations
Analytic governance
Analytic infrastructure

ABSTRACT

We introduce a framework called the Analytic Processes Maturity Model (APMM) for evaluating the analytic maturity of an organization. The APMM identifies analytic-related processes in six key process areas: i) building analytic models; ii) deploying analytic models; iii) managing and operating analytic infrastructure; iv) protecting analytic assets through appropriate policies and procedures; v) operating an analytic governance structure; and vi) identifying analytic opportunities, making decisions, and allocating resources based upon an analytic strategy. Based upon the maturity of these processes, the APMM divides organizations into five maturity levels: 1) organizations that can build reports; 2) organizations that can build and deploy models; 3) organizations that have repeatable processes for building and deploying analytics; 4) organizations that have consistent enterprise-wide processes for analytics; and 5) enterprises whose analytics is strategy driven. The APMM is broadly based upon the Capability Maturity Model that is the basis for measuring the maturity of processes for developing software.

1. Introduction

It is rare today for an organization to develop software that is critical to its products, services or operations without a software methodology being used; on the other hand, it is relatively common for an organization to build analytic models that are critical to its products, services or operations without using any analytic methodology.

We introduce a framework for evaluating the analytic maturity of an organization that consists of assigning an *Analytic Maturity Level or AML* score from 1 to 5. The higher the score the more likely that the organization's processes for building and deploying analytic models will result in analytic models that: i) that are statistically valid and are completed according to schedule; ii) can be deployed into an organization's products, services or operations; and, iii) meet the organization's goals for the model.

The framework is based on common challenges that organizations face when developing and deploying analytic models:

- Problems obtaining the data necessary for building models.
- Problems deploying models into an organization's products, services and operational systems.
- Problems quantifying the business value generated by models.
- Deployed models do not bring the business value that was expected.
- A lack of repeatability when building models.
- A lack of repeatability when deploying models.
- A lack of repeatability when testing and evaluating models.
- Difficulty integrating different models developed across an organization to meet the requirements of the organization as a whole.

There are also several common confusions that organizations face:

1. Not understanding the difference between reports generated from data and models built from data.
2. Not understanding the difference between models built from data and business rules.
3. Not understanding the difference between the *outputs* of models and the *actions* and *business processes* required so that products, services and operations achieve a desired business goal.

The greater the analytic maturity of an organization, the more likely that it is for an organization to meet these challenges and not face these confusions.

We note that there is no standard terminology yet in the discipline of analytics. For some time, the first confusion above has been described as the different between reports (or business intelligence reporting) and predictive models. More recently, developing reports from data that summarize the data has been called descriptive analytics, while the term predictive analytics has been used when statistical models are built from data, especially when these are used to make predictions about future events. Recently, the term prescriptive

* Correspondence to: Center for Data Intensive Science – KCBD 10142, University of Chicago, 900 East 57 Street, Chicago, IL 60637, United States.
E-mail address: robert.grossman@uchicago.edu.

http://dx.doi.org/10.1016/j.ijinfomgt.2017.08.005
Received 25 February 2017; Received in revised form 22 August 2017; Accepted 25 August 2017

analytics has begun to be used when the outputs of predictive models are used to derive actions that have business value (which is related to confusion 3 above). Prescriptive analytics has also been used more generally when optimization and related techniques are applied to the outputs of predictive models.

It may be helpful to look at the framework introduced here for evaluating the analytic maturity of an organization from the viewpoint of information or knowledge management. At a high level, one can think of analytics as using data to build models and then deploying the models within an organization's products, services or internal processes to achieve a desired outcome, such as increased revenue, higher retention, lower costs or reduced risks. If we think of knowledge management as the process of capturing, distributing, and effectively using knowledge, then the analytic framework described here relates to knowledge and practices around analytic models, including not only building and deploying them, but also related strategic, governance, security, privacy and risk issues. In the broader context of information management, the framework involves the information management required to manage the data and model assets associated with building, deploying and evaluating models.

2. Background

2.1. Analytic models, infrastructure and operations

The framework introduced here is based upon a few basic concepts: analytic models, analytic infrastructure and analytic operations. We describe each of these in turn. In addition, the framework specializes more general processes related to strategy, IT governance, and security and compliance to those specifically focused on analytics. We use the terms: analytic strategy, analytic governance, and analytic security and compliance for these specializations. See Fig. 1.

2.1.1. Analytic models

By analytic models we mean statistical or data mining models that are empirically derived from data using generally accepted statistical methodologies.[1] For simplicity, we generally use the term *model* below instead of *analytic model*. In contrast to model, we use the term *rule* to refer to a manually derived "if-then" statement that sets a variable or take a specified action based upon the "if" clause of the statement. Although some models can be described by one or more if-then statements, the essential difference is that with models statistical algorithms are used to create the if-then statements while with rules humans manually create rules.

2.1.2. Analytic infrastructure

Analytic infrastructure refers to the software components, software services, applications and platforms for managing data, processing data, producing analytic models, and using analytic models to generate business value through taking actions, making recommendations, and generating alerts (Grossman, 2009).

2.1.3. Analytic operations

Analytic operations refers to the various processes that result in the outputs of analytic models being used to make decisions and to take actions relevant to the business or enterprise, such as increasing revenues, decreasing costs, or improving operations. Analytic operations ensure that the results of analytic models are integrating into an organization's products, services and operations.

[1] We borrow this terminology from the Fair Credit Reporting Act (FCRA), codified at 15 U.S.C. Section 1681 et seq., which requires a credit model to be "empirically derived, demonstrably and statistically sound."

2.1.4. Analytic strategy

Although there is an extensive literature on strategy and there are several articles that stress the importance of analytic strategy, we have not found a commonly accepted definition of analytic strategy. For the purposes of the APMM, we define *analytic strategy* as the long term decisions an organization makes about how it uses its data to take actions that satisfies its organizational vision and mission; specifically, the selection of analytic opportunities by an organization and the integration of its analytic operations, analytic infrastructure and analytic models to achieve its mission and vision.

Note that this definition is modeled on a standard definition of corporate strategy, which is sometimes defined as "Strategy is the direction and scope of an organization over the long-term: which achieves advantage for the organization through its configuration of resources within a challenging environment, to meet the needs of markets and to fulfill stakeholder expectations" (Johnson, Scholes, & Whittington, 2017).

2.1.5. Analytic governance

As with strategy, although there is a large literature on IT governance, there is no commonly accepted definition of analytic governance. A common definition of IT governance is (Brown & Grant, 2005): 1) Ensure that the investments in IT generate business value. 2) Mitigate the risks that are associated with IT. 3) Operate in such a way as to make good long-term decisions with accountability and traceability to those funding IT resources, those developing and supporting IT resources, and those using IT resources. This suggests that the goals of Analytic Governance should include:

1. Ensure that good long-term decisions about analytics are reached and that investments in analytics generate business value.
2. Operate in such a way that data, derived data and analytic products are protected and managed in a secure and compliant fashion.
3. Operate in such a way as to make sure that there is accountability, transparency, and traceability to those funding analytic resources, to those developing and supporting analytic resources, and to those making use of analytic resources.
4. Provide an organization structure to ensure that the necessary analytic resources are available; that data is available to those building analytic models; that analytic models can be deployed; that the impact of analytic models is quantified and tracked; and that data, derived data and data products are managed in a secure and compliant fashion.

2.2. Building and deploying models

In general, models do not generate value for an organization until they are deployed. They can be deployed into products, into services, or to improve operations. For this reason, it is helpful when evaluating the analytic maturity of an organization's processes to be aware of certain choices when building models.

Usually, a modeling group using statistical or other specialized applications develops an analytic model. Once the model is developed, the IT group deploys the model into the appropriate product, service or operational system. Since they are two environments (the modeling environment and the deployment environment) and two teams (the modeling team and the IT team), it is important that there be an efficient mechanism for moving the models between the environments. There are a few common approaches:

1. The same application may be used in both environments. This approach is sometimes used but not very often since applications that are designed to be used by modelers are not in general designed to be deployed into operational systems.
2. The model may be applied to data in the development environment to produce a table of outputs, which are then loaded into a database

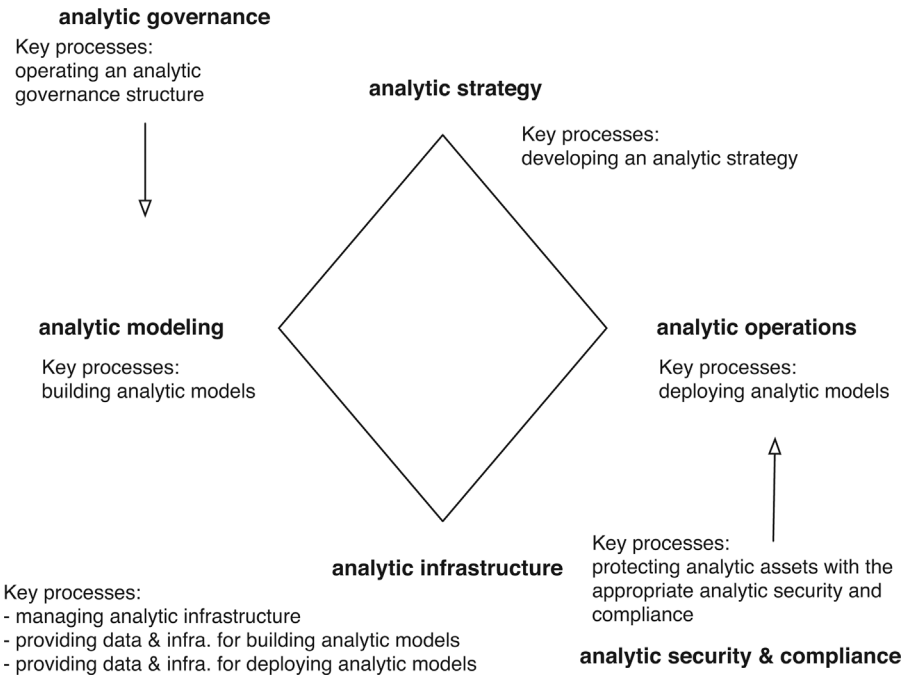

Fig. 1. Key processes for analytics can be divided into five functional areas: analytic modeling, analytic operations, analytic infrastructure, analytic strategy, analytic governance, and analytic security and compliance.

in the deployment environment. Since most deployment environments have integrated databases, this is a popular option.

3. The model is exported in the development environment and imported in the deployment environment. In this case, the model is expressed in some language or format, such as SQL (Date & Darwen, 2017), the Predictive Model Markup Language (PMML) (Data Mining Group (DMG), 2017a), the Portable Format for Analytics (PFA) (Pivarski, Bennett, & Grossman, 2017), or a custom language developed by the organization.
4. The model may be manually coded by the IT team and integrated into the deployment environment. In this case, the model will be expressed using a computer language, such as Java, Python, C or C++.

2.3. Model producers and consumers

Since approximately 2000, a not-for-profit consortium called the Data Mining Group has been developing a language for encapsulating statistical and data mining models called the Predictive Model Markup Language or PMML (Data Mining Group (DMG), 2017a). Models were initially defined in terms of XSDs and later with XML schemas. A central tenet of the DMG approach for defining PMML is the recognition that: 1) Models should be considered first class objects and defined with a language (such as XML or JSON). 2) There should be an architecture that separates cleanly environments in which models are produced, such as development environments, and environments, in which models are consumed, such as in products or in operational systems. PMML can be exported by applications running in the first environment and then imported by applications running in the second environment. With this approach models can be easily updated in operational environments simply by having the application read the PMML file. More recently, the Data Mining Group has developed a JSON based format for describing analytic workflows called the Portable Format for Analytics (PFA) (Data Mining Group (DMG), 2017b). PFA can be used describe the compositions of analytic models, as well as the pre-processing and post-processing commonly required when deploying analytic models (Pivarski et al., 2017).

2.4. Life cycle management for analytic models

Analytic models usually have a life cycle: 1) building models; 2) deploying models; 3) refreshing models; 4) rebuilding models; 5) retiring models. We have already discussed the first two. Refreshing usually refers to the process of re-estimating models using new data, while rebuilding models is the more labor intensive process that can add additional features, additional sources of data, change the model type, etc. Decisions for refreshing, rebuilding and retiring models are made based upon the performance of the deployed model and business requirements for the deployed model. See Fig. 2.

3. Analytic processes maturity model (APMM)

3.1. APMM framework

The APMM described here is based upon a framework for analytics that divides analytic processes into six areas:

1. *Building analytic models.* The process of building an analytic model takes as the input: i) data and ii) the appropriate business requirements and produces an analytic model as the output.
2. *Deploying analytic models*. These set of processes take an analytic model that has been developed and integrates it into an organization's products, services, and operations in such a way as to deliver the desired business value.
3. *Managing analytic infrastructure*. Managing the IT infrastructure required to build and deploy analytic models (as mentioned above we call this *analytic infrastructure* following (Grossman, 2009)) has historically been challenging for many organizations, but is becoming even more so as the volume, velocity and variety of data grows. Analytic infrastructure includes both the IT infrastructure for building and deploying models.
4. *Operating an analytic governance structure*. Operating a governance structure to support analytics is critical since those selecting analytic opportunities, building analytic models, deploying analytic models and managing data required for building and deploying analytic models are usually in different organizations. Without analytic governance, it is difficult for most organizations to successfully build and deploy analytic models, much less do this with a

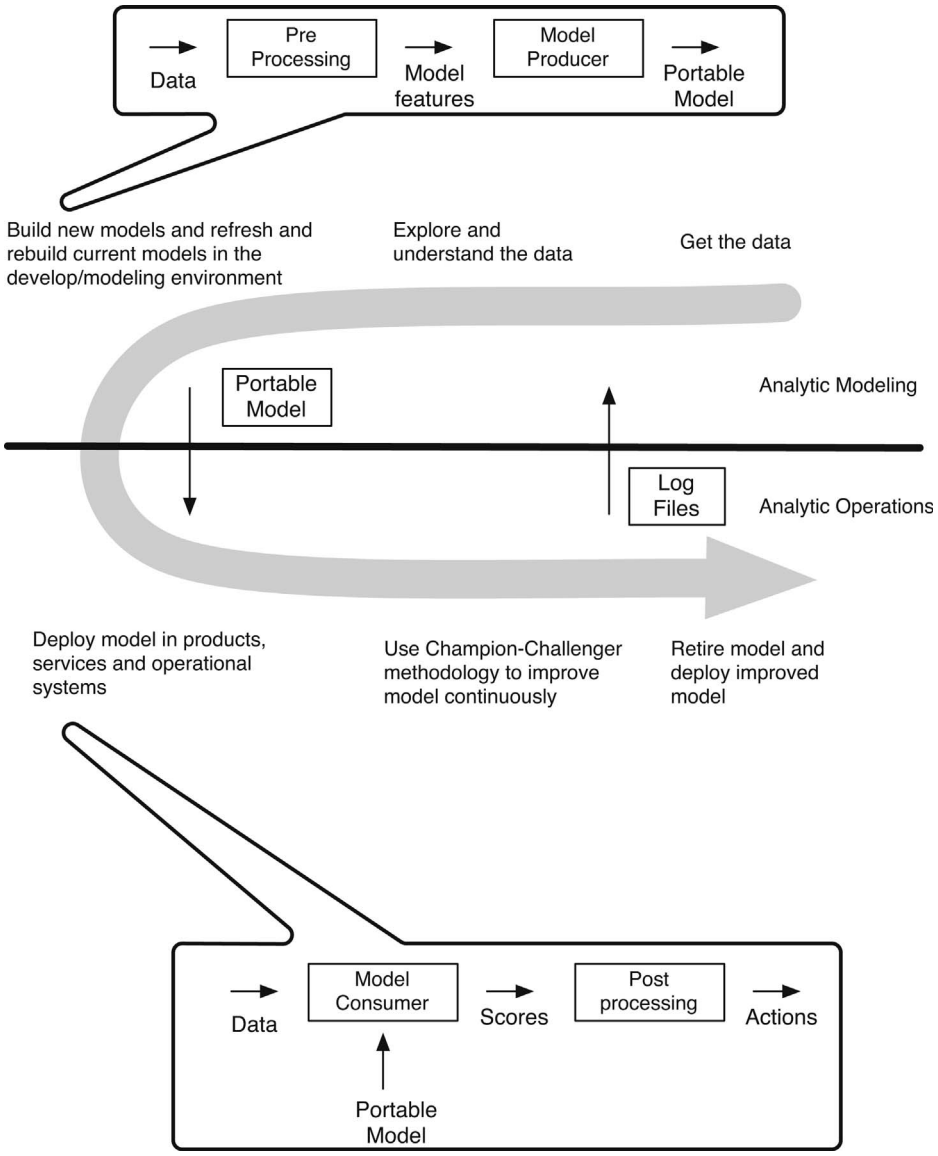

Fig. 2. The life cycle of analytic model.

repeatable process (Fig. 3).

5. *Providing security and compliance for analytic assets.* A growing priority of IT organizations has been IT security. Analytics and big data present some additional challenges, such as: i) protecting data privacy when side channel attacks on data are growing increasingly easy; ii) managing analytic infrastructure for big data, which can be so large that manual processes for infrastructure provisioning are no longer adequate; iii) and following appropriate security and privacy procedures when working with third party data. Setting up appropriate security and compliance processes to reduce risk, protect analytic assets, and to satisfy any required regulations is an important component of a mature analytic organization.
6. *Developing an analytic strategy*. Developing an analytic strategy and using the analytic strategy to select appropriate analytic opportunities. Almost all organizations have more analytic opportunities than the resources required by the opportunities and the first set of processes involve selecting which analytic opportunities to pursue based upon the short and long terms requirements and opportunities of the organization.

3.2. Overview of analytic maturity model

We now describe the five levels in the Analytic Processes Maturity Model (APMM), which we call Analytic Maturity Level 1 through Analytic Maturity Level 5. We abbreviate *Analytic Maturity Level by AML*. With our definition of AML, an organization of maturity level n, must also have reached analytic maturity levels 1, 2, …, n^{-1}. The five levels are:

1. **Build reports.** An AML 1 organization can analyze data, build reports summarizing the data, and make use of the reports to further the goals of the organization.
2. **Build models.** An AML 2 organization can analyze data, build and validate analytic models from the data, and deploy a model.
3. **Repeatable analytics**. An AML 3 organization follows a repeatable process for building, deploying and updating analytic models. In our experience, a repeatable process usually requires a functioning analytic governance process.
4. **Enterprise analytics**. An AML 4 organization uses analytics throughout the organization and analytic models in the organization are built with a common infrastructure and process whenever possible, deployed with a common infrastructure and process whenever possible, and the outputs of the analytic models integrated together as required to optimize the goals of the organization as a whole. Analytics across the enterprise are coordinated by an analytic governance structure.

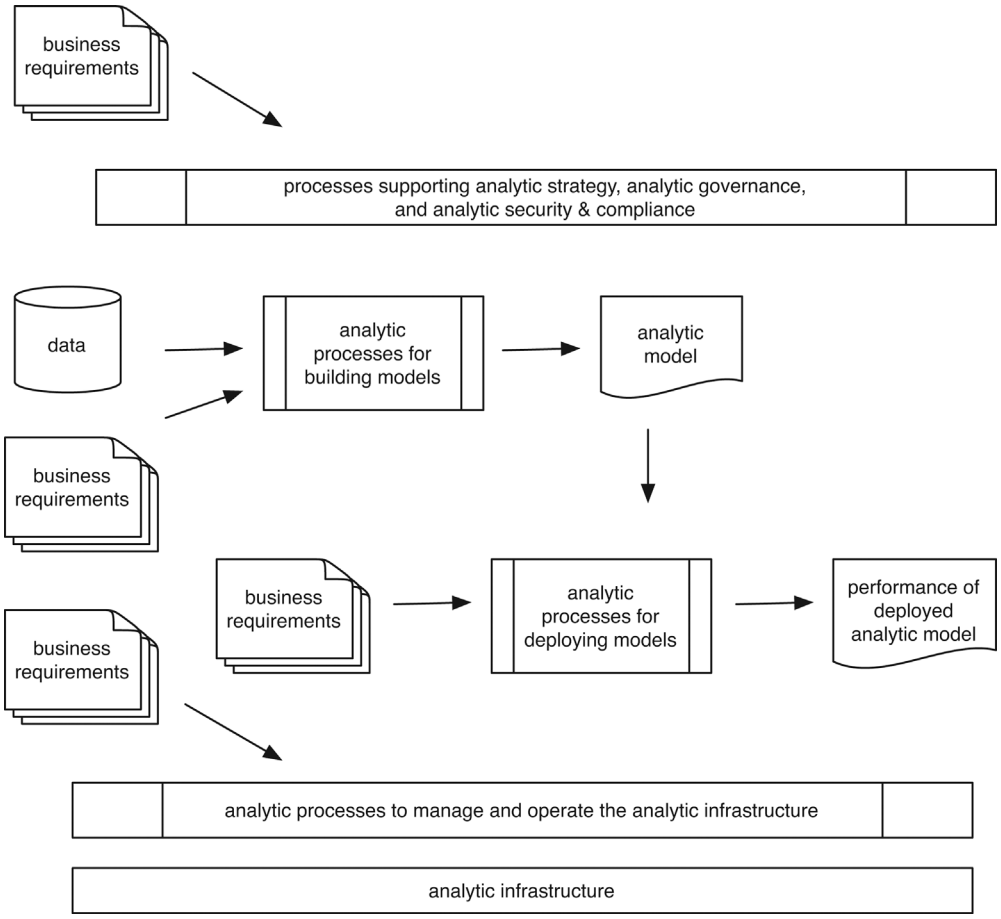

Fig. 3. Some of the analytic processes required for building and deploying analytic models.

5. **Strategy-driven analytics**. An AML 5 organization has defined an analytic strategy, has aligned the analytic strategy with the overall strategy of the organization, and uses the analytic strategy to select appropriate analytic opportunities and to develop and implement analytic processes that support the overall vision and mission of the organization.

3.3. Key process areas and goals

The Capability Maturity Model (CMM) (Humphrey, 1989; Paulk et al., 1993) uses the term *key process areas* or KPAs for related groups of processes. Using this terminology, the APMM is described in terms of the six KPA described above. According the CMM, an organization develops goals for each KPA as a basis for determining appropriate policies and procedures (Humphrey, 1989).

Based upon common practice, we have identified a preliminary list of goals for each of the key process areas. There is not a community consensus yet on what a definitive list of goals is for each of the areas, so we have limited the goals below to be what an expert generally versed in the field of analytics would typically consider a general accepted goal.

3.3.1. Goals for building analytic models

1. Models should be built from data ("empirically derived") and use generally accepted statistical procedures.
2. The performance of models should be quantified with metrics and a process developed so that new models can be developed that outperform the current models with respect to these metrics.
3. When rules are used that are not empirically derived using generally accepted statistical procedures, then the business, compliance, or other reason for the rule should be known and managed, and the performance impact on the model of the rule should be quantified if possible.
4. Models should be robust in the sense that small changes to the data result in substantially similar models.
5. The processes used to clean and transform data to create the features of models should be separately managed, automated and documented, as should any pre- and post-processing required.

3.3.2. Goals for deploying analytic models

1. The performance and business impact of the model in operations should be quantified and monitored on a regular basis.
2. It should be possible to update the model without writing code that impacts operational products, services or systems.
3. There should be a process for validating and verifying models before they are deployed broadly.
4. There should be a mechanism for checking that models are being used as per compliance policies. In particular, it is important that the models and the resulting actions as deployed in the operational systems are consistent with the organization's security, privacy, and data use policies.
5. Deployed models should be robust in the sense that missing or malformed data, delayed feeds, etc. do not disrupt the systems or operations associated with the model.

3.3.3. Goals for managing and operating analytic infrastructure

1. The analytic infrastructure for managing the data required for analytics should be adequate given the volume, velocity and variety of the data and the analytic objectives and strategy of the

organization.
2. The analytic infrastructure available to the modeling group should be such that the data required for building analytic models is available in a timely fashion to those that build the models.
3. The analytic infrastructure for deploying models should allow analytic models to be deployed efficiently and reliably into operational systems, products and services.
4. The analytic infrastructure should support the management of models over their entire life cycle.
5. The analytic infrastructure should integrate the security and compliance needed to protect the data as required.

3.3.4. Goals for operating an analytic governance structure

1. The analytic governance structure should include the groups responsible for building models, deploying models, and managing the analytic infrastructure, and include the appropriate stakeholders and business owners from these organizations.
2. The analytic governance structure should include *executive committees* that involve the appropriate businesses owners and stakeholders for making the decisions required so that the analytic strategy developed can be developed and executed.
3. The analytic governance structure should include *technical committees* for evaluating and making recommendations on analytic processes and technology that span more than one group or impact more than one stakeholder or business owner.
4. The analytic governance structure should include the necessary stakeholders, decision makers, and executives so that the policies required for the security and compliance for analytic assets can be developed and implemented.
5. The analytic governance structure should include a process for assessing the analytic competence of the organization and improving the analytic maturity of the organization.

3.3.5. Goals for developing an analytic strategy and for selecting analytic opportunities

1. Analytics should used by the organization to help differentiate itself from competitors and to provide a competitive advantage.
2. The analytic strategy should identify long-range analytic directions for the organization.
3. There should be a process for selecting analytic opportunities that optimizes the value to the organization as a whole, given the limited resources that most organizations have for building and deploying models.
4. The value brought by the analytic opportunities selected should be quantified and tracked.
5. The analytic strategy should manage data as corporate assets.

3.3.6. Goals for providing security and compliance for analytic assets

Our assumption is that the company or organization has a Chief Information Security Officer (CISO) and perhaps a Chief Compliance Officer or Chief Risk Officer and that the goals below are supplementary to the organization's security and compliance policies and procedures. Some specific goals for security and compliance related to analytics include:

1. Protecting data assets used in, and produced by, analytics should be integrated into the organization's security plans, policies, procedures and controls. By protecting data, we mean protecting the confidentiality, integrity and availability (Information Technology Laboratory (National Institute of Standards and Technology), 2004) of data assets.
2. The analytic group should work with the company's inside or outside counsel so that the collection of data assets, modeling practices, and the deployment of analytic models are compliant with all relevant local, state and national and international laws, regulations and policies.
3. As the size of data grows, it is important to make greater and greater use of automation and continuous monitoring (Dempsey et al., 2011) to ensure that data is being properly protected and relevant policies, procedures and controls are being followed.
4. Analytic security and compliance should cover not only analytics within the company, but also the compliance of data made available by the company to third parties through service contracts and other contractual relationships.
5. When the company sells data to third parties, then it is important that there is a system for monitoring how third parties use the data and whether it is consistent with the terms of the sale.

4. Discussion

4.1. Significant differences between organizations at different AMLs

The most important difference between Analytic Maturity Level 1 and Level 2 organizations is that AML 2 organizations can build models that make predictions about future events instead of summarizing past events. AML 2 Organizations know the difference between (business) rules and analytics and integrate both of them into deployed systems.

The most important differences between AML 2 and 3 organizations is that AML 3 organizations remove barriers to building models, such as when modelers to do not have easy access to the data. AML 3 organizations also remove barriers to deploying models.

The most important differences between AML 3 and 4 organizations is that AML 4 organizations have processes and structures in place, including a governance structure, so that there are *uniform processes across the organization* whenever possible for selecting analytic opportunities, efficiently building, deploying and integrating analytic models, and for reducing analytic. AML 4 organizations integrate analytic processes with other business processes throughout the organization to support the overall goals and objectives of the organization. AML 4 organizations also integrate analytic models from across the organization to support the overall goals and objectives of the organization.

4.2. Related work

A *Capability Maturity Model (CMM)* (Humphrey, 1989), was originally developed to quantify the level of an organization's business processes to develop software and complete a software project. The use of key process areas and key practices in the APMM was based on the CMM since it is one of the most widely used maturity models. Standards for CMMs are part of ISO 15504.

The APMM is also based upon a clear separation between model producers and model consumers. As mentioned, this separation is one of the foundations for the Predictive Model Markup Language (PMML) (Data Mining Group (DMG), 2017a).

There are some similarities between the APMM described here and what is usually known as the KDD or data mining process model (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). The KDD process consists of the following steps: i) selecting target data from the available data; ii) pre-processing the target data; iii) transforming the preprocessed data; iv) using data mining on the transformed data to produce patterns; and v) interpreting and evaluating the patterns to produce knowledge. The loop as a whole and individual components are repeated as necessary. Perhaps the most important difference between the APMM and the KDD process model is that the latter focuses on activities within the modeling or data mining group, while the former identifies activities that span the organization (IT, modeling, operations and strategy).

4.3. APMM and big data

There is as yet no commonly accepted definition of big data, but for

the purposes here, the following draft definition developed by a NIST working group is useful:

> Big Data refers to digital data volume, velocity and/or variety [,veracity] that: enable novel approaches to frontier questions previously inaccessible or impractical using current or conventional methods; and/or exceed the capacity or capability of current or conventional methods and systems. (bigdatawg.nist.gov)

For big data, analytic maturity becomes particularly important for several reasons. First, as the volume, velocity and variety of the data grows, having the appropriate analytic infrastructure grows in importance. Second, with big data, multiple models are more likely to be used. As the number of models grows (to hundreds or thousands or more), it becomes important to have an analytic infrastructure that can build, manage and deploy these models. Using a language for expressing models, such as PMML or PFA, becomes very important. Third, if big data is to have value to organization, it is because models built from it can be deployed into products, services and operations to increase revenues, decrease costs, reduce risk and optimize operations. The greater the analytic maturity of an organization the more likely this is to occur.

5. Summary and conclusions

It took some time for organizations to put in place formal policies for developing software and over time quite a few different methodologies for software development have been developed. Although there are strong opinions today about which software development methodology should be used, it is generally recognized as much more important that *some* software development methodology be used versus which *particular* one is chosen. This point of view is not yet common when developing analytic models. In this article, we have identified six key process areas in analytics and distinguished between five analytic maturity levels based upon the maturity of these processes.

Although it is fairly arbitrary how we choose to distinguish between the various analytic maturity levels in an APMM, the five levels in the APMM described here were chosen to separate organizations into different groups based upon the following distinctions:

- Organizations that build models versus using rules.
- Organizations that have a repeatable process for developing analytic models.
- Organizations that have an analytic governance structure that support repeatable analytics and enable analytics to be extended across an enterprise in a uniform fashion.
- Organizations in which there is an analytic strategy that drives analytics.

References

Brown, A., & Grant, G. (2005). Framing the frameworks: A review of IT governance research. *Communications of the Association for Information Systems (Volume 15, 2005), 696*(712), 712.

Data Mining Group (DMG), The predictive model markukp language (PMML). Retrieved from www.dmg.org on February 14, 2017.

Data Mining Group (DMG), Portable format for analytics (PFA). Retrieved from www.dmg.org on February 14, 2017.

Date, C. J., & Darwen, H. (2017). *A guide to the SQL standard: a user's guide to the standard database language SQL*. 1997: Addison-Wesley.

Dempsey, K., et al. (2011). *Information security continuous monitoring (ISCM) for federal information systems and organizations.* NIST Special Publication (NIST SP)-800-137.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., et al. (1996). From data mining to knowledge discovery: an overview. In M. F. Usama (Ed.). *Advances in knowledge discovery and data mining* (pp. 1–34). American Association for Artificial Intelligence.

Grossman, R. L. (2009). What is analytic infrastructure and why should you care? *SIGKDD Explorations Newsletter, 11*(1), 5–9.

Humphrey, W. S. (1989). *Managing the software process. The SEI series in software engineering, xviii*, Reading, Mass: Addison-Wesley494.

Information Technology Laboratory (National Institute of Standards and Technology) (2004). *Standards for security categorization of federal information and information systems, in FIPS pub 199, iv*, Gaithersburg, MD: Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology9.

Johnson, G., Scholes, K., & Whittington, R. (2017). *Exploring corporate strategy*. 2008: Financial Times Prentice Hall.

Paulk, M. C., et al. (1993). Capability maturity model, version 1.1. *IEEE Software, 10*(4), 18–27.

Pivarski, J., Bennett, C., & Grossman, R. L. (2017). Deploying analytics with the portable format for analytics (PFA). *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)* (pp. 579–588).