A Modern Approach to Powerful, Transparent and Collaborative Modeling

An increasingly complex world demands a different type of modeling solution - one that prioritizes scope, collaboration, speed, accuracy, impartiality and transparency.

Unfortunately, that is not what permeates the data modeling world.

Most modeling tools are structured in a way that the subject matter expert is effectively excluded from the process of developing the model until after it is built — creating significant risk for the organization across a number of dimensions.

Why do organizations exclude subject matter experts when it leads to sub-optimal outcomes?

The reason has to do with how the modeling process works in most enterprises. Presently, 80% of the modeling timeline is dedicated to extracting and preparing data (ETL). What this means is that 20% of the modeling timeline goes into actually building and validating the model. Given the scheduling constraints associated with iteration, this leaves little time to explore the problem from a principled perspective. With no time and minimal business input, enterprises fall back on building an algorithmic black box because it is faster and easier than the alternative.

It doesn't have to be this way. Even within the time constraints imposed by lengthy ETL processes, robust models can be developed — they just need a better approach.

The Ayasdi Model Accelerator (AMA) is a software application designed to build models quickly, accurately, and perhaps most importantly in the regulated world of financial services — transparently.

This application, built on Ayasdi’s award-winning machine intelligence platform, is particularly well suited for complex, high-dimensional modeling problems where the goal is to principally reduce the number of features/variables being used for modeling purposes — thereby facilitating the process of building, documenting and deploying those models into production while creating exceptional transparency.

The application excels at a variety of modelling challenges but is particularly effective at risk modeling problems such as loss-given default (LGD), probability of default (PD), and other regulatory modeling problems. These models typically make use of simple regression techniques to support the simplicity and explainability requirements associated for regulatory oversight. However, AMA supports a range of approaches.

The AMA’s unique workflow is as follows:

1. The AMA starts by including pre-packaged transformations that derive from base data and are specific to these modeling problems, such as the calculation of log, square, delta lags, and so forth. These transformations create additional features that are detected and explained by Ayasdi’s machine intelligence approach. This is important as the model form (e.g. linear) is constrained in many regulatory settings. Without variable enrichment, it is exceptionally difficult to build accurate models. Since the AMA draws upon Ayasdi’s particular capabilities for distilling complexity, the addition of this additional resolution creates better outcomes where other modeling solutions struggle with high dimensional data.
2. In many cases, multiple models are required to achieve the modeling objectives. For example, the data might have a grouping variable (such as country) and the user might want to stratify the data based on the grouping variable and while considering constructing models for each strata independently. Ayasdi’s discovery capabilities can also be used to discover a stratification of the data in an unsupervised manner.

3. Ayasdi’s Model Accelerator then segments all variables— both base and transformed ones— into self-similar groups using semi-supervised learning. The subject matter expert can then let the application auto-select the best combination of features, or interact with an easy-to-use interface to select the variable(s) that are meaningful to the business in their estimation. This is as simple as checking a box. Every variable will have statistical significance and the list will be sorted by absolute correlation— but not every variable will be noteworthy to the business leader. Importantly, the application maintains a full audit trail of all the selections made by the software and the human operators. It should also be noted that the interaction between the data and business users is a forcing function to consider the various economic hypotheses and build intuition about the data and the space of models.

4. Finally, using the selections made by the business leader, the AMA will construct all possible models from among the down-selected variables— presenting a candidate model and a full suite of challenger models. Note that the form of the models created depends on the business challenge. AMA performs model validation by using the standard hold-out data methodology.

Besides dramatically reducing the time to generate candidate and challenger models, Ayasdi Model Accelerator automatically generates transparent and inspectable documentation at each step, supporting independent validation and regulatory compliance requirements.

This workflow may seem simple but there are several elements that make it the application of choice when modeling business complexity in a transparent and principled fashion. Below are some of the features that underpin AMA’s performance characteristics:

**VARIABLE GROUP DISCOVERY**

At the heart of Ayasdi’s Model Accelerator is a powerful approach to data science that encompasses both unsupervised learning capabilities as well as supervised learning capabilities. By leveraging the learners, AMA can identify groups of variables and explain why they were selected.
While the selection of an outcome variable presages an element of supervision, the AMA uses topological data analysis to look at all the possible relationships encoded within the enriched base-level data; finding hidden patterns that hold predictive value. These patterns are frequently obscured and cannot be detected by other approaches. By taking an unsupervised approach and looking at all the data enterprises can often validate or disprove organizational assumptions while becoming more sensitive to environmental changes impacting the business.

This approach also avoids modeling pitfalls such as multicollinearity and the tendency of forward or backward selection methods resulting in local minima.

While the AMA has unique capabilities in unsupervised learning, often model developers and line of business leaders want explicit control over the model through the selection of key variables. This approach is fully supported in AMA but goes one step further by documenting any bias associated with the selection of certain variables and the elimination of others. This leads to better models and greater transparency.

Given that most models are created by starting from the previous year’s model and adjusting based on the data they contain implicit bias associated with both variable and algorithm selection. While known initially, over time, this bias is often obscured through revision. AMA solves this challenge and can ensure that an enterprise has better insight into the behavior of models and their interaction with other models or processes within the business.
AUTOMATED DOCUMENTATION

During the model creation process, everything is documented automatically. The AMA institutionalizes both variable selection and modeling methodology, systematically and deterministically, to produce a repeatable process with consistent supporting reports on model lineage, variable selection and cross validation.

Not only are the initial selections recorded and documented, the entire modeling and approval process is tracked and cataloged for all model developers and collaborators—facilitating everything from review to model re-use.
COLLABORATIVE WORKFLOW

For the vast majority of modeling solutions on the market today, the only interface that a business user or regulator has to the model is PowerPoint or Word. This is particularly problematic for the business users. Those with the domain knowledge to think through the model, to guide its development and ensure its applicability to the business rarely ever interact with the model until the results are presented. Often it takes weeks or months to gain consensus using this approach of build, explain, evaluate as stakeholders must be scheduled to interact with the candidate model. This is sub-optimal on multiple levels and introduces risk for the organization that can be dramatically reduced or eliminated altogether.

In the AMA, carefully designed application workflow and collaboration features allow the model designer to anchor discussion while facilitating reviews with line of business owners and internal review managers on statistically defensible decisions. Once a base model has been created, a full suite of supporting or alternative models are being generated in parallel, enabling “real time” and interactive iteration while all stakeholders are in the same meeting.
ACCELERATING MODEL DESIGN AND DEVELOPMENT

The AMA can be used both as a model generation engine and as a model execution engine. This accelerates development cycle time considerably. In most scenarios, the model is created by a modeling or data science team, shared via PowerPoint or Word with the business owner and, if validated is sent to IT to be hand-coded, requiring a full software development lifecycle, which can be extremely expensive. With AMA’s workflow and collaboration features models can go from development directly into production.

BUILDING A PORTFOLIO OF MODELS

By default, Ayasdi’s software will present the candidate model (selected using Ayasdi’s unsupervised learning approach) as well as a number of viable challenger models.

This is particularly important in a regulatory setting where performance may dictate changing models over time. Further, the ability to demonstrate the rigor of evaluating challenger models is also of value to the enterprise when explaining the models.
GLOBAL AND LOCAL MODELS

Frequently, the modeling requirements demand both a global model and local models. This frequently occurs geographically (e.g. Asia and the component countries) but can be found in any number of situations. The local models have the same sets of variables as the global models, however the interaction among those variables can be fundamentally different. The AMA can successfully construct models at both granularities rapidly and with full documentation - enabling a more detailed understanding of both the entire region and the local markets.

CREATING OPERATIONAL FLEXIBILITY FOR MODEL DEPLOYMENT

The AMA can be deployed in number of configurations including high frequency configurations. Furthermore, AMA can output to other modeling execution engines that the enterprise has in production through industry-standard formats (PMML).

AMA IN ACTION: SIMPLIFYING CCAR/PPNR

The 2008-2009 financial collapse led to a Federal Reserve directive that banks with consolidated assets over $50 billion have additional risk assessment frameworks and budgetary oversight in place. To assess a bank’s financial foundation, the Federal Reserve oversees a number of scenarios (company-run stress tests). Referred to as the Comprehensive Capital Analysis and Review (CCAR) process, these tests are meant to measure the sources and use of capital under baseline as well as stressed economic and financial conditions to ensure capital adequacy in all market environments.

A Fortune 50 bank was consistently struggling to pass its annual stress test. The bank needed a way to rapidly create accurate, defensible models that would prove to the Federal Reserve that they could adequately forecast revenues and the capital reserve required to absorb losses under stressed economic conditions. The bank’s modeling approach left the business unit leads with little room and time to weigh in on the logic behind the choice of the variables selected. The result was that could not confidently defend the models that they included in the filings they presented to the Federal Reserve.

The bank selected Ayasdi to supplement its capital planning process. The process began with the leaders of the bank’s business units reviewing the macroeconomic variables stipulated by the Federal Reserve. Ayasdi enriched these variables using several techniques (e.g., time series transforms such as lags, differences, and percent changes) and created over two thousand variables. Ayasdi applied its machine intelligence software to rapidly correlate and analyze the impact of these variables on each business unit’s monthly revenue performance over a six-year period, uncovering statistically significant variables that were highly correlated with each business unit’s performance.
A comprehensive business review was conducted to screen the identified variables prior to inclusion in the models for each business unit. Ayasdi then conducted exhaustive statistical tests (including stationarity and multicollinearity tests) to validate these models’ ability to predict revenues for the business units.

The business leads then evaluated the candidate model and the challenger models, selecting those that best represented their business units. With a collection of accurate, transparent and defensible revenue forecast models, the bank could easily clear their most challenging regulatory hurdle.
Summary

The Ayasdi Model Accelerator (AMA) is a software application designed to build models quickly, accurately and, perhaps most importantly in the regulated world of financial services— transparently.

The key attributes of this modeling solution include:

- Automated feature selection to create bias-free, highly defensible models
- A consistent and predictable workflow to rapidly produce accurate, transparent models
- Designed for active collaboration and explainability with technical and non-technical counterparts alike
- Automatically generates documentation for internal validation and regulatory compliance
- Rapidly isolates an entire suite of defensible, high-quality candidate and challenger models
ABOUT AYASDI

Ayasdi is on a mission to help our customers gain transformative advantage through their big and complex data. Our revolutionary machine intelligence platform leverages automation, machine learning and topological data analysis to simplify the extraction of knowledge from even the largest and most complex data sets confronting organizations today and to facilitate the deployment of intelligent applications across the enterprise. Developed by Stanford computational mathematicians, Ayasdi’s unique approach to machine intelligence leverages breakthrough mathematics, highly automated software and scalable compute to revolutionize the process of converting big data into business impact. We are excited to count many of the Fortune 500, plus leading governments and research institutions as our clients and partners.