Why Predictive Analytics Should Be “A CPA Thing”
Table of Contents

Predictive Analytics: The Concept ............................................................. 3
The Challenges of Modeling ...................................................................... 4
Best Practices .......................................................................................... 5
The CPA’s Role ....................................................................................... 5
IMTA Resources ....................................................................................... 6
External Resources .................................................................................. 6
“Predictive Analytics” (PA) means using facts about the past to make inferences and then using these inferences to make assertions about the future. The purpose of this paper is to explain the value of PA and explain why CPAs are well positioned to be crucial participants in its adoption. The paper will highlight the growing importance of PA, the challenges of modeling and also define best practices, including how to initiate PA projects, how a CPA provides value and resources for further research and learning.
**Predictive Analytics: The Concept**

The primary difference between fact-based/historical reporting and predictive analytics is **inference**. Historical reporting (by definition) tells us where we are and where we have been. How much budget is left? What was the volatility of the portfolio last month? How has year-over-year sales revenue changed by product and region? This type of reporting is a summary of facts and any uncertainties in this summary typically stem from data accuracy, completeness and timing.

**Predictions** combine facts about the past with inferences to anticipate the future. How much budget will be left in three months given that my business units typically underestimate Q3 spending by 10%? How much is portfolio volatility likely to decrease next month given that volatility is at historically high levels and typically reverts to average levels? How will sales revenue of a core product change given the combination a recent slowdown but increased marketing expenditures on the product?

The simplest predictive model asserts that what has happened in the past will continue to happen in the future. Even a simple model of this type can create powerful explanatory value. For example, electrical consumption, retail spending and the commute time between two addresses in a city follow powerful and consistent cycles. In fact, much of the variation in these variables can be explained by the time of year (electrical consumption or retail spending) and time of day/week (commute time). Additional predictive or “independent” variables can be layered into a model to increase its stability and power and specialized modeling techniques can be used to extract more value. The application of predictive analytics ranges from these simple models that we use every day to complex multivariate models implemented in specialized systems such as SAS, R and SPSS.

The potential for improved decision-making has created growth in predictive analytics applications that is not restricted to private enterprise — as adoption by charitable organizations, government and regulatory bodies is increasing as well. Some common active uses of PA are audit risk models, actuarial science and insurance policy risk modeling, predictive failure of computer equipment and bridges, marketing analytics, customized coupon and advertisement placement, fraud risk modeling and stock and option valuation modeling.

However, the boom also has increased dependence upon and risk resulting from predictive analytics. Dependence on PA has grown because PA techniques provide higher predictive value than simple traditional techniques such as gut instinct, trend lines and flat reporting. Business risk from model failure increases as models are trusted for direct input into the decision-making process. The combination of these factors means that model failure is more likely to shake the house and there is an ever-growing need for business professionals who can help build, validate and responsibly use PA. Advanced business users with PA concepts fluency are needed to help build and validate models along with their IT partners. When decision-makers trust a prediction, it will be used to improve tactics and strategy, enhancing decision quality. Without this trust, investment in PA will be limited and inferences will be ignored.

“Trust” is a familiar concept to CPAs, who often are recognized as their clients or organizations most trusted advisers. Integrity, objectivity and subject matter expertise within business are optimal characteristics for effective product owners in the PA space. CPAs are disciplined consumers and producers of information and are adept at documenting the reasoning process. These behaviors are prerequisites for trust building.
The Challenges of Modeling

Data quality is one of the most significant reporting problems. Missing or biased data can result in erroneous reporting results and ultimately can lead to bad decisions. We have to understand and account for data quality issues before the results of our reports can be considered reliable. Perhaps the “sales revenue” column that we mentioned earlier contains values in different currencies. We have to find a way to convert these values to a standard base before our report is meaningful.

PA also depends on solid data quality but also is dependent on model choice. Let’s look at a simple model predicting future inflation as its five-year historical average. Is this model valuable? If there is a demonstrable relationship between inflation from period to period (and this is the case) then the use of a trailing average is reasonable. But should it be a five-year trailing period, a three-year trailing period or a one-year trailing period? How can one make this determination, and is it appropriate? There is nuance in finding relationships between the variable you want to predict (called the dependent variable) and the variables that will provide input to your model (called the independent variables). This requires specialized subject-matter expertise. It is necessary to show that independent variables are causal to the predicted outcome as opposed to being non-causally correlated.

Imagine that you are trying to predict the speed that a person is driving but that your model only contains information about the type of car that they own. The model might have some predictive power. However, adding a variable containing the number of speeding tickets the driver has received over the past three years may result in a more accurate prediction. Furthermore, adding this new independent variable to the model may illustrate that the variable containing the car type is not a significant predictor of speed.

This all sounds easy enough. But, in practice, it is difficult to build and validate a model. Some reasons for this difficulty include:

1. It often is difficult to understand causality.
   Does a fast car cause a driver to exceed the speed limit?

2. Data quality issues affect model results.
   One source rounded dollars up to millions but the other used standard rounding.

3. It often is difficult to obtain the data that you want.
   You did not collect historical data from one of the locations in your model.

4. The real world is the lab — you can’t always be flexible with experiment design.
   Would there have been fewer audit findings if you had used a different sampling system?

5. Independent variables affect each other.
   The type of customer who is complaining behaves differently during price increases.

6. The relationships between variables often are not linear.
   How much would a client care about a price change of 1%, 5% or 500%?

7. Relationships are not necessarily stable over time.
   Who is watching the same advertisement at 2pm and 6pm?

There are tools to test which model results in the best predictions and the stability of the model over time. This exceeds the scope of this paper. But the resources listed in the last section provide more technical information.
Best Practices

A tool that improves the outcome of high-stakes decision-making is the goal. A powerful and accurate tool has no value if it never is used but the worst possible case is an enthusiastic adoption of a bad tool. What are the best practices that reduce the risk of landing on the wrong end of this spectrum?

1. Wait for it (find the right opportunity).
   Successful mobilization depends on a strong clear case for predictive value. Create a well-reasoned business case that justifies the value of the project.

2. Advocate for a strong team that represents a wide range of perspectives and knowledge.
   A respected cross-functional team will enable trust in the product that they deliver.

3. Determine whether available data are sufficient to support the project.
   Consider quantity, continuity, completeness, accuracy and quality.

4. Advocate for or against the project based on its merits.
   It often is the case that several project participants have a disproportionate amount of sway with senior stakeholders. Speak truth to power and don’t underestimate your importance.

5. Assemble and clean data needed for the project.
   Garbage in, garbage out can seriously understate risk. If your project will inform business strategy, then it is, “Garbage in, nuclear waste out.”

6. Validate the model on historical “holdback” data.
   Always set aside data before developing your model and use this for validation. This helps prevent a serious pitfall.

7. Continuous validation and tuning of the model.
   Predicting the future based on the past depends on internally and temporally stable relationships in your model. Don’t assume that this is true — continuously demonstrate it and make revisions when necessary.

The CPA’s Role

CPAs are detail-oriented, trained to “prove” their assertions and document their work, cross-trained in multiple aspects of business and often are strong in one or more technical areas. These are the precise characteristics needed to champion, build and deploy solutions in the analytics space. The same methodical techniques that a CPA uses as an auditor (to determine material risk thresholds), a financial accountant (to analyze and draft footnotes to financial statements) or a tax practitioner (to research and implement optimal strategies) map to the skill sets needed to identify opportunities, implement, validate and use predictive models.

How does a CPA facilitate projects in PREDICTIVE ANALYTICS? Successful projects in predictive analytics typically begin with a written “business case” that justifies their launch. CPAs often are in a position to help identify new opportunities for the application of PA due to their general “situational awareness” of factors driving fundamental business performance. CPAs can contribute to the creation and validation of the business case for a project. After project mobilization, these same skills can be applied to identify, validate and cleanse data that the project will use.

Even the most accurate and powerful PA model is without value if it is not used. Perhaps the single most important role for the CPA in predictive analytics projects is that of trusted adviser. CPA involvement in PA projects increases the confidence level that key decision-makers have in the approach and implementation of predictive analytics. This confidence ultimately is what catalyzes tool adoption and enables the realization of what is otherwise only theoretical business value.
IMTA Resources

IMTA resources can be found at aicpa.org/IMTA.

- IMTA Web Page
- IMTA Store
- IMTA Membership
- IMTA Resource Center
- IMTA CPE and Events
- IMTA News and Publications
- IMTA Volunteer Opportunities

External Resources

- Case Studies
- “Using Predictive Analytics to Optimize the Premium Audit Process”
- Predictive Analytics Training and Testing Data Sets
- Predictive Business Analytics: Forward Looking Capabilities to Improve Business Performance