# The Five Myths of Predictive Analytics

Predictive Analytics is used by various organizations to analyze current and historical facts to make predictions about health care, financial collection activities, customer behavior, and customer retention. Although Predictive Analytics is a powerful optimization technique, it is often misunderstood, and thus misused. This paper examines common myths surrounding Predictive Analytics. Businesses are often under the misconception that Predictive Analytics and Predictive Models:

1. are new;
2. produce perfect prediction and are always the best technique;
3. are foolproof (good software tool = good models);
4. always deliver business results ; and
5. can be built and forgotten.

Myth: **Contrary to popular belief, Predictive Analytics is not a new technique.**
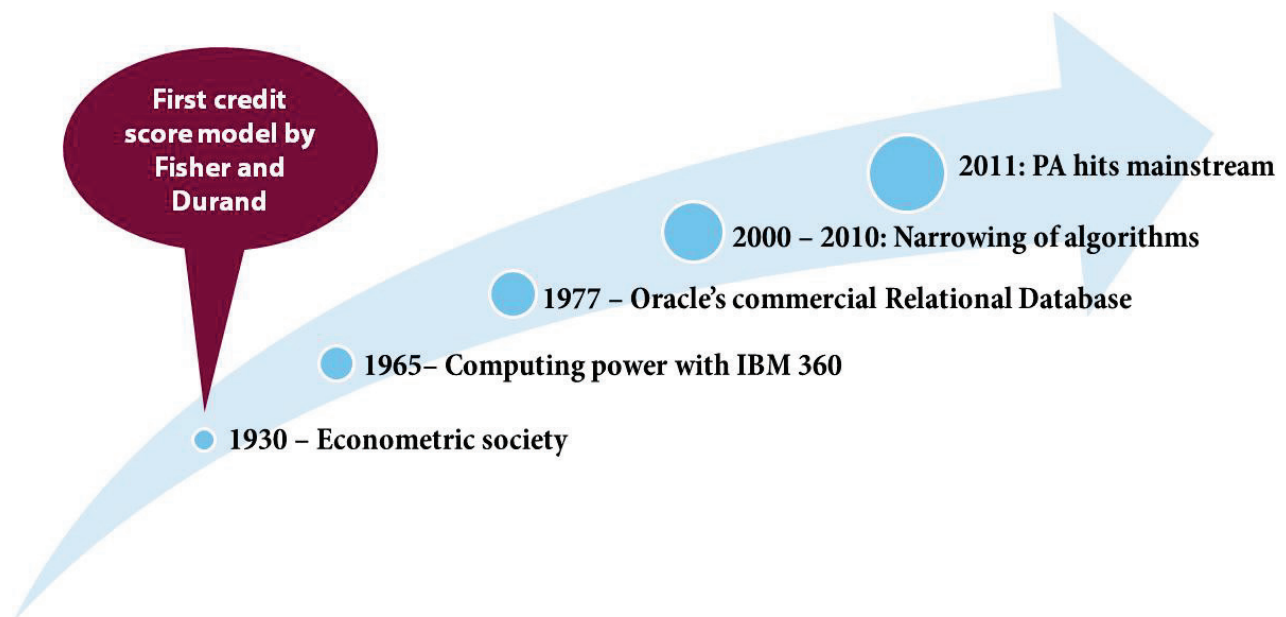
As demonstrated in the timeline above, beginning in the 1930s, models were built to make predictions about financial risk of credit holders. Over the years with increase in computing power and database development becoming mainstream, predictive analytics evolved to where we are using it quite expansively.

However, looking beyond this recent history, we can see its use centuries ago in the form of predicting good matches for arranged marriages. Today, we see the application of astrology in the western world when we read our horoscopes to see our daily predictions. This concept of astrology was first used in ancient India to create an individual's astrological chart, which contained 36 predictive attributes describing that person. These charts are based on mathematical calculations of input parameters, such as the birth time and birth place of individuals.

In the case of arranged marriages, the charts of prospective bride and groom are evaluated for compatibility and a higher rate of overlapping attributes predict a greater probability of a good marriage.

Bride: 36 Attributes   Groom: 36 Attributes

**Myth:** Predictive analytics produces "perfect" predictions and are always the best technique.

Even though we know that nothing in life is 100% guaranteed, we often overlook this fact of life in the way we use predictive analytics for our daily business use.

It's important to remember that even the best models are subject to significant misclassification. Illustrated below is an example of a model that is designed to predict the number of "green dots" in a box. As demonstrated, it shows misclassification where "red dots" as well as "green dots" are incorrectly predicted and classified.

Trying to predict "greens"
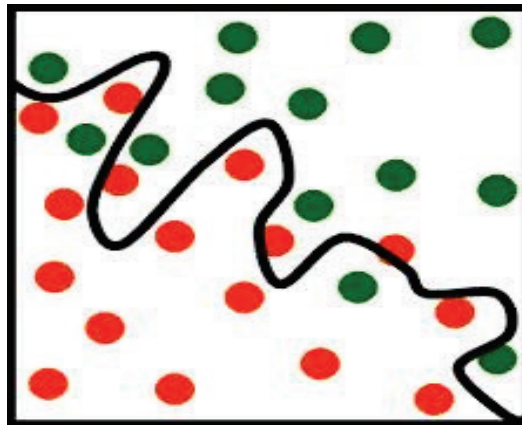
The other obvious attribute of predictive analytics models that we often overlook is that the models are estimations. The below decision tree illustrates a system for predicting risk associated with extending credit to individuals. The decision tree classifies risk on the basis of three variables (income, renter status and debt status). However, it does not take into account the multitudes of factors that would also impact risk. For example Bob, a struggling artist, with low debt ends up with extra alimony driving his total income above the $40K limit his classification changes from bad risk to good risk, which in reality is not the case.
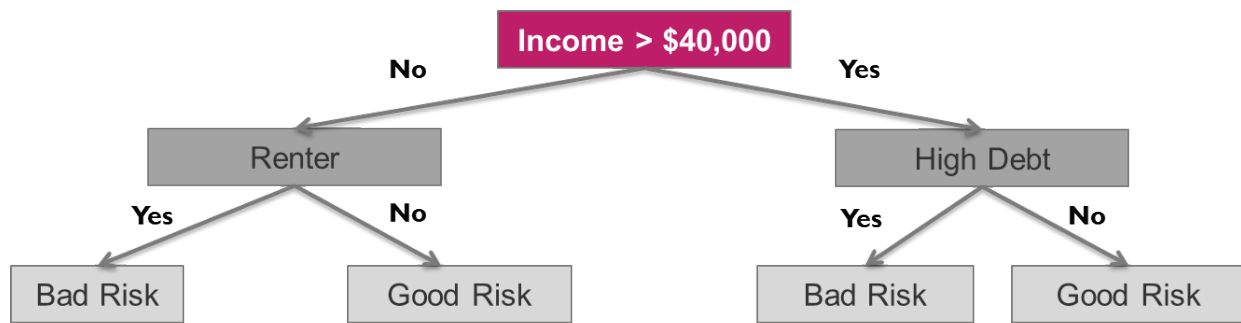
In addition to PA not being "perfect," the conditions under which one would want to use it need to be optimal. There is no need to jump on the predictive analytics bandwagon when there are other simpler techniques that are less resource intensive.

As illustrated below, PA is only effective when it is supported by management, increases ROI, and draws on existing historical data.
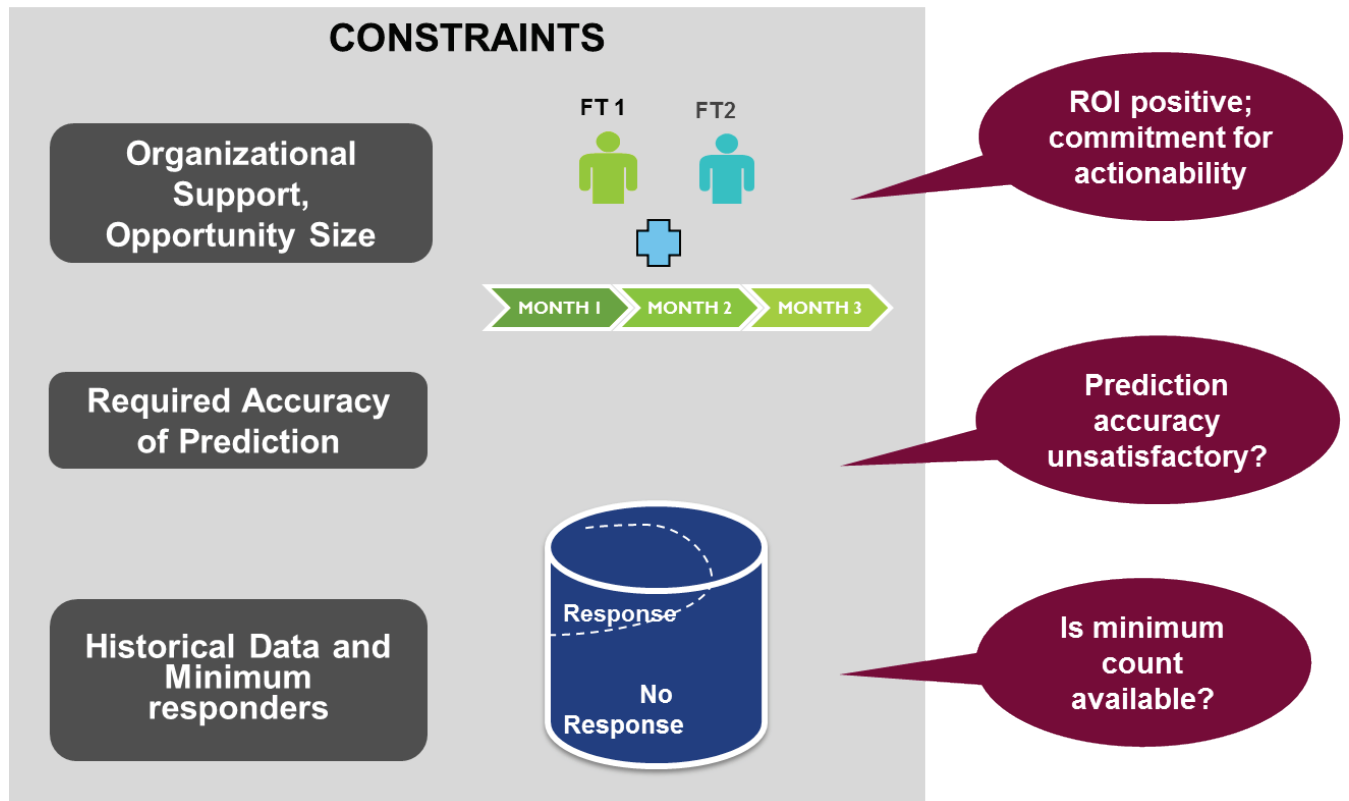
In the event we determine PA is not most appropriate, we can choose from a slew of simpler techniques as shown below that may provide us more bang for the buck, compared to advanced methods.
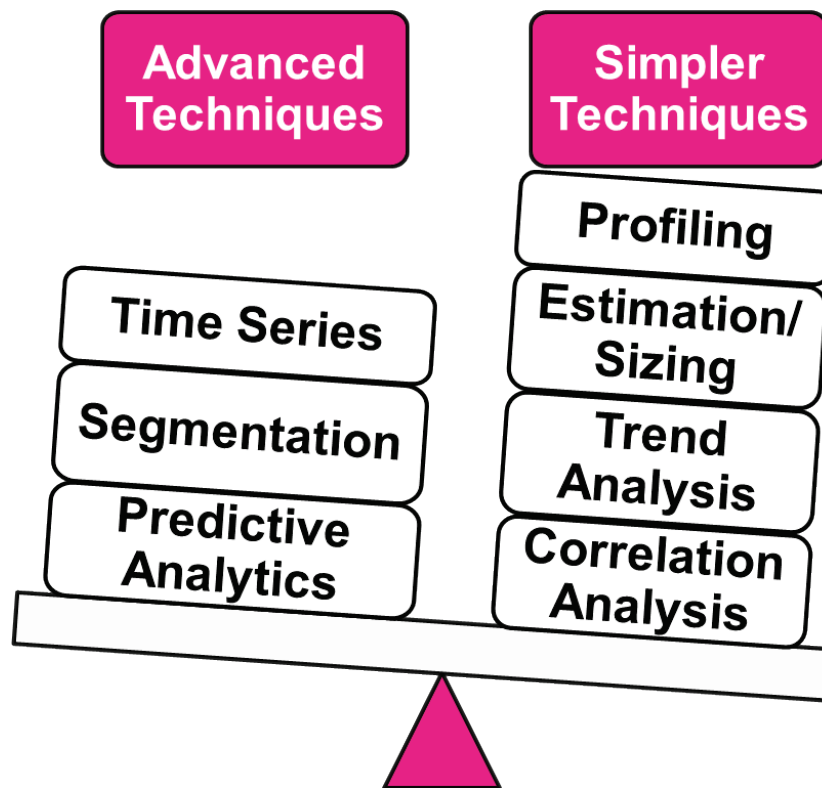
Myth: Predictive Models are foolproof, ie. Good software tools implies good models.

There is often a false sense of security that comes with using a good software tool. In our experience, building a good model is not a "press the button" solution. There are several key tasks to perform:

- **Data specification and pull**
- **Data cleaning and preparation**
- **Variable transformation and selection**
- **Model training and validation**
- **Choosing the best model based on business context**

In addition, there is a need to recruit team members with the appropriate skills and experience to execute a successful modeling project.

As a side, if one is considering developing their PA skills, we would suggest intensive hands on training of the following commonly used techniques as well as soliciting mentors to help develop models.
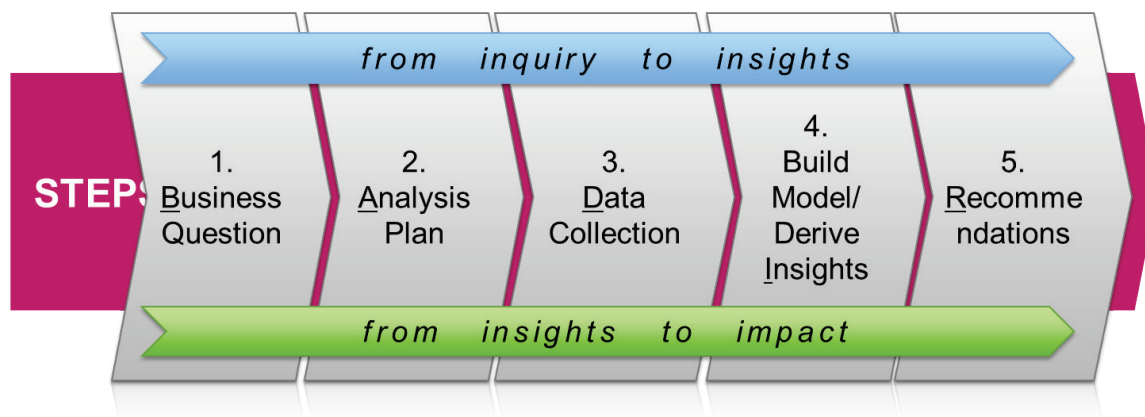
Commonly used Predictive/Advanced Techniques

Logistic Regression

Decision Tree

Linear Regression
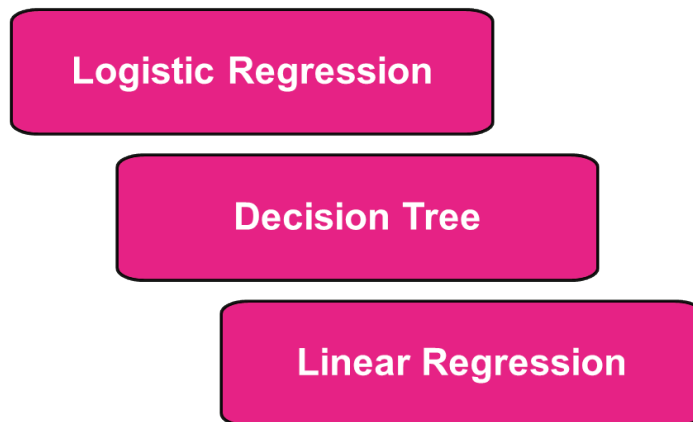
Other techniques: Clustering (K-means), PCA, Factor Analysis, Time Series, Survival Analysis, Neural Network

Myth: **Predictive Models always deliver business results**

There is a natural tendency among modelers to be in love with their models and to believe that that their models are going to generate incremental for the company and solve all the problems. In reality the projects that drive measurable business results encompass more than just good models. They typically incorporate an effective process such as Aryng's BADIR™: 5 steps from "data to decisions"™ framework.

Effective Predictive Analytics Framework

BADIR™: 5 steps from "data to decisions"™

from inquiry to insights

STEPS

1. Business Question

2. Analysis Plan

3. Data Collection

4. Build Model/ Derive Insights

5. Recommendations

from insights to impact

Step One: Business Question

Identify the real business questions behind an ask. This means closely validating with the stakeholders and understanding the impact of the problem.

Step Two: Analysis Plan

Use hypothesis driven planning to limit the scope of the analysis to only the core questions at hand. This allows for choosing the appropriate data and the correct analysis techniques.

Step Three: Data Collection

Collect data for the appropriate time period, applying rules to shunt the long tail in the data and performing a data audit.

Step Four: Derive Insights and Build Model

Analyze the data to gain insight into its significance as related to the hypotheses, build and validate model.

Step Five: Recommendations

The most important step which clearly defines concrete actions to address the business questions from step 1.

Myth: Can be built and forgotten

A common mistake that many organizations make is that once a model is built, it does not need to be proactively maintained and changed to keep up with the various changes in the surroundings. Some of the changes that would significantly impact models would be:

- ✓ Product, pricing, policy changes
- ✓ Competitive landscape change
- ✓ Economic condition changes
- ✓ Innovation …

For models to be relevant and effective, it is critical that it is updated and that modelers revisit the 5 steps BADIR framework to help them make the revisions.

Summary

Although Predictive Analytics is a powerful optimization technique, it is not always the best solution. Even though we know that nothing in life is 100% guaranteed, we often overlook this fact of life in the way we use predictive analytics for our daily business use. In reality the projects that drive measurable business results encompass more than just good models.   They typically incorporate an effective process such as Aryng's BADIR™: 5 steps from "data to decisions"™ framework. Key takeaways from this paper are:

1. Process (BADIR) is key to good business result
2. Right Talent + Good Tools = Great Models

Aryng

3. Models needs to be maintained
4. Models are not perfect
5. Use simpler techniques till PA can be justified.

About the Author:

Piyanka Jain, President & CEO, Aryng

Piyanka, founder of Aryng - a premier analytics training and consulting company, is a well-regarded industry thought leader in analytics, keynoting at business and analytics conferences including Predictive Analytics World, Data Science Summit, TDWI Big Data Conference, Google Analytics User Conference, Business Performance conference on data driven decision making in an organization.

With her 15 years of experience in analytics, she has had 150M+ demonstrated impact on business. Her prior roles include the head of NA Business Analytics at PayPal and senior marketing analytics position with Adobe.

About Aryng

Aryng is a premier analytics consulting company singularly focused on **in-sourcing of Analytics**.

We believe -

- Data has power to transform our day-to-day product, marketing and operations decisions.
- The people, who are most well placed to extract insights from the data, are those who are working within the organization in the respective product, marketing, sales and operations role.
- 80% of business problems can be solved using simpler techniques, which can be learnt by business professionals with no statistical background and can be performed in Microsoft Excel.
- Timely and relevant insight from data holds the key to drive up revenue and growth, and reduce cost and loss.

Aryng is about building Organization's capability – people, process, and tool, so the organizations can leverage data for smarter decision making. We do that through systematic **analytics maturity assessment** and then **addressing the gaps** through-

1. **Analytics training for business professionals**,
2. Setting up **decision making processes**,
3. **Executing analytics projects** while **teaching and mentoring**,
4. Executive coaching and
5. **Enabling right data tools**.

Contact Aryng

info@aryng.com |1.866.604.3092
3333 Bowers Avenue, Suite 130, Santa Clara, CA 95054

Aryng