**Best Practices for Predicative Analysis:**

Cathy Kiriakos

Regis University

MSDS 664\_X70 Predicative Analysis

March 16, 2020

The practice of predicative analytics has been around for ages, while taking the Ethics, Privacy and Social Justice course in the Fall of 2019 we were introduced to a book called “How to Lie with Statistics,” which was published in 1954; written by Daniel Huff. At the time of its publication, the mathematical and statistical methods to gain insight into trending had been well established. In present state we currently have enough computing power, along with programs and software to gain insight from real time analytics and machine learning approaches; allowing for organizations to gain a competitive edge against their competitors. We will delve into current state of analytics and the technology built around end state goals, the various modeling approaches that be utilized to accomplish them, the skills required to meet these goals, and best practices while doing so.

As mentioned in previously, the mathematics and statistics required for predicative analytics have been around for quite some time as the advent of such computing (on paper at the time) came into fruition during the 1940’s (Foote 2014). As of recently we’re come into an era of technology and development donning more processing power, speed, and distributed cloud processing; it has allowed organizations to scale build out analytics and forecasting will a low barrier to entry. This has allowed organizations to compete more effectively, having a strong grip on trending and insight based on indicators of what might be to come, allowing for a proactive approach to business operations versus reactive to events that have already occurred. More organizations are leveraging predicative analytics in their everyday business use, and those organizations that are behind on this trend are building teams and skillsets to adapt the capabilities into their operations which is where my workplace falls into.

There are many approaches to predicative analysis, in 2014 at the time Predicative Analysis for Business Advantages was published linear regression and decision trees were highly utilized (Harper 2014). However, the methods adopted vary far and wide based on the type of data and analysis, and best use case for an organizations particular need. In my studies and business experience, I have seen utilization of linear regression, and logistic regression in scenarios where you have data that is binary, time series models are highly leveraged in the world of finance. Unsupervised machine learning algorithms are highly leveraged in the targeted marketing used by many online retailers, and service industries. Medical institutions are also heavily leveraging data; applying clustering to imaging; specifically, being able to identify patterns of cancer in the body. Its obvious that most industries, corporations, and institutions are using machine learning and statistics and data to further enhance their abilities to reach targeted results to be competitive in todays economy.

Currently there is a wide array of technologies, software and modeling approaches available to businesses and organizations. However, a knowledge gap between understanding the capabilities of predicative analysis and having the ability to successfully implement scaled modeling is very real barrier to entry for some organizations (Harper 2014). Having adequately trained staff with both technical skills to run programs and the logical ability to trouble shoot issues and find gaps in models remains to be a challenge. According to a survey of 330 respondents, the most desired skills to apply successful predicative analytics into an organizations tool kit start at in order of importance; business knowledge, strong critical thinking skills, knowledge of source data, knowledge of predicative analysis, strong communication skills, quantitative degree, and the required software skills (Harper 2014). From this ordered list of importance, we can clearly see that organizations wishing to implement more quantitative analytical approaches to their business processes, weigh heavily on business knowledge and a general analytical ability. Which highlights the ability for organizations to build organic growth of their analytics, by leveraging current talent and enhancing their predictive analytics prowess via targeted training based on specific need. In addition to talent, organizations will also be challenged with building out the appropriate infrastructure to support enhanced analytics and predicative modeling techniques. This involves taking a deep dive into current and future needs of database functionality; having a strong understanding about the need to scale and how that will be approached whether it involves building out on premise architecture or going to the cloud. Decisions on the software and tools deployed to enhance analytics is also a topic that will need to be thoroughly investigated, organizations will need to determine the types of tools that will be leveraged for an example the choice to leverage Python or R or both, and the benefits and downfalls of those choice will need to be examined and understood and then scaled out.

Through this deep dive into the current state of predicative analysis and it business application we have gained a better understanding of the why’s, helping us understand the importance of adopting these methodologies in various types of settings from consumer marketing, to finance and medical advances. We have also gained a betted understanding of the types of tools and skills that are required to scale out predicative analytical modeling, along with infrastructure needs and the choices industry faces when determining the best approach to scaling out technology to remain competitive in our current business environment.

Citations

Harper, Fern. Predicative Analytics for Business Advantages. Q1 2014. <https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/general_format.html>

Foote, Keith. A Brief History of Analytics. September 25, 2014.<https://www.dataversity.net/brief-history-analytics/>