Governing Energy

Big Data Revisited

There is a common, but old story about the male who buys a six pack of beer and package of diapers at the local convenience store—some organizations lost millions because of belief in this correlation.[i] Data analytics at the time believed that patterns such as this one would predict future buying habits. Many were wrong!

Almost two decades later, this consumer routinely gets information, "based on what you recently did/bought/viewed, we recommend …" Marketers assess recent behavior and make a Big Data driven "pick" for me. Most turn out to be bogus at best.

Big Data Analytics have attained a *State of the Unquestionable*; that condition whereby the "science/law is settled" and we must all march to the tune of the Data Scientists. Such a mindset may cost shareholders billions.

In 2014, we commented in this blog about the need for the computational mathematics of computing to be correct in the equation being calculated.[ii] Issues surrounding the accuracy of algorithms have been known for decades.[iii]

Additionally, how the problem is set up is important too. For example, many polls in the recent election were inaccurate and we raised the issue of the possible poor construction and perhaps bias.[iv]

Another example, can be found in statistics. Most do not realize that there are three different types of variables; categorical, ordinal and interval. Certain calculations depend on knowledge of the variable type. For example, when using statistical where the variable is defined as interval, the assumption is made that the data set is equally spaced.[v]

Similarly, how data is distributed is important. Many assume data almost always has a normal distribution (fits the bell curve). Statisticians and other mathematician understand that is not always the case but can be approximated by using the appropriate sample size.[vi]

Another axiom, "Let your data speak for themselves" is not as straightforward as it sounds with the pundits either.[vii] For example, there are still issue with pattern recognition and whether a view of an animal is just that or a cloud formation.[viii]

The author has previously made the case that Mobility enabled Apps and the Cloud are just beginning a long run as base information technologies. We expect that the industrial sector will see rapid development and adoption of Operations Management Systems (OMS). Disruptive Big Data and the Industrial Internet of Things (IIoT).[ix]

Organizations can attain significant economic value from better decisions using these tools. According to one source, the value derived from operational efficiency from the IIoT can be "as high as $15 trillion of global GDP by 2030."[x]

However, when the numbers are this high, making the beer—diapers correlation mistakes may actually destroy shareholder value and possibly reshape sector winners and losers. Data Scientists can help assure catastrophic mistakes are not made; however, as always IT must be *aligned* with the business. If it is not aligned, no amount of Big Data Analytics can add value.

With great power comes great responsibility. Make sure your Big Data Analysis truly adds value to your organization. The phrase, "Lies, damn lies and statistics" has some basis in fact.[xi]

Are You Sure Your Organization's Big Data Strategy is Appropriate?

About the Author

Dr. Scott M. Shemwell has over 30 years technical and executive management experience primarily in the energy sector. He is the author of six books and has written extensively about the field of operations. Shemwell is the Managing Director of The Rapid Response Institute, a firm that focuses on providing its customers with solutions enabling Operational Excellence and regulatory compliance management. He has studied cultural interactions for more than 30 years—his dissertation; *Cross Cultural Negotiations Between Japanese and American Businessmen: A Systems Analysis (Exploratory Study)* is an early peer reviewed manuscript addressing the systemic structure of social relationships.

Free Economic Value Proposition Matrix version 2.0 (Realize the value of your investment)

End Notes

[i] http://www.forbes.com/forbes/1998/0406/6107128a.html
[ii] Shemwell, Scott M. (2014, September 22). Man Machine Codependency. Governing Energy. PennEnergy.
[iii] https://docs.oracle.com/cd/E19957-01/806-3568/ncg_goldberg.html
[iv] Shemwell, Scott M. (2016, November 17). All the Pundits Were Wrong—Again! Governing Energy. PennEnergy.
[v] http://www.ats.ucla.edu/stat/mult_pkg/whatstat/nominal_ordinal_interval.htm
[vi] Ibid.
[vii] http://www.burness.com/let-the-data-speak-for-themselves-is-bad-advice/
[viii] http://www.theepochtimes.com/n3/1401624-googles-dreaming-ai-shows-what-real-machine-learning-looks-like/
[ix] Shemwell, Scott M. (2016, May 18) Implementing a Systemic Culture of Safety: The Role of IT. PNEC 20th International Conference on Petroleum Data Integration, Information and Data Management. Houston.

[x] _____ (2016, April). Achieving Operational Excellence: Capitalizing on the Cloud. Petroleum Africa Magazine. pp. 32-34.

[xi] https://en.wikipedia.org/wiki/Lies,_damned_lies,_and_statistics