MetricVision

Becoming Data-Driven: a Culture of Experimentation

Steps Businesses Need to Take in Order to Become Data Driven

Dr. Al Nigl

According to some, 90% of all the data available for analysis today has emerged in the past two decades

This explosive expansion of data to become what we now call “Big Data” has occurred due to many factors, including the maturation and domination of the Internet as the primary source of information transmission, the rapid development of social media, the transformation of purely textual information to multi-media formats, especially video and the digitization of everything.

The explosive growth of IOT (Internet of Things) will create more data to process in the coming decade.

The most successful companies today are or rapidly becoming Data Driven; examples include Google, Amazon, Capital One and Facebook.

And each of these companies devotes a significant amount of data science time and energy to running data experiments to support effective predictive model development.

Data Science is an emerging discipline that has been described as having four main components: Data Mining, Machine Learning, Big Data and Predictive Modeling

Some thought leaders have questioned this definition however, claiming that any scientific discipline must have Experimentation as a core component

And you can not have experimentation without statistics, otherwise how would one determine whether an experiment was successful or unsuccessful

The absence of rigorous experimentation and statistical tests to measure significance or even effect size characterizes most of what is termed “data science” today

The question then must be asked: “Can Data Science truly be a science without experimentation?”

Many knowledge experts have suggested that Data Experimentation is an indication of Data Science Maturity and many also believe experimentation is also the cornerstone of a Data Driven company

Data science is often first utilized by companies to understand the relationship between customer behavior and product adoption and purchases

As a company becomes more data driven, it uses its data science capabilities to answer questions such as “What if” not just “who, what, and where”

“What if” questions are tantamount to the application of the scientific method and data experimentation

What if questions can lead to explosive growth in data insight and increased proficiency in predictive modeling built on these new insights

Based on an analysis of the thought leadership in Data Science today, it appears that there are three main pillars that comprise a truly data-driven company:

I. Data Understanding and Integration (using CRISP_DM)

II. Employee Education

III. Data Experimentation (including Data Mining & Predictive Modeling)

Data Understanding and Integration

a. There is a company-wide understanding of what data is being collected, what different databases are being maintained and for what purpose and what type of data is stored within each database

b. There is an integrated approach to working with the data (if in fact the different databases are not unified or integrated----which would be optimal, of course)

c. A unifying process for data analysis and experimentation is in place e.g., the very influential and pervasive CRISP_DM process

Employee Education

There is a company-wide directive for supporting continuing education in data science, statistics, data mining, data experimentation and predictive analytics for all non-technical managers and some junior data analysts without a formal degree in computer science, statistics, data science or a formal scientific discipline

In a truly data-driven company, innovative ideas often flow upward and non-technical staff members (who have had course work in data science) are encouraged to submit ideas for data experiments in formal and non-formal staff meetings

Data Experimentation

Data driven companies not only have data mining processes in place which support predictive modeling but they also devote considerable technical staff resources (time and intellectual capital) to conducting rigorous data experiments

In fact, some authors and commentators have put forth the position that companies would be more successful and more innovative if they did more data experiments and fewer predictive models

Many believe that too much intellectual capital is being wasted on modestly effective or even ineffective models and not enough attention is given to randomized experimentation which would lead to greater data insight and more effective prediction

Data Experimentation

Non-data driven companies (which describes most companies today) often rely on off-the-shelf, proprietary predictive modeling packages which too often apply the same set of modeling algorithms to the data without much thought of fine tuning and matching the algorithm to the data

This automatic approach may yield modest success in marketing campaigns but is certainly not an approach which optimizes the value of the company’s data which can be realized through the combination of data experimentation and modeling

Some consulting companies now provide a virtual analytical laboratory for companies as a service so they add Data Experimentation as part of their Data Science capabilities

Consulting data scientists perform dozens of scientific experiments on a company's data to uncover connections which were previously unknown

The results of these experiments can then be used to create highly predictive models which outperform previously derived models with no experimental basis or science foundation

Data experiments can include Monte Carlo simulations, ensemble models, data transformations to permit the application of up to 28 different algorithms to find the most effective model that fits the data as well as rigorous model validation experiments including the statistical analysis of Decile Tables, Lift Charts, Target Reshuffling and Bootstrapping.