

THE DATA DEBATE

Remember the old Star Trek series, the character “DATA” was presented as a rather bland and unexciting person. With all the current hype and discussion surrounding “Big Data”, it would be interesting to see how the “DATA” character would be presented today. Nevertheless, DATA and the discussion of data represent the new cool. There isn’t one day now when at least one major mainstream publication does not mention “Big Data” or data.

But with all this discussion on data, what are the key burgeoning issues for business. If one looks at data, as the raw material for decision-making, then one begins to obtain more clarity around this challenge.

Companies have no problem amassing or gathering data which has exponentially increased due to the ever-increasing digital environment. But how to make sense of all the data? The data is often in very distinct solos which are often unconnected to each other. Furthermore, the “raw material” or source information of this data is meaningless unless it is converted into information that is more meaningful. Good examples of this are the unstructured information pertaining to emails, posts, blogs which represent textual type information. Through text mining type tools, this unstructured textual information can be converted to meaningful information. The interest in extracting meaningful information from text is not new. It is the capability of being able to do this which is new today and that has resulted in an entire new discipline referred to as text mining.

Even within the structured data environment, examples abound of the need to transform raw data into meaningful data. For example, postal code information by itself is useless for data mining purposes. However, the ability to group postal codes into categories such as regions suddenly transforms this information into a more meaningful variable for analysis. Yet, the most classic example of converting raw structured data into meaningful information represents the use of “transaction” data. Here the data miner or data scientist uses their skills to create and derive all kinds of different “transaction” behavior variables. Raw transaction data can be summarized in an infinite number of permutations from basically three fields:

- Amount
- Date
- Type

Besides the infinite number of summarization options that exists, another key element or behavior that the data scientist /data miner is trying to capture is the notion of change. Using the data field, the analyst can create how behavior is changing overtime as well as looking at this change behavior by type of transaction.

Another often overlooked skill is the ability to link data. As Big Data continues to monopolize discussions, it is easy to forget that in most situations data needs to be linked or integrated. For example, information pertaining to online behavior, purchase/transaction information, campaign history, demographic behavior all needs to be linked together in order to gain a more holistic view of that individual or customer. Here, technical skills are required in not only determining the relationship between files and tables but also what is required to link the tables. In many cases, the fields to link these tables need to be constructed by the analyst. Once again, a certain level of technical expertise is required in order to build the appropriate link field.

Certainly, any discussion of transforming raw data into more meaningful variables does not nearly communicate the complexity or hard work which is required to successfully complete this process. Yet, what are the key ingredients to achieve success? Going back to our 4 step data mining process, the first step in being successful is identifying the business problem. Once the business problem or challenge is identified, then the analyst needs to work their “magic” with the data. The essence of this magic derives from two areas:

- Domain knowledge of the business and what is required for business to be successful
- Mapping of current data/information environment into meaningful derived variables based on above problem/challenge

So far this discussion has tended to focus on what is important from the data scientist perspective. But the other equally important perspective arises from the value architect or key business stakeholder. The points discussed above within this data debate do not change but what really changes is the relative strengths of both the data scientist and value architect concerning each of these points. For example, the data scientist will clearly have a much stronger focus on the information environment while the value architect will focus on the business domain side focus. Yet, overlap will and should arise as the data scientist will have knowledge and insight on the business domain area while the value architect will most certainly bring a perspective concerning the data. In fact, one might argue that the most successful solutions arise where this overlap in

areas is more pronounced but at the same recognizing the specific core functional strengths of each area.

Needless to say, data is the new cool. Its growing recognition as a key corporate asset will just continue to expand the significance of roles such as data scientists and value architects within organizations. It's a great time to be a "data" nerd as it is these roles for people beginning their careers that may actually lead to a career path that ends up in the "C" suite of organizations.