

## Data and Analytics Skills for Your Career Security

*Keeping it simple. . .  
only the skills you're likely to use*



**Richard G. Lamb**

There is tremendous hyperbole around data and analytics. To get beyond the confusion it causes, it is necessary to narrow the hyperbole to the meaningful few definitions. Of all of the hyperbolic terminology, the essential can be narrowed to four: data and big data, machine learning, artificial intelligence and algorithms.

Excerpt:

**2.2.3. Essential Definitions**



Additional "Look Inside" at <https://analytics4strategy.com/book-look-inside>

Book is available from Amazon

Paperback, 508 pages, 300 figures and outputs, 7.5 x 1.15 x 9.25 inches, 2.37 pounds

### **2.2.3. Essential Definitions**

There is tremendous hyperbole around data and analytics. Ironically, the hyperbole may be a cause for managements' instinctive aversion to the discussion of data-driven operations. It has also caused tremendous confusion as an obstacle to become data-driven. To get beyond the aversion and confusion it is now necessary to narrow the hyperbole to the meaningful few definitions. With respect to them, the terminology of the hyperbole are expressions of the same thing but with different and flashy wording.

The essential definitions can be narrowed to four. They are data and big data, machine learning, artificial intelligence and algorithms. Many terms have come and gone over the last several years, but the four will remain as the language of data-driven asset management.

**Data and Big Data.** Data and big data are distinctively different with respect to necessity, technology and organizational abilities. It is an important distinction. This is because big data entails high-tech systems and infrastructure, specialized skills and capital. Data does not.

We tend to think of “big data” in a colloquial sense from working with Excel. In the context of Excel, thousands or hundreds of thousands of rows are “BIG.” It is difficult to work with that much data in an Excel worksheet. Things are laborious once we get beyond a few hundred rows.

However, lots of data does not make it big data. Big data is the case in which data are prohibitively massive or unstructured. A professor of data science proposed a simple litmus test of data versus big data. It is big data if the data or the analytics of the data cannot be worked on our notebook computers.

Unstructured data include disparate types of data such as streaming sensor data, e-mail, document, video, photo, audio and webpage. This is compared to the structured alpha and numeric data that is predominate to operating systems. The purpose of analytics of unstructured data is to transform them to be structured data with which other analytics can be conducted.

The types of data analytics conducted in either arena are the same. The type does not decide whether it is data or big data. However, the important notable reality for data-driven asset management is that there are very few imaginable needs for big data.

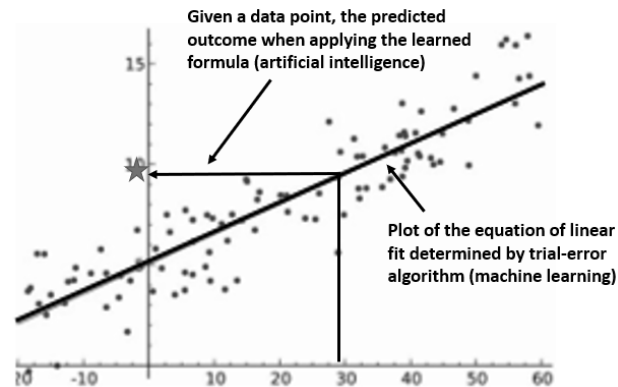
Let’s look at a situation that is not big data, but the definition of unstructured data may cause us to assume it is. Other than the massive, streaming data of CBM systems, the only unstructured data in most operating systems are the free-text variables of work order descriptions and notes. Text mining analytics may be used to classify the work orders by the failure mode that triggered them. Remembering the professor’s litmus test, although free text is unstructured, it is possible to conduct text mining on our notebook computers.

The IIoT-supported CBM of Figure 2-3 is an example of big data. Streaming data is massive because the data points are almost infinite in number. Additionally, for some types of monitoring, the data is not structured data and must be transformed by analytics before it can be subjected to the ultimately planned analytics.

Therefore, when the expression “big data” is casually tossed about, we must ask an inspectorial question. Is it hyperbole or big data? If it is truly big data, then we cannot take the grassroot strategy as it is otherwise possible with the triad of software.

And once again, a very important point must be repeated. Almost never will the insight deliverables to asset management entail big data. This is a fundamental reason that data-driven asset management can be grown from the grassroots.

**Machine learning, artificial intelligence and algorithms.** We need to clarify the interrelated terminology of machine learning, artificial intelligence and algorithms. We will use the two-variable regression analysis shown Figure 2-4 as a frame of reference.



**Figure 2-4: A two-variable regression demonstrates machine learning, artificial intelligence and algorithms.**

We have all at some time touched regression modeling. However, the concept is the same regardless of the type of model and the number of predictor variables placed in the model.

Machine learning (ML) takes place when we feed the predictor and outcome variables to the regression. The gut algorithm conducts a trial-and-error calculation until “learning” the best fit and returns a formula of an intercept and slope coefficient. The predictive and outcome variables are the axes and the points are their cross-plots. The line fit to the plotted points is the learned outcome of the algorithm.

Most often our interest ends with the returned coefficients for each variable (one in this case) and associated inferences for how strongly, if at all, the predictor variable is related to the outcome variable. Confidence interval to the coefficient will also get our attention. In contrast, artificial intelligence (AI) feeds new cases to the fitted model to predict outcomes upon the “learned” formula.

AI does not distinguish the model. All types of models entail machine learning, and most can be deployed as AI.

When the model is to be deployed as AI, the learning process for some types entail an additional stage of analytics. A portion of the original dataset is held out from the machine learning stage to be a test set. The remaining portion is fed to the model for learning. The test set is subsequently fed to the learned model to evaluate how accurately the “trained” model estimates or classifies the actual outcome of each case in the test set.

## 22 | Chapter 2

If accuracy is acceptable to the intended use, the model is deployed to serve its purpose—augment human experience and judgement. Results better than 85 percent are typically considered acceptable. This is why we must always think of AI in terms of “augmenting” rather than “supplanting” experience and judgement. Hyperbole may lead us to unrealistically expect “supplant.”

## Data and Analytics Skills for Your Career Security

*Keeping it simple, only the skills you're likely to need*

**Richard G. Lamb**

For those of us who are role holders in enterprise functioning, the personal purpose of acquiring practical working skills in data and analytics is to be able to better do what we already do and find new ways to do better yet. It follows that if you are a role holder who brings and incorporates data and analytics methods in your thoughts and tasks, your career outlook will be more secure and exciting. The book is written to be your gateway to the skills and to be the templates with which you will install the methods in your operational roles.

We all know that the field of data and analytics is huge and intimidating. It is a long slog to becoming comfortable. During the author's own long slog until arriving at the book, something exciting bubbled to the surface. There is a big difference between what we need to know and everything there is to know. We need to know what is possible as insight for decisions and functioning, we need to know how to get to the insight and, finally, we need to be able to interpret the insight. Just as the book does, we can leave the rest to the data scientists.

**About the Author:** In 2003, Richard Lamb, while struggling to get at the history captured in the databases of operational systems, found the skills to extract datasets of related history and join them in a super table of variables to make possible what was being envisioned for operational effectiveness. In 2014, Richard realized that, with statistical analytics and free enabling powerful pc-level software, an enterprise could ask and answer questions of operational effectiveness that are otherwise not possible. His activism to bring the epiphanies into the careers of role holders in the mainstream of operations has arrived at this book to explain data and analytics through the demonstration of methods.

Richard is a Registered Professional Engineer and Certified Public Accountant. He has previously authored two books: Availability Engineering and Management for Manufacturing Plant Performance, and Maintenance Reinvented for Business Performance. He has a BSCE, BBA and MBA from the University of Houston and a graduate-level Applied Statistics Certificate from the Texas A&M University.

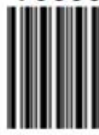
<https://analytics4strategy.com/data-and-code>

**Analytics4Strategy**

ISBN 978-1-7343947-0-2



90000 >



9 781734 394702