Building a Basic Model and Telling a Story

Marc S. Galli

Walden University

Professor Dr. Mike Collins

DSCI 2010: Data Science Essentials

August 15, 2020

Building a Basic Model and Telling a Story

The ultimate goal of data science is to create the methods for extrapolating insights from data for a business purpose. Within the data science methodology, all of the necessary steps are incorporated which will take a rudimentary business understanding and through articulated analysis of data requirements, collection methods, thorough understanding of the data set, and finally preparation of the data, the team can model, evaluate, deploy, and gather feedback on the business insights obtained. Data science methodology is an iterative process and needs multiple revisits to refine the insights to suit the needs of the team. As seen within the Nutri Mondo case study, the data science team began to evaluate whether their model met their goal, and they found sometimes it is necessary to have an objective third party review the models to see if insights are clearly obtained and elucidated through the visualizations (Laureate, 2018). In review of data science methodology, I will focus on the modeling and evaluation aspects in this report. In the data modeling step, models are developed based upon the datasets acquired which will either tell what has-happened or what might-happen (IBM, 2016). To this end, and through these means, data modeling can either be descriptive or predictive. Ultimately the models are reliant on having taken an analytic approach from inception. This is paramount.

As concerns data modeling, first and foremost, the keen data scientist must recognize that models, or visualizations if you will, are principally designed to answer a question and present that understanding in an immediately cognizable manner: visually. With an initial business understanding, the right dataset, and the correct visualization, the model will describe or predict and depict what has- or might- happen, respectively. It should, then, not come as a surprise that data modeling must come after data scientists have understood and prepared their data. The models themselves are based on an analytic approach where data scientists seek to comprehend a deeper grasp than a cursory look could ever reveal about the data. Models can be statistics-driven or machine-learning driven. If the data scientist chooses to formulate a predictive model, the data scientist must use what is termed, in the industry, a “training set” (IBM, 2016). This training set is comprised of historical data where variables are controlled and the outcomes of the visualization

and model are already known. When the training set is applied to the visualization, it acts as an instrument to conclude if the model must first be adjusted or calibrated. Said calibration is a means to control for certain constants so variables can be accurately gauged. It is in this step of the iterative process, modeling, wherein the data scientist will make several adjustments to any computational algorithms which may be involved to ensure they're using the most simplified versions of said algorithms. It is with this understanding of the modeling process that the data scientist concludes that successful data compilation and modeling depend on having first understood and prepared the data.

The data scientist must, at all times, keep the question for which the model seeks to answer in the forefront of his or her consideration. After all, as previously stated, the model (visualization) is principally designed to answer the question by presenting the answer visually to bring forth immediate cognition. A great example of this is seen in how bar graphs and line graphs immediately convey change over time where review of the dataset would require time multiplied by the vastness of that dataset. In the data science methodology, the steps of modeling, evaluation, deployment, and feedback are iterative and here in lies the method by which data scientists ensure that modeling responds to the original business understanding concluded from step one: As feedback is received, the team returns to modeling as needed (International Business Machines, 2016).

I have already explained a revelatory purpose for data modeling. That is, the ultimate visualization provides a speedy way to take in a voluminous dataset and achieve an instant opinion about its conclusions. However, prior to arriving at the ultimate model and visualization of the dataset, the modeling step serves as a method for the data scientist to assess the appropriate application of the training set, to see if all constants are appropriately controlled and if the min and max values for data described by the axis labels are within view of the data's graphical representation. This is called parameter tuning and is the additional aspect of model building (Big Data University, 2016). Iterative evaluation leads to the perfect model.

References:

Big Data University. (2016). *Modeling Case Study* [Video file]. Armonk, NY: Author.

IBM. (2016). *Modeling – Concepts* [Video file]. Armonk, NY: Author.

International Business Machines. (2016). *Evaluation* [Video file]. Armonk, NY: Author.

Laureate. (2018). *Nutri Mondo*. Laureate International Universities. https://mym.cdn.laureate-media.com/2dett4d/Walden/DSCI/2010/WorldOfDataScience/index.html