Deploying a Basic Model and Refining Methods

Marc S. Galli

Walden University

Professor Dr. Mike Collins

DSCI 2010: Data Science Essentials

August 22, 2020

Deploying a Basic Model and Refining Methods

In review, a data scientist's primary function is to organize large amounts of data, be it structured, unstructured, or semi-structured, and then to analyze it to discover insights which can be interpolated from the data (White, 2018). The keen data scientist may even uncover insights, *prima facie,* utilizing conventional methods such as reviewing charts and graphs which he or she creates from the data and through perceiving patterns and trends readily comprehended visually. The step-by-step procedural process which begins with a comprehensive business understanding and ends with received feedback on a deployed model, is known colloquially as the data science methodology. As reviewed, an initial and comprehensive business understanding is crucial, such that, the data scientist can collect and analyze data, then model insights, which are actually relevant to the business' needs. The process of refining data requirements, collection, understanding, and preparation is iterative and promotes the relevance of obtainable insights. Once modeling is underway, the subsequent process of evaluating said models, deploying them, and receiving feedback is also an iterative process as the data scientist strives to refine the models and visualizations.

In the ongoing analysis of our principle case study, Nutri Mondo has successfully modeled the data science team's data. The data science team worked hard to evaluate the models. They sought to clarify and structure governmentally sourced data and removed outliers from the scope of the graphic visualizations to ensure constituent data was appropriately represented and visuals depicted trends readily. Each horizontal and vertical axis was carefully chosen and labeled to give the best exemplification. Subsequently our Nutri Mondo team began to outline plans for deployment. The team determined that the visualizations would be data-driven and descriptive. Predictive models were not employed.

Data science model deployment typically takes on two primary forms depending on the purpose for modeling which has been contemplated. Either the models may be deployed to an initial and limited group of staff or deployed in a test environment where confidence can first be built and an assessment can be made as to the broader application of said models. Several choices were contemplated. First, deployment was considered within the context of a very limited initial release which would only be to directors at the Nutri Mondo Miami headquarters. Secondly, the

team contended that deployment could be made as to a limited release and sampling of models to the directors at offices world-wide. Thirdly, deployment to directors in select regional offices in the U.S. could be made with select deployment in other countries where Nutri Mondo has offices. Lastly, deployment could made in an 'open release' to the entire organization at once (Laureate Education, 2018).

After careful consideration of each choice laid before the Nutri Mondo team, I have analyzed the alternatives and decided how best I would advise Nutri Mondo director Susana Maciel if the decision were mine to make. I have weighed-in the additional contemplation that we are a world-wide organization and am most favorably inclined to agree with Karen Pond. Karen Pond has been stationed at the São Paulo, Brazil location and comes from an objective viewpoint on account of the fact that the data science team used primarily United States governmentally-sourced data. She is experienced in data analysis and prior to Nutri Mondo worked for a legal firm investigating bank fraud (Laureate, 2018). She was behind development of custom financial audit systems. With credentials and a degree in computer science and her over four and a half years of experience working for Nutri Mondo, her insights and expertise can be relied on.

Therefore, the third deployment option up for consideration, developed by Karen Pond is the method I embrace and endorse. It will be the most beneficial for us to deploy to directors selected from regional offices in the United States and in some other countries. Karen correctly notes that we would need director approval before this release but releasing to directors globally falls short of addressing adequate considerations for timeliness. Alternatively, if we had released only to Miami, Florida, USA directors, it would take considerably longer for our data to be usable by the very Nutri Mondo staff we developed it for. Our data may be significantly less impactful, or worse, outdated and irrelevant. By cherry-picking specific Nutri Mondo directors to receive our releases, our team will be working with directors who will be making adequate use of our data models and we will get the benefit of direct feedback and general oversight. Receiving this feedback will enable us to enhance our models incrementally and will aid us in resolving errors before the landmark release to our entire organization. Lastly, I assert that limited deployment would bring in too much feedback. The Nutri Mondo team assigned to this project consists of only four team members. The feedback would be too overwhelming and since it would be coming from

directors who couldn't fully utilize our regionally tailored data, it would be irrelevant and would not be helpful.

In spite of a global overabundance of food, millions are still underfed (Laureate, 2016). The Nutri Mondo organization needs to address causes of hunger. After our models were deployed, feedback began to pour in. The Nutri Mondo data science team need not be caught up on studies linking Supplemental Nutrition Assistance Program enrollment to obesity. The directors are seeking assistance in order to work individually and independently with our team's visualizations and models (Laureate International Universities, 2016). They desire to create custom charts. They are willing to use commercially and commonly available software such as Microsoft Excel but request access to our organized datasets. They want to graph using our dataset and tweak the models for their individual regions. There is a common desire in received feedback for directors' ability to conduct independent analysis. They wanted help with presenting the data (Laureate International Universities, 2016). In fact, the principle complaint I have noticed in the emailed feedback referenced in the class learning resource was that the feedback from each director echoed the request for more local region-specific detail in the datasets and a common complaint among foreign national offices that the United States data would not apply to their regions. Hints abounded that we may be missing out by settling on just governmentally sourced data.

To conclude and with final analysis, the data science methodology, as intended, will be best when used iteratively. The resulting feedback should be used to return to the modeling step and repeat evaluation and deployment phases to further refine as needed. A data scientist's goal is to organize large amounts of data and to analyze it to discover insights which can be interpolated from the data. Nutri Mondo intends to use data scientists and modeling to aid directors world-wide in working to solve global food insecurity.

References:

Laureate. (2018). *Data Scientists*. Baltimore, MD: Author. https://mym.cdn.laureate-media.com/2dett4d/IBM/LIDS0000/EN/nutrimondo/index.html

Laureate Education (Producer). (2018). *Case Study: Deployment* [Video file]. Baltimore, MD: Author. https://mym.cdn.laureate-media.com/2dett4d/IBM/LIDS0000/EN/Deployment/index.html

Laureate International Universities. (2016). *Nutri Mondo Email Feedback.* https://mym.cdn.laureate-media.com/2dett4d/IBM/LIDS0000/EN/emails_feedback/emails_feedback.pdf

White, S. K. (2018, May 24). *What is a data scientist? A key data analytics role and a lucrative career*. CIO. https://www.cio.com/article/3217026/what-is-a-data-scientist-a-key-data-analytics-role-and-a-lucrative-career.html