Ryan Johnson

Design Thinking Statement

Data Science is a growing, complex, and often misunderstood field. Many people think of data science as a mathematical/scientific process to mine data. However, data science is not just about mathematically approaching a data problem. The main goal of data science is to solve a business problem and/or extract relevant insights to inform later business decisions. This not only involves mathematical algorithms, but also domain or business knowledge. This brings us to the topic of design thinking, which can be defined as creatively experimenting to solve a business problem.

Design thinking involves seeing a business problem as a unique situation. Not every problem needs to be solved with a complex, multi-layered algorithm. Often, the importance of domain experts is overlooked. The incorporation of domain experts to simplify a business problem, accompanied with a proper algorithm, can lead to a much more relevant and successful data science project.

Having a design thinking approach will also help when working through end-to-end solutions. This type of solution sees a data scientist help develop a business problem, develop a model, test this model, deploy this model, and continually evaluate this model for necessary changes. For example, COVID-19 caused rapid changes in many industries, such as the toilet paper industry. At the beginning of the pandemic, toilet paper was a hot commodity, and stores struggled to keep the shelves stocked. Price optimization models would not be trained for such wild occurrences, so this model would need to be re-trained in order to produce accurate results.

Another important aspect of proper model design is that it is not overfit to a specific dataset, meaning it is not programmed to only be able to address the current dataset. Having your model become a REST API, where it acts as an endpoint of sorts, provides validity to your model. Within the adoption of the model as a REST API, other data scientists will be able to send data to your model, and your model will send back the proper results.

When designing a model, it is unrealistic to expect a perfect model on the first iteration. Many iterations of a model will be needed before it can be trusted, which is the process known as rapid prototyping. Within this process, feedback loops are crucial, as this allows the data scientist to obtain relevant feedback from the user on the validity of the model. One example of a feedback loop is Facebook, who use a thumbs up or thumbs down feature that makes it easy for the data scientist to determine which elements of the model are working, and which aren’t. Sometimes there are relevant variables missing within the model that should be used to determine specific outcomes. For example, Dave Wentzel, a data scientist for Microsoft, has discussed a previous feedback loop with an insurance company that allowed him to realize that certain items were missing from the revenue projection model. Design thinking overall is about helping a client solve their problem, so it is necessary to gather feedback and create many prototypes to find the best solution for your client.

Feedback loops are also integral to helping with the ML Ops process. ML Ops, or Machine Learning Operations, is a process of developing and deploying a machine learning model for business use. ML Ops are very similar to the previous end-to-end solutions we discussed previously. It involves identifying a problem, gathering relevant data, building a model, testing the model, deploying the model, and monitoring the model. Integrating a feedback loop will allow the data scientist to better monitor this model. The most important part of ML Ops, which is also very similar to Dev Ops, is the application of the model for business use. Again, as discussed with design thinking, the math and statistics part of the model development are much less important than the application and use of the model to solve certain business issues. Overall, design thinking allows a data scientist to better address problems and assist a business in making better, more data-driven decisions in the future.