Taylor Moorman

Syracuse University

Master of Science in Applied Data Science

June 18, 2019

LINK TO GITHUB: <https://github.com/tmoorman95/SyracusePortfolioMilestone>

Portfolio Milestone

My time within the Applied Data Science program at Syracuse University has been an experience of incredible growth within my chosen area of expertise and trade, a copious amount of work and time management, and a learning experience which has taken my confidence and my knowledge of data, data architecture, machine learning, and coding/programming to heights that I was not sure that I could reach. I chose to do this program coming straight out of my undergraduate studies at Arizona State University for two reasons. One, I got a great big taste of it all within my internships and undergraduate coursework, but I knew that there was a much deeper level of knowledge and experience that I simply did not have upon the close of my time in undergrad, and that was a level of understanding that I sought to reach. Secondly, after much research, I knew that attaining this degree could help fast-forward my career and place me on a fast track to more responsibility, greater pay, and much more important and potentially more fulfilling work. Therefore, I decided to take this program and accelerate it as fast as possible. I started in October of 2018, and will be done by the end of September 2019. I have maintained a full-time job as a Business Intelligence Analyst while taking three classes a semester my entire time in the program. Something that a few counselors along the way tried to give me forewarning about. For all of this, I am incredibly proud. During my studies in the program, there have been a number of different projects and milestones that truly ingrained a lot of the learning and techniques that I have come to understand at this point in time. Certainly, there have been numerous opportunities for me to show what I have learned and apply the methods that I have been taught by the phenomenal teachers in tis program, but there were definitely a few that stuck out amongst all the rest.

**Learning Goals**

To meet the overarching learning goals of the Applied Data Science program, a number of projects were incredibly formative in my experience. I will discuss these below and the goals that they helped me accomplish. First, in IST 652, Scripting for Data Analysis, I undertook a project that attempted to predict the outcomes of mixed martial arts bouts. This project was written in Python and helped me accomplish showcasing a myriad of data science areas and challenged me to collect and organize data. The second project was in my IST 707 Data Analytics course where my group and I predicted the economic impacts of hosting the FIFA World Cup. This project showcased being able to identify insights from data visualization and statistical analysis, developing alternate strategies based on the data, and in communicating results. The next significant project was in my MAR 653 Marketing Analytics class where my group and I predicted sales for a bank based on their marketing outreach efforts and other demographical factors. This project showcased my ability to develop a plan to implement business decisions from the analyses. The last projects was from my IST 623 Introduction to Cyber Security course which helped me build the ability to synthesize the ethical dimensions of data science like privacy.

1. The first project of mention is my final project in IST 652, Scripting for Data Analysis.

For this work, I created a program in python that would read in a number of csv data files from online sites, and clean, transform, and join them together to create one master dataset, then this program takes that data set and trains a RandomForest model to output predictions of UFC fights and whether any given fighter given a number of features would win or lose any fight. During the course of this project, I needed to showcase my skills in every area of data science.

First, I had to go out and collect data, I needed to find data and format it into a method to which I could then run analysis on. Second, I had to then use data transformation and data cleansing techniques to get the data into an actionable form to where it could be useful for analysis and ultimately prediction. Next, machine learning and other data mining techniques such as correlation analysis, data visualization, and performance metrics were used to analyze the data and get to an answer of the questions posed in the project guidelines. After all of the actual programming, coding, and analyses were completed, I then had to explain my findings in presentation form and within a business context. This meant presenting very deeply technical work in a context to where those without deep data science knowledge could understand and interpret the work that I had completed. This project allowed me to touch on every aspect of the data science cycle, from initial project inception, into the nitty-gritty of the programming the solution, and then to the end of the cycle where the findings are presented to business users.

1. Sticking with the same project, my final project in IST 652, Scripting for Data

Analysis, this project was incredibly difficult in the organization of data. More so than any project that I have done during my time in this program, this project required by far the most work in transforming the data to get it to a working point before the analysis even started. My project necessitated the use of three different data sources. However, because I was predicting singular entities of fights with multiple entities inside of them (fighters) I had to combine the sets to make any logical sense out the predictions. I needed to be able to predict outcomes of one fight and one fighter per fight. In the program that I wrote for the function, now only is each individual set pared down to the necessary columns, but then they are cleaned and joined together. Lastly, some supplementary or calculated columns are added to the data set, and other columns are discretized or encoded into binaries for easier feeding into the RandomForest model. The ability to transform data in all of these different ways has certainly become a strength of mine during my time in the program. Before enrolling, I had no idea about the various types of ways to manipulate data to get it into a better form for modeling. I didn’t know what types of data are best with specific models, and I would have had no great idea how to best get predictions from data. Lastly, my imagination of calculated columns and conditions that can be put on various features when feeding data into a model would not be where it is now, had I not been in this program, and had to tackle projects like this one.

1. My IST 707 Data Mining final project is a great example of identifying patterns within

analyzing data. For this project, my team and I sought to use a Kaggle dataset of FIFA world cup matches and teams, along with a data set of financial winnings from past FIFA tournaments to attempt to predict if a given team (country) should want to host the world cup. Our criteria for making this decision was ultimately if they came out ahead after hosting. In other words, was hosting the tournament a good or bad financial decision for an individual country. During our analyses, we looked at the different attendance numbers for the games by country host for historical world cup matches. We found out that host countries are much more likely to not only place in the tournament, but also to win individual matches, and more likely to win the tournament overall. We then looked at various financial data for FIFA and the given World Cup prize amounts for every tournament that there was data for. We determined that the prize money went up every year, but also that the cost to host was tremendous. Ultimately, by combing our match and tournament analysis with the financial work that we did, it became clear that at best a host country could hope to do is break even when things are all said and done, but that most countries lose a substantial amount of money when hosting the tournament. This project certainly helped my ability to draw conclusions based upon mining and visualizations, however this is something that I had serious experience with in my working profession, but also in my undergraduate studies as well and something I would have already considered a strength.

1. In the same project for IST 707 Data Mining, my team and I had to make a critical

pivot in our goals as we got deep into our analysis. Our original plan had been to attempt to predict every country’s end financial gain or loss of hosting the World Cup in any given tournament. Unfortunately, we discovered that we could not find the prerequisite consumerism and tourism data for the time periods we needed for each country in order to make this goal a reality. Not only that, but the exact monetary value of the spend of each country in order to host the tournament was not clear in every case. This made us pivot into an alternative strategy where we would forgo attempting to outright predict whether each country would breakeven, but rather analyze the tournament as a whole in its history and get to the sum-all answer of whether it was a positive decision financial to host the tournament at all. In addition, we wanted to see what countries or what conditions would factor in most to the end result of hosting. In the end, we were able to ascertain that only if host won the tournament, did hosting the tournament approach being anywhere near a zero-sum activity, but that it was nearly impossible to come out ahead.

1. I am thankful that my time in this program has also come with a number of classes

from the Whitman School of Management. This is because, in the end, whether a data scientist has discovered something extraordinary or spent 6 months developing a new model, it all has to be for a business purpose, and it all needs to make business sense. Ultimately, the decisions we make in working with data needs to lead to an enhancement for the business. In this thought, my work in my MAR 653 Marketing Analytics class definitely brought this teaching to light. Every deliverable for that class needed to come in the form of recommendations for the end business. In my final project for this class, my team and I looked at telemarketing campaign data for a Portuguese bank. Through a variety of campaign factors, previous campaign results, socio-economic data, and demographic data about its clients, the bank sough to predict whether or not their customers would choose to invest in a new product or not. The dataset all boiled down to a binary yes or no target variable of investing or not investing.

Our team took on that task, and then also looked at which factors were the most important in the client’s ultimate decision. As an end deliverable, we needed to provide recommendations for this bank’s marketing efforts. Through the use of binary logit regression, visualization, and other statistical analysis techniques we came up with the end recommendations that quite simply, more money spent on campaigns equaled more client investments. Luckily the most important set of factors as to whether a client would invest was that of the campaign and previous campaign. This led us to recommend that they focus on increasing that spend and touching as many leads as possible. In addition, we recommended that they lower their campaigning in April, May, and June and making sure to stay as engaged with their clients as possible because the higher the duration of their phone calls the more likely a person is to invest. Also, the need to focus on renewals and targeting those who have already purchased or invested. Explaining these types of data-driven business decisions was not something I had a ton of experience doing, but I do have a ton of experience in addressing people and speaking to a crowd, so it was not too rough of a transition for me to do. The biggest challenge was making sure to keep every recommendation grounded in what the data was telling us from our analysis, rather than making hypothesis that extrapolated off of what our findings were.

1. IST 707 has been one of the most impactful classes for me during my time in the

program. One of the reasons for this was the need to write every homework and report in a business sense, taking myself out of the analysis and getting straight to the heart of what all of the analysis that I did equated to for the end-user and effectively communicating that and the key takeaways. For my final project in this class we had to give a presentation that touched on every area of our project. The struggle was in doing so in a manner to which any individual in an organization, from IT, to managers, to sales people, to developers, and execs could all understand what was done and why and what the actionable items are from the presentation. Our team was able to accomplish this by using a slightly longer presentation and bringing in outside visualizations to create a firm contextual standing where everyone could then grasp what we were explaining. We dove into a bit of the innerworkings of our program but focused mostly on results and what they meant overall. We then included real-life examples of host countries and rotting stadiums and huge host costs to further drive home our points. Lastly, we concluded with our main takeaways. Again, speaking to others or in front of a lot of people has always been a strength for me, so for this it was not a challenge. The challenge came it constantly keeping the end-user in mind and making sure to include relevant information on every slide, and in every word coming out of our mouths.

1. During my coursework in IST 623 Introduction to Information Security, I sought to

expand on a limited knowledge that I had on how cyber-security worked. I knew what it was but had absolutely no idea how it worked, through what mechanisms, and what tools professionals used to protect computers and protect companies from breaches. In our final project, my team researched and presentation on the infamous Marriot Data breach that happened within the last year. We spoke at length on how the breach happened, and what through what channels the information was accessed through. We also hypothesized exactly how the hackers would have gotten in. Lastly, we took a look at the ramifications, the legality, and the consequences that Marriot and its customers would have to face because of this situation. Throughout the course and specifically during this project, I learned just how accessible data is and how easy it is to be breached if a company is not vigilant and constantly aware of its security protocols and systems. My favorite quote that we used in our presentation I think sums this up very well, “A company has to correct every single time (in regard to its security) a hacker only has to be correct once”. This project also brought up conversations about the ethical side of data science. Just how much data is too much to collect on people and how are companies using our data. In the ramifications portion of our presentation, we touched on the GDPR and how it is changing privacy laws in Europe and how the Marriot breach could be the first breach to have punishment levied on it through the GDPR legislation.

Over the course of my time in this program, the amount that I have learned has been incredible. I am so thankful that I have done/am doing this program because when I think about the knowledge I had a year ago today, before I started, to where I sit now, there is no comparison. I believe that having this program and the knowledge it has afforded me will enable be to be a very competitive applicant to jobs in the future. Furthermore, I think these skills will serve me for the rest of my career. Lastly, this program has opened so many more avenues of data science for me to explore and has certainly put me on a path of a lifetime of learning.