Individual Capstone Assessment – Eli Pappas

My senior design project aims to create a data-driven tool for sports statistical analysis, combining data, machine learning, and web presentation. For our senior design project, I knew that I wanted to do something pertaining to statistics, data analysis and, along the same lines, machine learning. I also have a passion for sports as I’ve always loved looking through box scores from my favorite teams, so this was the perfect intersection of the two. This project is going to be about creating a tool and a site where people can go to see projection data on what we believe is a likely performance both from teams as well as individual athletes. We plan on using a machine learning model created either from TensorFlow or PyTorch to utilize large datasets from public sources such as MLB Savant or other APIs which offer this data for free. We also plan to incorporate sports betting lines, such as over/unders, or money lines and compare them to our projections to determine whether or not people should bet on a certain thing. Finally, I would like to take all our data we gather and put them out there on a publicly hosted website where we can give easy to read visualizations of our data so that other people can use it as well.

Throughout my time at UC, I have not been able to take any courses covering machine learning, but this is a topic which I have long been interested in. I have however taken courses such as Intelligent Data Analysis (CS 5152) where I learned the basics of important topics like data mining. I also took Visual Interfaces Data (CS 5124) where I learned the best practices for presenting data in an easy to read way. I am currently taking User Interfaces (CS 5167) where I have been learning the basics for web design and how to create a good and easy to use interface. These classes have all driven me in the direction I’m heading in with this project, but I have taken classes for my minor in mathematics as well such as Data Science and Statistics (STAT 3041) which will help me to understand how to clean and use the data which we will be importing. I believe that my classes have all prepared me for this project and served as a jumping off point for learning the new skills required, such as the basics in web-design and interpreting data from UI and VID, or the python data manipulation skills from Data Science and Statistics, as well as the data literacy which I developed through IDA.

My Co-op experiences have also formed my learning style as well as my skillset. I have done all 4 of my rotations throughout college with Siemens Digital Industries and I was a part of 2 different teams. My first team I learned how to work professionally as well as work well with others outside of a school environment. I also learned about how to manage deadlines and picked up programming skills in C++, Java, and C#. I had no prior experience in C# or Java, so I had to pick them up quickly, which is another skill I need to use for my senior design project. My last two rotations were on a software engineering team where I made changes to the core functionality of some features in NX, a 3D modeling software. I had to work through different sections of code, work independently on projects, and use my time learning and developing at the same time. I also was expected to jump from project to project and I was able to come up to speed quickly. This experience has taught me how to learn efficiently and how to manage myself and my time in the most productive ways. While I don’t plan on using many of the technical skills I picked up, I believe that the thought processes I used and learned through work will help me tremendously on this project.

I sort of jumped into my motivations for this project a bit in the first paragraph, but sports have been a big part of my life since I was younger. Since I’ve gotten older and more experienced in math and data, I have been more and more intrigued by statistics and their impacts on how players perform, such as ERA’s in baseball vs lefty or righty batters which can have large discrepancies or even how certain players bat against specific pitches (like a slider). This can be really good to predict how that player may perform, and I believe that the market for data driven decision making has only expanded, especially so with the rise in popularity of sport betting as it continues to be legalized in more and more states. I am really excited to start on this project because like I said, I really enjoy data and analysis based off of it, and this provides a way for me to learn even more techniques that are necessary to get better insights into the games which I love to watch. This also is a great resume project for me and a great learning opportunity, especially as I aim to break into the data science field out of college with no experience to speak of. Finally, I am super excited to test my models out myself and share them with friends to try to get an edge on sports books in my own personal endeavors in that part of sports.

Our preliminary approach to this project is to start out by gathering as much data as possible, starting out with one sport and one part of the sport, such as batting stats for the MLB, or win percentages in the NFL, then to clean the data and build a simple ML model to predict upcoming events. After we have that accomplished I think the next step would be to test our predictions and improve the model and while one person is doing that, the other will be starting on the website and putting our predictions up on the web for others to see. If we are able to accomplish that in a good amount of time, then I think we would move on to creating more models for different stats and expanding for as much time as we have. I think to self evaluate my contributions I will look at the complexity of the part of the project as well as the time I spent working on it and compare it to my teammates. I will know that we are done when we have a working website with models that can accurately display projections and I would like for those projections to be somewhere around 75-80% accurate. I don’t think you can ask for much more than that especially in sports where literally anything could happen. I will be really happy if we are able to create something that I personally use weekly and if it’s something that I want to further perfect even after I graduate.