**Capstone 2 Project Ideas**

**Pick a problem, identify the issue to address**

1. **Streaming Services Analysis and Genre Prediction**

There has been a rise of streaming services over the last few years. Some of the more popular ones are Netflix, Hulu, Disney, and Prime Video. I think it would be interesting to get an idea of the types of movies on these streaming services. What is their average rating? Are they relatively older or newer movies? Which streaming service has more foreign movies available? I also think it would be useful to employ some machine learning models to predict a movie genre from its title.

Data:

<https://www.kaggle.com/ruchi798/movies-on-netflix-prime-video-hulu-and-disney>

1. **Tennis Match Exploration**

Sports are always interesting to analyze. I particularly enjoy tennis, and I think it would be interesting to explore a tennis dataset. One important factor in a tennis match is the surface of the court. Is there a correlation between court surface and number of minutes played? In tennis, it’s not always the case that the winner of the match has won the most total points. How often does this occur? The dataset I found has data for the years 2000-2019. Have any statistics changed over that time period? Finally, I could employ machine learning models to predict the winner of a match based on the match statistics (number of aces, number of break points, etc.)

Data:

<https://www.kaggle.com/pablodroca/atp-tennis-matches-20002019?select=atp_matches_2000.csv>

[https://archive.ics.uci.edu/ml/datasets/Tennis+Major+Tournament+Match+Statistics#](https://archive.ics.uci.edu/ml/datasets/Tennis+Major+Tournament+Match+Statistics)

**3.  Knowledge Tracing of student knowledge over time**

Track knowledge states of 1M+ students learning through the internet

Background: In 2018, 260 million children weren't attending school. At the same time, more than half of these young students didn't meet minimum reading and math standards. The task is to create algorithms for "Knowledge Tracing," the modeling of student knowledge over time, the goal of which is to accurately predict how the students will perform in future interactions.

* Data Description: The same sorts of information a complete education app would have, including student's historic performance, the performance of other students on the same question, metadata about the question itself, and more.

Data: <https://www.kaggle.com/c/riiid-test-answer-prediction/overview>