Athlete Forecast Writeup

<u>Initial Thought Process</u>:

As a team of four women, we wanted to tackle a track that would challenge us while making a statement in a field typically dominated by men. We brainstormed various ideas in the sports track, ultimately landing on a topic with which our teammates each had personal insights towards.

<u>Problem</u>:

Bianca Jayaraman, a valued member of our team and an integral part of the Georgia Tech Football Media Team, identified a critical challenge in the recruitment of high school athletes. Despite the substantial revenue college football generates for universities across the U.S., the current recruitment process suffers from a lack of standardization in athletic scouting. This deficiency has resulted in an increased number of athlete transfers and cuts, highlighting the urgent need for a more systematic and effective approach to scouting and recruitment.

<u>Solution</u>:

Our aim is to establish a more reliable benchmark for recruiters to evaluate potential college recruits. Recruiters often rely on highlight tapes of varying quality, a discrepancy that can often be traced back to socioeconomic disparities among recruits. Recognizing this, we sought to develop a tool that enables recruiters to objectively assess a high school athlete's potential through quantifiable data: height, weight, star rankings, and overall ratings. This approach ensures a fair and standardized method of evaluation, leveling the playing field for recruits from diverse backgrounds.

<u>Implementation</u>:

To simplify the introductory development of this application, we focused on the position of Wide Receiver.

We began by extrapolating datasets from the [College Football Database 2019-2023](#) to gather the four years of recruitment data that would comprise the current College Football Data 2023 NCAA Wide Receiver players. We cleaned the dataset to include only notable statistics from this website: height, weight, stars, and rating associated with each high school recruit that went on to play in the college.

Our application accepts the input of a high schooler with these four criteria. It then compares these statistics with existing NCAA players' statistics from their time of college recruitment (available in our dataset). When a collegiate player with statistics that resembles those of the potential recruit is found, the application outputs this player. The application also displays the players' current performance in the 2023 season, including rushing yards and touchdowns. We extracted a separate [ESPN 2023 Wide Receiver Stats](#) dataset in order to access this information.

For the front-end design of our web application, we utilized HTML, CSS, JavaScript, and Python to craft a user interface that stands out in its simplicity and vibrant color scheme. Initiating the process, we meticulously sketched out our vision for the background, bringing it to life with custom imagery created in Adobe Photoshop. Our primary goal was to curate an intuitive user experience, steering clear of conventional gender-normative design principles prevalent in athletic applications. This approach ensures our platform not only captivates but also inclusively serves a diverse audience, setting a new standard for design in the domain.

We began the implementation of our algorithm by experimenting with similarity algorithms such as K Nearest Neighbor and Cosine Similarity models, but eventually found Euclidean Distance Calculations to be the most accurate and efficient means of analysis. This algorithm converts the given input and data into vectors in $\mathbb{R}^4$. These vectors are then normalized to remove all weight from different units of measure. Following this, we aimed to provide the BMI benchmark by increasing the magnitude of axes related to height and weight. Similarly, we weighted rating and star axes for the Skill benchmark. The Euclidean Distance between the weighted input vectors and their respective weighted data points were calculated and placed in a dictionary. Finally, the minimum distance between the weighted data point and the weighted inputs was returned, and the college athlete that mapped to this minimum was outputted.

Future Expansion:

Our ambition is to evolve our application into a comprehensive tool that supports all positions within American college football. Moreover, we aim to integrate advanced machine learning algorithms to refine the process of evaluating the significance of various statistics in athlete comparisons. This enhancement will empower recruiters nationwide with the ability to swiftly and accurately assess the potential of high school athletes using readily accessible data, streamlining the recruitment process and ensuring a more effective talent identification strategy.