GitHub Link: <https://github.com/jfiengo/Michael_Jordan_ML_Project>

In our current time, machine learning has become an extremely powerful tool that we rely upon in many areas of our lives. It has the capability to recognize patterns in data that exceed far beyond the capability of humans. The purpose of our project was to showcase the incredible power of machine learning in the context of sports by applying it to a player to see if it could accurately predict the outcome of the game.

We chose Giannis Antetokounmpo, an NBA player for the Milwaukee Bucks as our dataset and took statistics from his 2021-22 season as the training set, as well as his 2022-23 season for the testing set. Our goal was to see if a machine learning model could accurately predict the outcome of a game based on Giannis’ statistics during the game. We chose Python’s scikit learn package, because it contains many different machine learning models that we could experiment with. In our X lists, we put several stats including home/away, points, assists, steals, blocks, and field goal percentage. In our Y lists, we put whether that specific game was a win or loss. To help us with all the data manipulation required, we utilized the pandas package. The head() method allowed us to display the data after each change. We needed to change the way some of the data was represented, such as extracting W/L out of a column containing additional information such as the point spread.

Once we had our datasets ready, we could finally insert them into a machine learning model to evaluate its effectiveness at predicting the outcome of the game. We chose a logistic regression model as well as a random forest cluster model, yielding very similar results. We wrote code to make a list of all the different combinations of data features that could be used to test and train the model. We then tried every combination of features, by fitting the model to the training lists and running the predict() function on the test lists. The data feature that proved the most useful was the ‘+/-‘ statistic. This is a representation of how many points the Bucks scored over their opponents when Giannis was in the game. Utilizing this statistic in the X list allowed us to get an average mean error of 0.14% using both logistic regression and random forest cluster algorithms.

We split up participation so that both members utilized their abilities well. Michael wrote starter code implementing data manipulation and a scikit learn machine learning model. Jack improved upon the data manipulation by using a .csv instead of a .txt file. He also added the pandas package to help with the data manipulation, as well as trying different machine learning packages. Michael wrote the documentation. Jack wrote a function to test different feature combinations and choose the one with the least mean absolute error.

This project highlights the incredible power of machine learning algorithms. Data analysis through machine learning can be used in sports betting to maximize chances of receiving a payout. Many sports bettors place combined bets including a win/loss bet along with one or multiple bets on player statistics. A machine learning model can help a bettor pick the right stats to include. Similarly, machine learning can be utilized by the actual teams playing the sport. Coaching staffs can use data analysis to decide whether to sub in a certain player late in the game, or if they should start the player.