//background and intro

The UEFA Champions League is a yearly international club soccer tournament featuring the best teams in Europe. Several machine learning models have been used to predict the results of sports games. Bunker and Thabtah [1] found that using an ANN with backpropagation performs with a slightly higher accuracy than expert predictions, beating ESPN sportscasters 75% to 63% accuracy when predicting NFL games. According to Zhang et al [2], RNNs can be unable to capture long-term dependencies. They recommend the use of LSTM, a variant of RNN, to get more accurate results. Yue et al [3] discuss the use of a Bayesian approach called GLICKO to predict strength of tennis players and therefore outcomes of tennis matches. In our particular use case, the unique format of the Champions League allows for more complex predictions of team knockout stage performance based on domestic league and group stage performance.

We want to create a model that accurately predicts the team knockout stage. Despite soccer being the most popular sport worldwide, more money is put into analytics for the NBA and NFL in the United States. The field of predicting soccer matches is scarce. Additionally, since UEFA only has European teams, it receives less resources than other soccer events such as the World Cup. An accurate model could help teams allocate their resources, knowing which factors are important for a winning team. It can also allow a team to see where weaknesses lie, allowing them to improve on their deficiencies.

//methods

We plan to use a neural network to implement our solution. More specifically, we will first look into using an RNN since it can predict future results based on past data. We can implement this using an existing library like TensorFlow with Python, which does well implementing RNNs. If the results are not sufficient, we can consider switching to an alternative method like LSTM.

//potential results and discussion

While we are still determining all the factors that we want to consider, we do know we want to analyze season performance, how home vs. away affects teams, and team’s past performance in the tournament. We expect that a team who has had a winning season will do well in the tournament and that being the home team would give a slight advantage. Additionally, teams who have a history of doing well in previous tournaments perform better than those who seldom make the tournament, as they are used to the pressure and environment.

//checkpoint

At this point in the process, we have defined the problem, brainstormed machine learning methods, and found our dataset. The [dataset](https://www.soccerbase.com) contains results from UEFA games and group stage results, as well as modern games’ statistics like possession and shots. By the midterm report, we hope to scrape and clean the data, create one supervised learning model, and determine how we will improve on this model. By the final report, we hope to deliver multiple supervised models, one of which can accurately predict knockout stage performance.

[1] R. P. Bunker and F. Thabtah, “A machine learning framework for sport result prediction,”

*Applied Computing and Informatics*, vol. 15, no. 1, pp. 27–33, 2019.

doi:10.1016/j.aci.2017.09.005

[2] Q. Zhang *et al.*, “Sports match prediction model for training and exercise using

attention-based LSTM network,” *Digital Communications and Networks*, vol. 8, no. 4,

pp. 508–515, 2022. doi:10.1016/j.dcan.2021.08.008

[3] J. C. Yue, E. P. Chou, M.-H. Hsieh, and L.-C. Hsiao, “A study of forecasting tennis matches

via the GLICKO model,” *PLOS ONE*, vol. 17, no. 4, Apr. 2022.

doi:10.1371/journal.pone.0266838