Leveraging Machine Learning to Forecast NBA Season MVPs

# Identify Need:

Creating a machine learning model to predict the NBA Season’s MVP (Most Valuable Player) can be important to various stakeholders such as society, sports enthusiasts, and the overall economy. Sports have been a huge part of modern society and being able to predict the NBA MVP can help increase engagement to the sport of basketball. By using data to effectively predict the future NBA MVP of the season this project can help generate more excitement and interest among fans. This also connects to the growing field of sports betting industry (valued at over 195 billion), being able to use data driven modelling allows fans to make better decisions with their bets or fantasy teams. They can benefit from these accurate predictions to make informed decisions when choosing players for their bets or teams. Increasing fan engagement can benefit the sport industry and the economy. In addition, sports enthusiasts, coaches or even players can benefit as the model can provide a more quantitative measurement of a player performance instead of relying on votes that are currently used to determine who is MVP of that season. The NBA is a multi-billion-dollar industry and accurate MVP predictions and more data driven decisions can increase engagement, viewership which positively increases the economy due to their economic contributions. There are tons of people that benefit from this project, the NBA, coaches, players, sports fans and even the overall economy can benefit from this initiative. The NBA can see much more engagement with their fans, coaches and players have a better metric to gauge their performance across a season as well as sports fans have a better way to make informed decisions on players they chose to support or bet on.

# Approach:

Using the nba\_api, an API client package designed to access NBA.com that can retrieve official NBA statistics, such as players’ career stats and living data like the scoreboard. Predicting NBA MVP involves a supervised learning problem where the model is trained on a labeled dataset. This dataset associates input features, such as player statistics and team performance metrics, with whether a player wins the MVP award. Assuming a linear relationship between features and the MVP outcome, linear regression is a potential ML approach. However, sports analytics often involve complex patterns or non-linear factors. In such case, decision trees can be a suitable ML approach. Decision trees handle non-linearity and measure feature importance, helping identify which player statistics or metrics have the most significant impact on the MVP prediction. It is important to consider these factors when developing predictive models for the dynamic landscape of NBA performance.

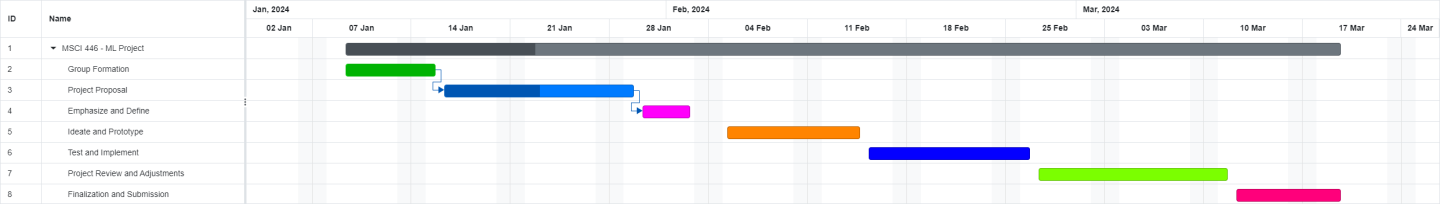
Comparison and Validation:

There are a few ways we can compare and validate the test results of our ML model. Firstly, we can implement k-fold cross-validation to assess the model’s performance on different subsets of data (try different combinations of training and test splits), to reduce the risk of overfitting or underfitting. Similarly, we can apply temporal validation on the data since it may span multiple seasons over a few decades. We can train the model on data from earlier seasons and validate it on more recent seasons to simulate real-world scenarios and prevent recency bias. We can also apply the statistical analysis to our training data and the test observations for different models and compare the bias-variance trade-off as the flexibility increases. Another approach is hyperparameter tuning by performing hypothesis testing and understanding which parameters have a larger influence on the model and seeing how they interact with one another.

# Potential Risks:

Other determining factors for the MVP, besides scores and team success, which are usually not easily extracted from datasets, include leadership and media votes. These factors can contribute to variations between the model's predictions and actual results. While quantifying the overall impact of players is possible, it inevitably introduces bias and inaccuracy due to subjective assessments. Digging deeper into these factors can be time-consuming and have a risk creating an overfitting model with results that will not generalize well. Furthermore, the model's predictions are vulnerable to unforeseen circumstances, such as injuries to potential candidates or negative news from players. Ignoring such outlier events may compromise the reliability of the model. The interpretation of these events still requires human judgment to determine their impact on a player's candidacy for the MVP.

# Design Thinking and Project Management:

Our group will combine efficient project management techniques with design thinking concepts to guarantee a successful project submission. Our methodology starts with understanding the project's goals and success criteria. Next, we conduct brainstorming sessions to produce original concepts for a innovative machine learning model that predicts the NBA Season MVPs. Effective communication is essential and can be encouraged by holding regular team meetings and using collaboration tools to ensure an honest and open discussion. Tasks will be assigned based on the team members' areas of technical and/or non-technical expertise. Using a Gantt chart, we will work through ideation and prototyping, task definition, testing, and machine learning model implementation. Feedback loops and ongoing testing will drive iterative improvements. Before concluding, the last steps entail assessing the project's progress, making any required modifications, and communicating clearly. This structured plan adheres to design thinking principles, emphasizing a user-centric approach, creativity, collaboration, and adaptability throughout the project lifecycle.