

# MATH 748 Project Proposal : Predicting the outcomes of NBA games

## I. Team

Mihir Kotecha - 923641248

## II. Description of the problem

The objective of my project is to develop a predictive model that can forecast the outcome of an NBA game using detailed statistics and player performance metrics. This model will aim to predict which team will win each game, potentially offering insights for applications such as fantasy sports strategies and betting odds. By analyzing key team and player statistics, my project seeks to identify which variables, such as points scored, assists, rebounds, or turnovers, are most consequential for a win. Additionally, this project will analyze how contextual factors, such as player injuries or home court advantage, can influence match outcomes. Another important aspect of the research will be to determine whether it's possible to predict the margin of victory based on pre-game statistics.

## III. Description of the data

The dataset contains detailed information on NBA games from the 2004 season to the 2020 season, including game dates, teams involved, and the number of points scored. Additionally, it provides player specific statistics for each game, such as points, rebounds, and assists, along with general player details. The dataset also includes the rankings of NBA teams by conference on a given day, providing insight into team performance across the season. Team-specific information is also available, providing a wider view of the NBA's structure, which is ideal for predicting match outcomes based on their historical performance.

## IV. Supervised or unsupervised?

The project will use a **supervised learning** approach. Since the task involves predicting match outcomes, which makes it a classification problem where the target variable is whether the team wins or loses. Historical game data, including player and team statistics, will work as input features, and the model is trained to recognize patterns. Regression techniques can also be used to predict the point margin of victory.

## V. Additional comments

A potential challenge with this project is handling any missing or incomplete data, mostly for the player-level statistics. There can also be some imbalance in the outcomes, such as games against teams with a strong record, which could lead to biased predictions. Another concern I have is that external factors like player injuries or even rest days may not be fully captured in the dataset. Also, feature selection will be critical to ensure that the model can capture the most relevant variables.