# <u>Logan Laszewski - March Madness Project March 2025 - ReadME</u>

# **Project Goal**

I've always loved watching March Madness and eagerly anticipated the first two days of the tournament — when anything can happen, and one unexpected upset could destroy your bracket.

As I've developed a passion for Machine Learning (ML), I wanted to try my hand at building a model to predict these unpredictable games. I knew it would be challenging since many upsets may not be visible in the data. However, this project was an exciting opportunity to combine my love for sports with my interest in data analytics and data science.

### **Project Overview**

I used the March Madness Kaggle data from the March Machine Learning Mania challenge. While I'm not submitting entries to the competition, the datasets provided a solid foundation for building my models and making predictions.

## **Data Split**

Training Data: 2003-2020 tournaments
Test Data: 2021-2024 tournaments

This setup allowed me to evaluate how accurate my predictions were for recent

tournaments.

### **Data Cleaning and Transformation**

Before running the models, significant data cleaning and data transformation were required. I worked extensively with team-level and opponent-level stats to prepare the data for modeling.

# Handling missing values and inconsistencies

Creating new features to capture team performance dynamics Normalizing and scaling variables for optimal model performance

#### **Model Choice**

For the initial predictions, I used XGBoost The model was trained on historical data to predict whether a team would win a given matchup. Also, experimented with h20 to look at other models and how they performed.

### **Datasets**

MNCAATourneySeeds.csv - To get seeding in the Tournament

MTeams.csv - To get all unique Team IDs and Team Names

<u>MRegularSeasonDetailedResults.csv</u> - To look into stats like Points for/against, Rebounds, Turnovers, Fouls, and more

<u>MTeamConferences.csv</u> - To grab the conference team is in

<u>MNCAATourneyCompactResults.csv</u> - To see actual tournament matchups and outcomes which will be used to train and test the model

MNCAATourneySlots.csv- To get this year's March Madness matchups

#### **Locate Files**

The datasets, my code, and my Diary are all located in my GitHub repository for the project. <a href="https://github.com/Logan142414/MarchMadness2025Model">https://github.com/Logan142414/MarchMadness2025Model</a>

# **Packages Needed**

Many packages are needed for this project. Common ones just like pandas and numpy. Align with sklearn packages for accuracy\_score, precision\_score, log\_loss. Furthermore, xgboost, matplotlib, ParameterGrid, RFE from sklearn, and h20.