Team Name: Bo Nix

Group Members:

- Benjamin Getches
- Brandon Reeves
- Ethan Epperson

Problem Space

Our team is stuck between two ideas that we are passionate about solving:

Predicting NBA/NFL Stats

- Problem to Solve:
    - Given a dataset of historical statistics across a large set of sample games, how can we develop a model to make predictions about player / team performance?
    - What features of sporting statistics offer the most value in predicting performance?
- Interest in Problem:
    - Predicting stat lines, player performance, and or team performance has usage in several contexts (*e.g. fantasy sports, sports betting, etc.)*
        - Strong model performance can provide go-ahead value in these contexts
    - There are already solutions to this problem, providing tried-and-true solutions to compare our model to
- Usefulness of Machine Learning in Solving Problem:
    - There are many quantitative features that can be leveraged to develop predictive models
    - Sporting games, like many other games, have patterns which provide indication of performance
        - *e.g. in soccer the team with more shots is often the winner of the match.*

Reinforcement Learning for a 2D Game

- Problem Being Solved:
    - Given a game with a clear objective and simple controls, how can an ML model win the game, and iterate/improve upon its solution?

- Interest in Problem:

    - It would be interesting to see objectively what strategies and approaches result in the best outcomes when playing.
    - Inspiration for this idea comes from videos like this where models learn and adapt to succeed in different games: https://youtu.be/kopoLzvh5jY?si=Ad-MmFwX_D-0fvPJ
    - And https://www.youtube.com/watch?v=pJPdW8WWAso

- Usefulness of Machine Learning in Solving Problem:
    - By setting correct reward functions and allowing our ML model to learn from itself, we can show that their ability can come close to or even surpass humans
    - ML models should be able to identify and create unique solutions to problems in a game as their lack of exposure compared to humans will create less bias, and the way they learn is different to a human player. In addition, since its ML, the input they receive must also be modified.
    - Traditional programming fails to capture complexity needed to use bots in games.

Data / Data Plan

Reinforcement Learning:

- Reinforcement learning based on a game would require us to create an environment and run iterations of the model playing a game to generate results that would then be used to train the model.
    - Features input into the model would mainly revolve around game state information.
        - Agent's position: Features like its velocity along with its position relative to things in the world around it would be useful. If this were trained off of a 2D platformer, it would be useful to have information like distance from end flag or from different platforms
        - Rewards: We would want to reward/punish different behaviors that either contribute or hurt the agent's progress. In a 2D platformer, this could look like rewarding for level completion and going in the right direction, while punishing for deaths and time taken.
    - We would want thousands, if not tens of thousands of iterations of training data.
    - The output of the model would consist of the agent's actions at each timestep

    - We believe the model will be based on policy learning and a reward function over time. Ideally, we want the model to be able to "play" the game and not just memorize solutions. So we need to be wise about the amount of outputs we give our bot.
    - RL means we don't have supervised/unsupervised learning, and instead the bot must learn from itself using trial and error and iterations.
        - We could let the bots learn in evolutions where top performers survive and worse bots die off.

Predicting Sports Stats:

- Potential data sources:
    - Pro Football Reference: https://www.pro-football-reference.com/
        - Large, well-maintained database with historical data dating back decades
        - Provides player and team statistics, game logs, draft history, etc.
        - Critical Features:
            - Examples from 2024 Advanced Passing Metrics:
                - Completed Air Yards per completion
                - Percentage of bad throws per pass attempt
                - On target percentage
                - Time pressured per drop back
            - Generally, statistics measures that have a large impact on player performance
- Outcome (regardless of college / professional football statistics):
    - Supervised
        - Utilize historical data with input features (football statistics) paired with target variables (fantasy points, player rating, etc.) to allow model to learn relationship between input and output features
    - Regression process
        - Focus of model is on predicting a continuous variable like fantasy points
- Dataset:
    - Although we do not have a concrete dataset in place, here is some general key information
        - Each team plays 17 games across the NFL season
        - There are roughly 1700 active players in the NFL
        - For each player there are 10+ key statistics that may provide predictive value

- These pieces of information suggest that are dataset may have the following shape (10000, 10)
    - This shape takes a conservative approach under the impression that we may not evaluate every single player in the NFL
    - We likely should remove players who get injured under our training data.