Weekly Report (1/3/25)

This week, we explored different approaches and how they can be potentially implemented in our project.

ANFIS

ANFIS is a hybrid AI model that combines Fuzzy Logic and Neural Networks. It can handle uncertainty in player performance, learn decision-making rules from data, and provide interpretable results for recruiters and coaches. Unlike traditional machine learning models, ANFIS allows for a flexible approach where player performance is evaluated on a range rather than as a fixed classification.

Planned Implementation of ANFIS:

1. Identifying Team Strengths & Weaknesses (Offense/Defense)
We will define fuzzy membership functions for offensive and defensive performance categories (e.g., weak, moderate, strong). ANFIS will be trained to analyze how different statistical combinations contribute to overall team effectiveness.

2. Profiling Individual Athletes
Player performance metrics such as points, assists, and rebounds will be used to train ANFIS. The model will assign a Player Strength Score, capturing both offensive and defensive impact.

3. Predicting Optimal Lineups Against Opponents
ANFIS will analyze past matchups, player performance trends, and opponent strategies to generate lineup recommendations. Fuzzy rules will determine which players should start, rotate, or be benched.

4. Athlete Clustering – Identifying Patterns
ANFIS will group players into categories like Elite Scorers, Defensive Specialists, and All-Rounders, helping recruiters find players that best fit team needs.

Pros & Cons of Using ANFIS

Pros

- Handles Uncertainty – Accounts for variations in player performance.
- Interpretable Results – Provides insights into why a player is suited for a role.
- Works with Small Datasets – Suitable for NCAA data

Cons

- Computationally Expensive – Training takes longer than traditional models.

- Requires Careful Feature Selection – Too many inputs can slow learning. We will have to reduce the features ourselves.
- Not Ideal for Pure Score Prediction – Traditional models like XGBoost may better predict exact game scores.

XGBoost

XGBoost (Extreme Gradient Boosting) is a machine learning algorithm optimized for structured data, widely used for predictive modeling in sports analytics. It builds multiple decision trees sequentially, learning from previous errors to enhance prediction accuracy. Its ability to handle missing values, large datasets, and complex feature interactions makes it well-suited for analyzing athletic performance.

Planned Implementation of XGBoost

1. **Identifying Key Performance Factors**
   We will train XGBoost on historical player statistics to determine the most influential features affecting performance. Using SHAP (SHapley Additive exPlanations) values, we can interpret which factors—such as sprint speed, reaction time, shooting accuracy, or endurance—have the greatest impact on overall performance.

2. **Predicting Match Outcomes & Player Contributions**
   By analyzing past match data, XGBoost will predict game outcomes based on team composition, player form, and tactical playstyles. It will also estimate the influence of individual athletes, providing probability-based insights into how their performance affects game results.

3. **Real-Time Performance Evaluation**
   We will integrate XGBoost with real-time performance tracking data, allowing coaches to make data-driven decisions during matches. Features such as player fatigue, heart rate, and movement efficiency can be used to predict when substitutions might be necessary.

4. **Athlete Profiling & Role Classification**
   XGBoost can be used to classify athletes into categories like Defensive Specialists, High-Impact Scorers, or Playmakers based on statistical patterns. This will assist in scouting, lineup selection, and game strategy optimization.

5. **Momentum Analysis in Competitive Matches**
   Inspired by research in tennis match dynamics, we will analyze momentum shifts in team sports using XGBoost. By tracking scoring streaks, defensive efficiency, and player

workload over time, the model can detect shifts in team dominance and recommend strategic adjustments.

Pros & Cons of Using XGBoost

Pros:
High Accuracy – Excels in numerical prediction tasks, such as game scores and player rankings.
Feature Importance Analysis – Provides insights into which factors contribute most to performance.
Handles Missing Data – Learns effectively even with incomplete player statistics.
Efficient on Large Datasets – Can process extensive game records without significant performance drops.

Cons:
Less Interpretability – While SHAP values can improve interpretability, decision trees lack the natural readability of rule-based models.
Data-Intensive – Requires high-quality training data for optimal results.
Feature Engineering Required – Performance depends on selecting and transforming the right input features.

By leveraging XGBoost, we aim to develop a **robust athlete performance prediction system** that provides **data-driven insights for coaches, analysts, and recruiters**.

Neurosymbolic AI

Neurosymbolic AI refers to AI systems that seek to integrate neural network-based methods with symbolic knowledge-based approaches. This integration aims to develop AI systems capable of learning from data and performing complex reasoning tasks, thereby enhancing robustness, interpretability, and versatility.

Neural networks excel at pattern recognition and learning from data. On the other hand, symbolic reasoning involves the manipulation of symbols and rules to represent knowledge and perform logical reasoning, enabling tasks such as problem-solving and understanding abstract concepts. By combining these two paradigms, neurosymbolic AI seeks to create systems that can learn efficiently from data while also engaging in high-level reasoning.

In our project, Neurosymbolic AI can be implemented using neural insights to find patterns in the player dataset and combine them with symbolic reasoning to evaluate different line-up configurations. This approach would help provide coaches with explainable recommendations for the optimal lineup, detailing the rationale behind each suggestion.

References for Neurosymbolic AI:

1. [2305.00813] Neurosymbolic AI -- Why, What, and How
2. Neuro-Symbolic AI in 2024: A systematic review
3. Surveying neuro-symbolic approaches for reliable artificial intelligence of things

Plan for Next Week

We will be unable to make progress for the upcoming week due to mid semester exams.