Weekly Report (22-02-25)

Overview

This week our team focused on exploring the datasets provided and literature survey to gain insight on potential techniques to use for the project.

Feature details

Abbreviation	Statistic	Description
PTS	Points Scored	Total points scored by the player in a game.
MIN	Minutes Played	Total time the player was on the court.
FGM	Field Goals Made	Successful two- or three-point shots.
FGA	Field Goals Attempted	Total shots taken.
PM	3-Point Field Goals Made	Successful three-point shots.
PA	3-Point Field Goals Attempted	Total three-point shots taken.
FTM	Free Throws Made	Successful free throws.
FTA	Free Throws Attempted	Total free throw shots taken.
OREB	Offensive Rebounds	Times player regained possession after a missed shot.
DREB	Defensive Rebounds	Times player secured the ball after opponent's missed shot.
REB	Total Rebounds	Sum of offensive and defensive rebounds.
AST	Assists	Passes leading directly to a made basket.
BLK	Blocks	Shots blocked by the player.
STL	Steals	Times the player took the ball from an opponent.
TO	Turnovers	Times the player lost possession of the ball.
PF	Personal Fouls	Number of fouls committed.
Team Score	Team's Total Points	Total points scored by the team in the match.
Win	Win Indicator	1 if the team won, 0 if the team lost.

Game Score Performance metric based on Hollinger's Formula (PTS, AST, STL, etc.).

Literature Survey

Sharma, S. U., Divakaran, S., Kaya, T., & Raval, M. (2022, July). A hybrid approach for interpretable game performance prediction in basketball. In 2022 International Joint Conference on Neural Networks (IJCNN) (pp. 01-08). IEEE.

Sharma et al. propose a Decision Tree-based approach to predict and interpret game performance in basketball. The study focused on a Women's Division-I basketball team, analysing data from 16 athletes over 25 weeks, encompassing 2,800 records with 37 attributes. These attributes included sleep and recovery metrics from wearable devices, training statistics, and cognitive assessments. The researchers developed a hybrid model combining factor analysis with decision tree-based methods, including classification/regression trees and random forests. This approach first reduced the dataset into seven latent factors through factor analysis, then utilised these factors to construct decision trees for game score prediction.

Taber, C. B., Sharma, S., Raval, M. S., Senbel, S., Keefe, A., Shah, J., Patterson, E., Nolan, J., Sertac Artan, N., & Kaya, T. (2024). A holistic approach to performance prediction in collegiate athletics: Player, team, and conference perspectives. Scientific Reports, 14(1), 1-10.

The study focused on data from 16 Division-I female basketball players over an entire season, tracking training workload, sleep, stress, and in-game performance. The players were monitored using various data sources: WHOOP straps for training, subjective stress, sleep, and recovery, Polar Team Pro monitors for in-game stats, and countermovement jumps measuring Reactive Strength Index Modified (RSImod). Players also completed bi-weekly stress questionnaires, and performance was quantified using Game Score (GS) at the team level and Player Efficiency Rating (PER) at the conference level. The study used Extreme Gradient Boosting (XGB) models to predict RSImod and GS. An ensemble approach combining Random Forest, XGB, and correlation analysis identified key features influencing performance at

each level. Partial Dependence Plots (PDPs) were used to interpret the impact of training load, sleep quality, and in-game statistics on performance.

Tasks for Next Week

We are planning to make a baseline implementation using clustering techniques like PCA. We will also try to identify the most prominent statistics affecting individual atheletes' performance via feature-selection models, and then try to apply K-means clustering.