NBA GAME PREDICTIONS BASED ON PLAYER CHEMISTRY

Machine Learning Approach to Team Synergy and Game Outcome Forecasting

Varalakshmi K M (PES1UG24AM815) | Malashree H K (PES1UG24AM807)

Introduction

Does Player Chemistry Affect Winning?

Traditional game prediction models rely heavily on aggregated statistics (PPG, RPG, etc.). However, basketball is a team sport where player interactions often determine the outcome.

Problem: To design a predictive model that *explicitly* incorporates **player synergy and rivalry** to predict NBA game win/loss outcomes, going beyond individual player metrics.

Goal: Compare the performance of a custom chemistry model against strong machine learning baselines to quantify the impact of "team chemistry."

Methodology

The Quadratic Chemistry Model (QCM)

We leverage the Quadratic Classifier approach, inspired by Stanford CS229 research, to model interactions between players on the court.The prediction for a game involving a team with lineup vector x is calculated as:

$$h(\mathbf{x}) = g(\mathbf{w}^T\mathbf{x} + \mathbf{x}^T\mathbf{Q}\mathbf{x})$$

Where $g(\cdot)$ is the sigmoid function, and $\mathbf{Q}$ is the core **Chemistry Matrix**.

Decomposition of the Chemistry Matrix ($\mathbf{Q}$)

The chemistry matrix $\mathbf{Q}$ is decomposed into two parts, which are learned during training:

Term	Symbol	Property	Interpretation
Synergy	$\mathbf{S}$	Symmetric ($\mathbf{S} = \mathbf{S}^T$)	Captures **mutual synergy** or rivalry. Player A playing with B is weighted the same as B playing with A.
Anti-Synergy	$\mathbf{A}$	Anti-Symmetric ($\mathbf{A} = -\mathbf{A}^T$)	Captures **directional effect** or rivalry. Player A's performance with B is different from B's performance with A.
Combined	$\mathbf{Q}$	$\mathbf{Q} = \mathbf{S} + \mathbf{A}$	The total interaction effect.

Explainability & Impact

By inspecting the learned matrices S and A, we can gain insight into specific player effects.

Individual Player Impact (w)

The linear weights (w) show a player's general, context-independent value.

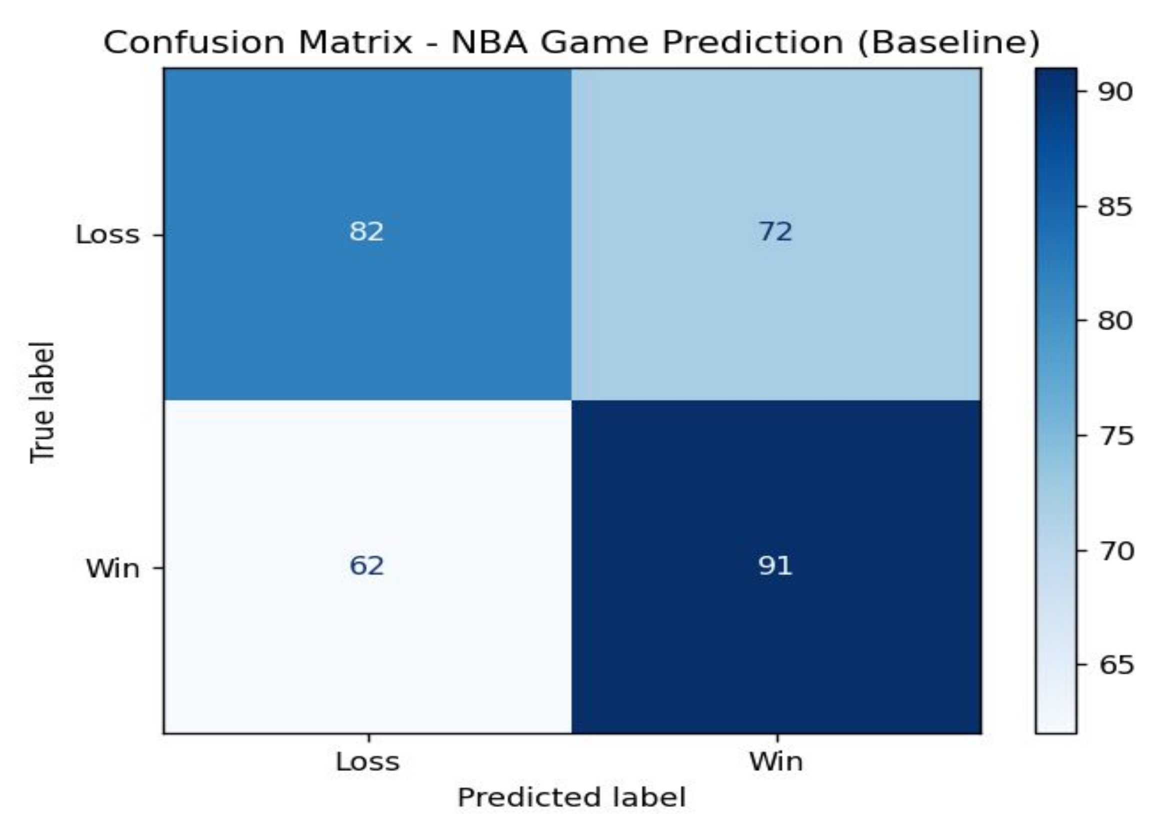
- **High** w**:** Players whose mere presence strongly correlates with a win (e.g., superstars).
- **TCI Feature Weight:** The w weight for the "Chemistry Score Feature" (from `preprocess.py`) was **positive** (approx 0.003), confirming that having more players participate in a game generally increases the win probability.

Synergy Matrix (S) Insight

Analyzing off-diagonal elements of S reveals specific pairwise relationships:

- *Example:* A highly positive value for Si,j(Player i and Player j) indicates a strong, mutual synergy that positively impacts the win probability when they play together.

Confusion Matrix



Model Comparison

Model	Type	Features	Test Accuracy
Logistic Regression	Linear (Baseline)	Player One-Hot + TCI	~ 60.0%
QCM (Static)	Custom Quadratic	Player One-Hot + TCI	~ 62.1%
Dynamic QCM	Custom Quadratic (Time-Weighted)	Player One-Hot + TCI	~ 63.0%
Random Forest	Non-Linear (Benchmark)	Player One-Hot + TCI	~ 64.7%

Key Findings

- **Chemistry Works:** Both custom Quadratic Chemistry Models significantly outperformed the linear baseline.
- **Dynamic Improvement:** The Dynamic QCM, which weights older games less heavily, showed a measurable improvement over the static QCM, suggesting **chemistry effects change over time**.
- **Non-Linear Power:** The Random Forest Classifier achieved the highest accuracy, confirming the complex, non-linear nature of player interactions.

VISUALIZATION INSIGHTS

Top 10 High Synergy Pairs (S Matrix Values)

Top Synergy (Positive S-Score)

	Player 1	Player 2	Synergy Score (S)
157531	Wendell Carter Jr.	Zach LaVine	0.004415
84355	Gary Payton	PJ Dozier	0.004351
70506	Devin Booker	Dyson Daniels	0.004322
152787	Robert Williams	Ziaire Williams	0.004290
22550	Ben Sheppard	Nick Smith Jr.	0.004269
1558	AJ Johnson	Paul George	0.004239
62506	Davion Mitchell	Jaylen Clark	0.004131
125675	Justin Minaya	Scoot Henderson	0.003997
11419	Alperen Şengün	Nick Smith	0.003986
57213	Damion Lee	Jaylon Tyson	0.003985

Top Rivalry (Negative S-Score)

	Player 1	Player 2	Synergy Score (S)
87329	Grant Williams	MarJon Beauchamp	-0.004975
88246	Guerschon Yabusele	Wendell Carter	-0.004912
65737	Dean Wade	Royce O'Neale	-0.004459
19387	Armel Traoré	Matas Buzelis	-0.004245
49446	Cole Swider	Jared McCain	-0.004069
68145	Dereck Lively	Kel'el Ware	-0.004032
77703	Duop Reath	Orlando Robinson	-0.003908
89875	Herbert Jones	Kenrich Williams	-0.003829
90345	Hunter Tyson	Ousmane Dieng	-0.003811
151226	Quenton Jackson	Scottie Barnes	-0.003752

Dataset and Feature Engineering

Dataset

- **Source:** NBA Game Logs (`nba_games_24_25.csv`).
- **Scope:** NBA 2024-2025 Regular Season Games.
- **Target (y):** Binary classification (Home Team Win = 1 / Loss = 0).

Features (X)

The feature vector x for a single team in a game has two components:

1. **Player One-Hot Encoding:** A binary vector where each dimension represents a unique player in the league. 1 if the player was in the lineup for that game, 0 otherwise.
2. **Team Chemistry Index (TCI) Feature:** An additional dimension (used in `preprocess.py`) representing the count of unique players in the team's lineup for that game. This acts as a simple, explicit feature for team cohesion/depth.