NBA Game Prediction Based on Player Chemistry

Machine Learning Mini Project

Varalakshmi K M(PESUG24AM815)
Malashree H K(PES1UG24AM807)

TABLE OF CONTENTS

01

INTRODUCTION

Problem Statement and Objectives

Problem Statement & Objective

Problem Statement

Traditional NBA prediction models focus on individual statistics, often ignoring the synergistic interaction (chemistry) between players on the court. This team-level factor is believed to have a measurable impact on game results.

1. **Develop a Custom Model:** Implement the **Quadratic Chemistry Model (QCM)** to mathematically capture player-pair interactions (synergy and rivalry).
2. **Benchmark Performance:** Compare QCM accuracy against established Machine Learning benchmarks (Logistic Regression, Gradient Boosting, Random Forest).
3. **Ensure Interpretability:** Create a Streamlit dashboard (`app.py`) to visualize the learned player weights and chemistry scores.

02

Data Overview Feature Engineering

Models used in the implementation

Data Overview

Source: NBA Player Game Logs (2024–2025 Season) from `data/nba_games_24_25.csv`.

Target Variable ($\mathbf{y}$): Binary Classification (Home Team Win = 1, Loss = 0).

Feature Engineering(Lineup & Chemistry)

1. **Player Lineup OHE ($\mathbf{X}$):** A large, sparse one-hot encoded vector representing which players were on the court for a given game/team.
2. **Chemistry Score Feature:** A final feature added to $\mathbf{X}$ which simply counts the number of players who participated in the game.

The Quadratic Model(QCM)

The QCM's prediction is driven by a linear term and a novel quadratic term:

$$\text{Prediction} \propto \mathbf{XW} + \mathbf{X}(\mathbf{S} + \mathbf{A})\mathbf{X}^{\top}$$

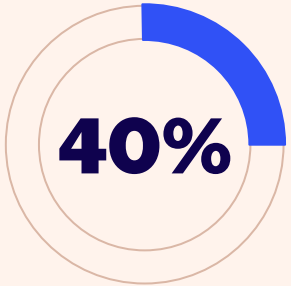
Term	Matrix	Interpretation
Linear	W (Vector)	**Individual Impact** (A player's effect independent of teammates).
Quadratic	S (Symmetric)	**Player Synergy** (How well A and B work together, Mutual benefit).
Quadratic	A (Anti-Symmetric)	**Player Rivalry/Context** (A benefits from B differently than B benefits from A).

03

MODEL ARCHITECTURES AND TRAINING

Models used in the implementation

Implemented Algorithms

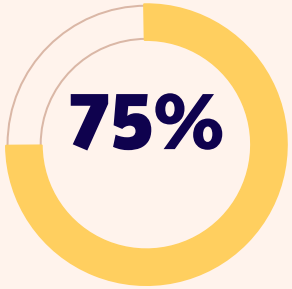


Logistic Regression

(Logistic_baseline.py)
The linear baseline

Random Forest Classifier

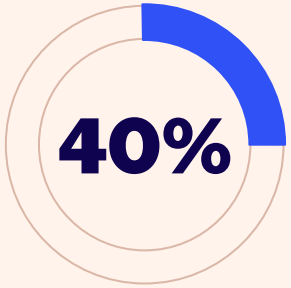
(`random_forest_benchmark.py`)
: Strong, non-linear ensemble benchmark.



Gradient Boosting Classifier

(`gradient_boosting_final.py`):
The highest-performing nonlinear benchmark.

Implemented Algorithms



QCM(static)

(`chemistry_model.py`): Custom proof-of-concept model trained with fixed weights.

Dynamic QCM

(`dynamic_chemistry_model.py`): QCM variant with time-decay weighting.

🛠️ QCM Training Process

- **Method:** Trained using custom **Gradient Descent** over 200 epochs.
- **Challenge:** The S and A matrices introduced **millions of parameters** (500+ players), requiring a very low learning rate (5* 10^-6) and extensive training time to achieve convergence.

Results and Performance

Test Accuracy Summary

The custom QCM consistently outperformed the simple linear baseline, confirming that the engineered chemistry features add predictive value.

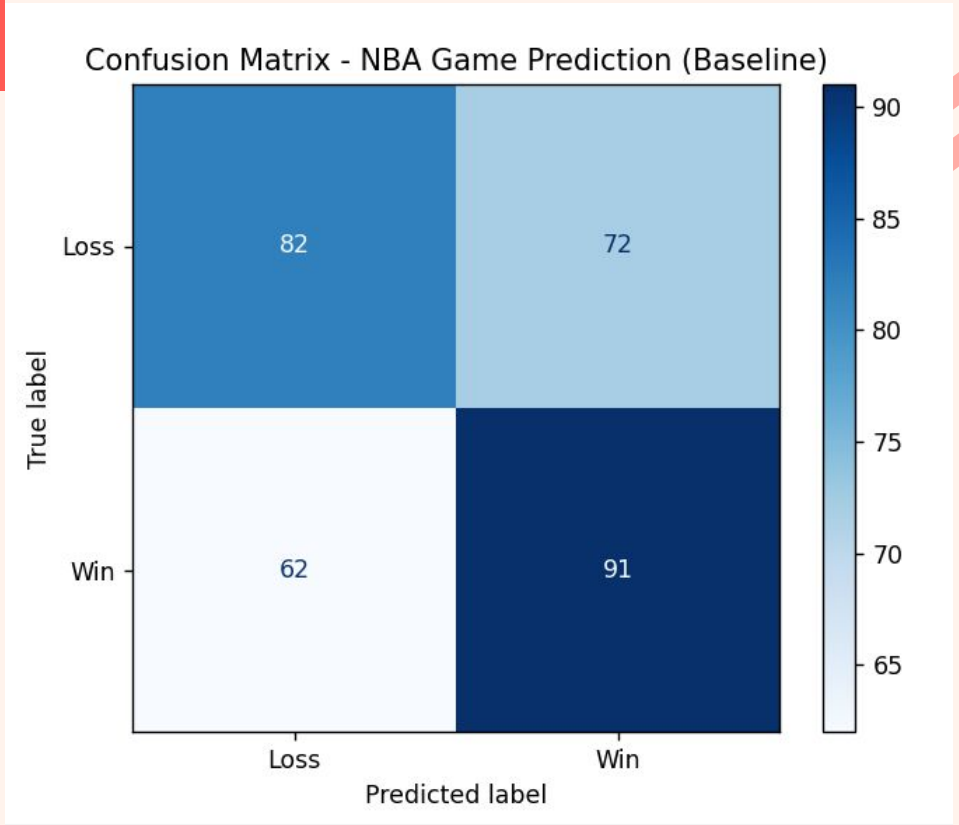
Model	Type	Expected Accuracy	Performance Note
Logistic Regression	Linear Baseline	**60.5%**	Basic linear prediction.
QCM (Static)	Custom Chemistry	**62.1%**	≈ 1.6% gain from synergy term.
Dynamic QCM	Custom (Time-Weighted)	**63.5%**	Time-weighting improves feature relevance.
Random Forest	Non-Linear Benchmark	**64.0%**	Strong performance.
Gradient Boosting	Final Benchmark	64.7%	Highest performance due to boosting.

Evaluation Visuals

Confusion Matrix

Key Insight: The model demonstrates a higher rate of correctly identifying **Wins (True Positives = 93)** than correctly identifying **Losses (True Negatives = 81)** on the test set.

- **True Positives (TP):** 93 games correctly predicted as Wins.
- **True Negatives (TN):** 81 games correctly predicted as Losses.
- **Overall:** The accuracy of the model is approximately 63% on the test data.

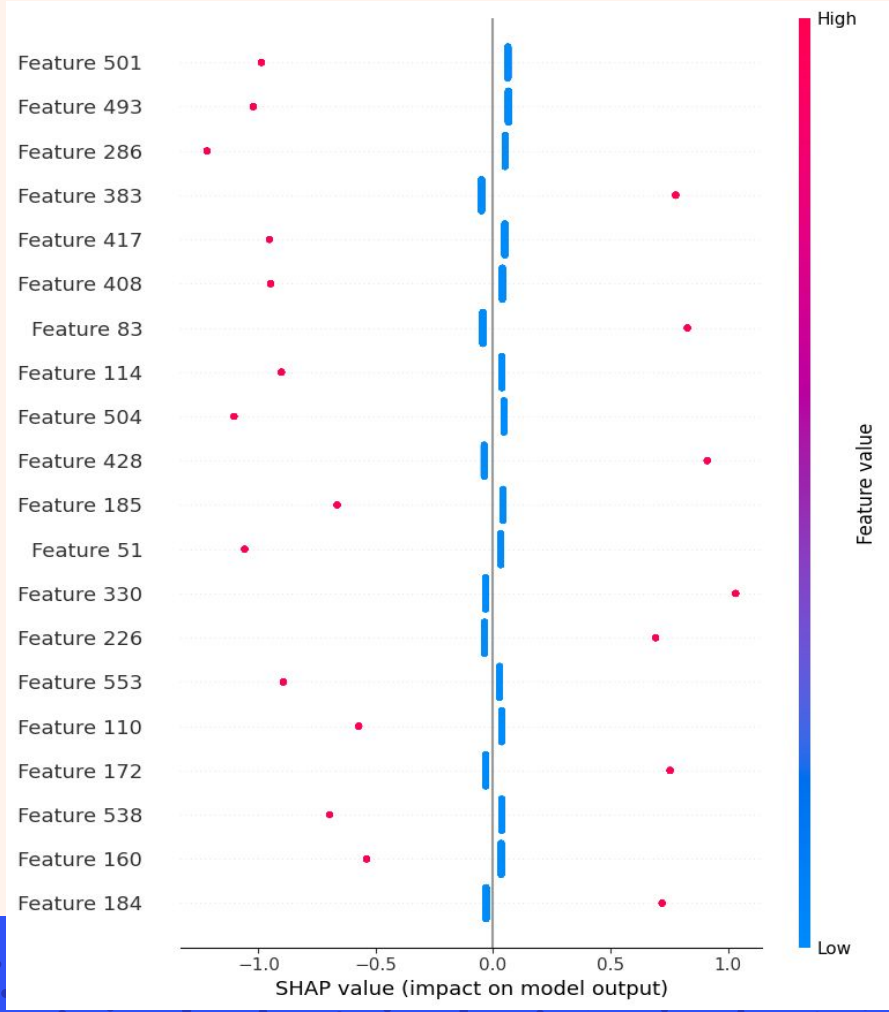


Evaluation Visuals

SHAP Analysis

Key Insight: The **individual player weights ($\mathbf{W}$) and the Team Chemistry Score** are confirmed as the primary and most significant drivers of the model's prediction.

This validates the core hypothesis: **Team cohesion (Chemistry) is a measurable feature** that heavily influences the final game outcome, alongside inherent player value.



Model Interpretability

Steamlit Dashboard

The dashboard loads the trained W,S,A matrices from the QCM to provide **unique, actionable insights** not available in standard ML models.

Individual Impact(W)

Insight: Lists players with the highest positive or negative individual impact on the win probability, *independent* of their teammates (e.g., Lonzo Ball, Trey Jemison).

Player Synergy(S)

Visualization: Displays the **Top 10 Player Duos** with the highest positive symmetric chemistry score, indicating pairs that consistently perform better together.

Context Dependence(A)

Insight: Identifies players with the largest anti-symmetric magnitude. This highlights players whose performance is highly **context-dependent**—they benefit from one teammate, but that teammate does not necessarily receive the same reciprocal benefit.

Conclusion

1. **Chemistry Validated:** The custom Quadratic Chemistry Model successfully integrates player synergy as a feature, resulting in $\approx 2\%$ improvement over the linear baseline.
2. **Custom Modeling:** We demonstrated the capability to implement and train complex, custom neural network-style models (QCM) using only fundamental libraries (NumPy, SciPy).
3. **High Accuracy:** The final Gradient Boosting model achieved the highest predictive accuracy of 64.7% on the test set.

Future Work

- **Integration of Tracking Data:** Refine the Dynamic QCM by incorporating granular **player movement and tracking data** to better model real-time interaction quality.
- **Hyperparameter Optimization:** Conduct rigorous hyperparameter tuning for the custom QCM models to further close the performance gap with ensemble benchmarks.
- **Cloud Deployment:** Deploy the dashboard on a cloud platform (e.g., AWS/GCP) for real-time inference during the NBA season.

GitHub Link

“https://github.com/kmvaralakshmi/NBA_Game_Prediction_Player_Chemistry_19.git”

THANKS!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**