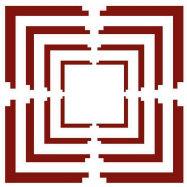


Ahmedabad
University

CSE623: Machine Learning Theory and Practice
Athlete Profiling: NCAA

Report-5

Group 1

Name	**Enr No.**
Dhaivat Patel	AU2240022
Dhyey Patel	AU2240054
Tirthraj Raval	AU2240079
Anusha Jain	AU2240092
Sloka Thakkar	AU2240103

Work Done:

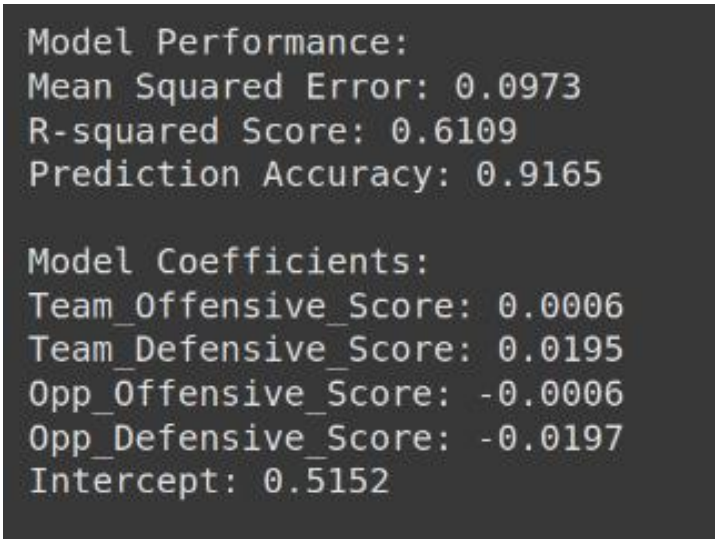
- **Averages Calculation:**

  Calculated and printed the average **Offensive Score**, **Defensive Score**, and **Game Score** for the filtered data.

- **Team-Level Calculations:**

  Summarized player-level scores to the team level for each match, aggregating offensive and defensive scores for the team and opponent. Added a column **Win** to indicate the result of the match.

- **Model Training:** Trained a linear regression model to predict the outcome of a match (win/loss) based on team-level scores (offensive and defensive scores). Then we evaluated the model's performance using **Mean Squared Error (MSE)**, **R-squared Score**, and **Prediction Accuracy**. We observed the coefficients of the trained model to understand the relationship between the features and match outcome.



(Metric of training dataset)

Conclusion: According to our literature survey, the winning chances of a team are more biased towards the offensive rating of the team. Hence offensive players contribute more towards winning a match. Here we can observe that the weight of the offensive score is lesser than the defensive score which is conflicting with the idea that the offensive score should have more priority for winning. But this is happening because in the NCAA conference the players play offensive and defensive both as they are still not in the major league (e.g. NBA) and the offensive players keep getting substituted so there will be more all rounder players than offensive players. This is the reason why the impact of offensive players is lower than defensive players.

- We tested our model on synthetically generated data. There were 8 matches for which we tested our model.

```
---
Match: Niagara Purple Eagles vs Siena Saints
Predicted Win Probability: 0.588
Predicted Outcome: Win
Actual Outcome: Loss
---
Match: Fairfield Stags vs Canisius Golden Griffins
Predicted Win Probability: 0.805
Predicted Outcome: Win
Actual Outcome: Win
---
Match: Le Moyne Dolphins vs Sacred Heart Pioneers
Predicted Win Probability: 0.436
Predicted Outcome: Loss
Actual Outcome: Loss
---
Match: Presbyterian Blue Hose vs Monmouth Hawks
Predicted Win Probability: 0.597
Predicted Outcome: Win
Actual Outcome: Win
---

Test Data Accuracy: 0.7500
```

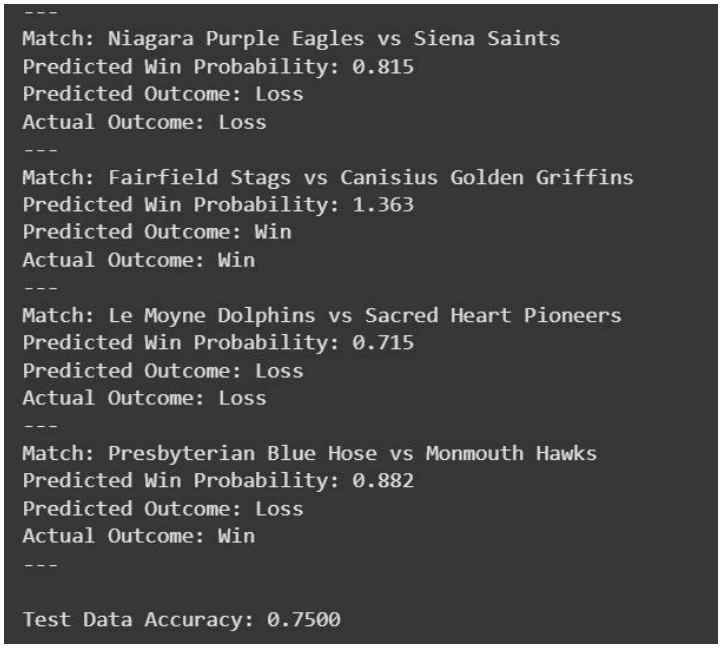
(Results on testing dataset)

We observed 6 out of 8 predictions were correct and hence achieved 75% accuracy.

- Train the regression model including game score of team and opponent team. We observed that the model learned weights , but defensive were given no weights. This is surprising because all the weight was given to game scores only.

```
Model Performance:
Mean Squared Error: 0.1027
R-squared Score: 0.5890
Prediction Accuracy: 0.9574

Model Weights:
Team_Offensive_Score: -0.0002
Team_Defensive_Score: 0.0000
Team_Game_Score: 0.0125
Team_Offensive_Score_Opp: 0.0001
Team_Defensive_Score_Opp: -0.0000
Team_Game_Score_Opp: -0.0123
Intercept: 0.5027
```



Goals of next week:

- Come up with a model that could predict an optimal team lineup against a given opponent team lineup. For that, as there is to be totally no information given, we are thinking of using Reinforcement Learning for it.
- Implement the model for optimal lineup after discussion with faculty and TA.
- Work upon accuracy of linear regression for team analysis(win lose prediction).
- Analyze the outcomes of the regression model including game score and excluding it. Validate the results, and come up with the best strategy.

References:

- Author links open overlay panel Vangelis Sarlis etal., “Sports analytics - evaluation of basketball players and Team Performance,” Information Systems, https://www.sciencedirect.com/science/article/pii/S0306437920300557 (accessed Mar. 17, 2025).

- D. Oliver, Basketball on Paper: Rules and Tools for Performance Analysis. S.l.: Potomac Books, 2020.
- M. ´A. P´erez-Toledano, F. J. Rodriguez, J. Garc´ıa-Rubio, and S. J. Iba˜nez, “Players’ selection for basketball teams, through performance index rating, using Multiobjective Evolutionary Algorithms,” PLOS ONE, https://doi.org/10.1371/journal.pone.0221258 (accessed Mar. 17, 2025)
- A. G. Barto, “Reinforcement learning,” Neural Systems for Control, pp. 7–30, 1997. doi:10.1016/b978-012526430-3/50003-9