Ahmedabad
University

CSE623: Machine Learning Theory and Practice

Report-6

Group 1

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

Work Done

This week, we focused on evaluating different machine learning models using various combinations of input features to predict the outcome of basketball games. We used three main types of input variables: offensive features, defensive features, and game score. These were tested with different machine learning models, including Linear Regression, Logistic Regression, Random Forest, and XGBoost, to assess which combination yields the best performance. The summary of insights derived is listed below.

- Combining All Available Features (Offensive, Defensive, and Game Score):
    - When we combined all three features and used Linear Regression, the model achieved a high accuracy of 0.9272. This indicates that combining all the relevant features gives us strong predictive performance.

```
Model Performance:
Mean Squared Error: 0.0966
R-squared Score: 0.6137
Prediction Accuracy: 0.9272

Model Weights:
Team_Offensive_Score: -0.0021
Team_Defensive_Score: 0.0004
Team_Game_Score: 0.1202
Team_Offensive_Score_Opp: 0.0019
Team_Defensive_Score_Opp: 0.0001
Team_Game_Score_Opp: -0.1186
Intercept: 0.4478
```

    - Additionally, we also tested the same combination with Logistic Regression, Random Forest, and XGBoost. Logistic regression accuracy 0.9272. Random Forest accuracy 0.9236 and XGBoost accuracy 0.9112.

```
Logistic Regression Accuracy: 0.9272
Random Forest Accuracy: 0.9236
XGBoost Accuracy: 0.9112
```

- Using Only Offensive and Defensive Features
  - Applying Linear Regression to just the offensive and defensive metrics resulted in a slightly lower accuracy of 0.7744. This indicated that while these features are informative, the game score adds valuable context.

```
Model Performance:
Mean Squared Error: 0.1752
R-squared Score: 0.2992
Prediction Accuracy: 0.7744

Model Weights:
Team_Offensive_Score: 0.0045
Team_Defensive_Score: -0.0129
Team_Offensive_Score_Opp: -0.0046
Team_Defensive_Score_Opp: 0.0129
Intercept: 0.5079
```

- Using only Game Score as feature
  - Using game score as the only feature resulted in the accuracy of 0.9272 indicating that the game score as a feature adds valuable strength to overall model prediction.
  - When linear regression is applied on game score

```
Model Performance:
Mean Squared Error: 0.0973
R-squared Score: 0.6109
Prediction Accuracy: 0.9272

Model Weights:
Team_Game_Score: 0.1091
Team_Game_Score_Opp: -0.1079
Intercept: 0.4955
```

  - Logistic regression, Random Forest and XGBoost Applied on game score

```
Logistic Regression Accuracy: 0.9272
Random Forest Accuracy: 0.8988
XGBoost Accuracy: 0.8988
```

- Using the parameters available in the dataset like PTS, MIN, FGM, FGA, PM, etc.
    - When we applied logistic regression, Random forest and XGBoost on parameters available then we attained the accuracies given below.

```
Logistic Regression Results:
Training Accuracy: 0.8729
Testing Accuracy: 0.8703
Confusion Matrix (Testing):
[[250  35]
 [ 38 240]]

Random Forest Results:
Training Accuracy: 1.0000
Testing Accuracy: 0.8135
Confusion Matrix (Testing):
[[236  49]
 [ 56 222]]

XGBoost Results:
Training Accuracy: 0.9924
Testing Accuracy: 0.8419
Confusion Matrix (Testing):
[[247  38]
 [ 51 227]]
```

- After reviewing different Reinforcement learning, we have started implementing **Proximal Policy Optimization (PPO)** to find the optimal team lineup. The system incorporates player statistics from past matches involving both teams that might have different team sizes. The team inputs with different sizes require processing using **mean pooling techniques** or **Set Transformer** to convert them into consistent fixed-length embeddings. From match to match the agent picks five players from an available total of N different players through combinations. Historical win conditions or individual player metrics will serve as the basis for selecting players for the team line up.

Key Findings

- The most consistent and high-performing approach involves using all three feature types together.

- Game score is a strong standalone predictor, but performs best when combined with offensive and defensive metrics.

Goals for Next Week

- Implement the Reinforcement learning by PPO to find the optimal team lineup.
- Further analyze and optimize the results received after applying reinforcement learning.