***Training Strategy Document:***

**Introduction:**

Predicting the number of goals a player is going to score in an upcoming match is a useful insight for the scouting department of the team. Also, knowing the goals that will be scored in an upcoming match by a player will help setting up the starting XI and prepare the tactics of the match. Moreover, it can help manage the team’s attacking players and help the scouting team make decisions on potential transfers. Lastly, it will assist fans bet on players scoring in upcoming matches

**1.Problem:**

Prediction of the number of goals an attacking player (RW, ST, LW) is going to score in an upcoming match of the Premier League based on his statistics in previous matches and the statistics of the opposing team.

**2.Dataset:**<https://www.kaggle.com/datasets/davidcariboo/player-scores/data>

We will include minutes played, shots per game, shots on target per game, expected goals per match, and position for a player from the Premier League and statistics of the opposing team like, goalkeeper saves, goals conceded, goals prevented.

We will handle missing values, normalize features, and address potential outliers.

**3.Steps:**

* Find data regarding the minutes played per match, shots per game, shots on target per game, expected goals per match, and position for a player from the Premier League that we input at the start.
* Import pandas to transform raw data from the website into group data into tables for each different team in the Premier League.
* Import scikit-learn.
* Import the dataset into our code
* We have found the average of each player, by using groupby.mean(), for the statistics regarding each player.
* Then we clean the data, handle missing values and outliers.
* We visualise the data by plotting bar charts for each numerical attribute in our dataframe.
* Find a correlation between this data to predict the goals that this player will score in the upcoming seasons by using simple regression models like linear regression or decision trees.
* Use evaluation matrix like mean absolute error (MAE) to access our models accuracy
* Output the result