Project Structure

A predictive analytics framework for forecasting soccer match outcomes using machine learning models

- Prediction_Paper.ipynb: A replication of the original paper's methodology. It establishes a baseline by using standard match statistics and weather data to predict results like in the original paper.
- Machine_Learning_Project.ipynb: The "Improved Version" of the paper. It introduces sophisticated features such as:
    - Team Momentum: Calculated using an Exponentially Weighted Moving Average (EWMA) of recent points.
    - Travel Distance: Calculated via the Haversine formula to estimate the distance away teams traveled to the stadium.
    - Full-time Weather: Expanding the use of weather data to better capture its impact on match outcomes
- Rolling_window.ipynb: An alternative approach that utilizes a rolling window to generate time-sensitive features, ensuring the model only uses data available prior to the kickoff.
- CatBoost.ipynb: Dedicated to the hyperparameter tuning and evaluation of the CatBoost classifier, one of the top-performing models in this project.
- 90_kickoff_API.ipynb: The initial data processing script responsible for loading raw CSV datasets and preparing the master dataframe.