Soccer League Prediction and Simulation Project

Project Overview

This project simulates a full soccer league season using machine learning models to predict match outcomes and goals. It generates accurate league standings based on simulated matches between all participating teams.

How It Works

1. Data Collection:
   - Historical data from reliable sources:
     • Transfermarkt: https://www.transfermarkt.com/
     • FBRef (2016-2017 Season): https://fbref.com/en/comps/12/2016-2017/
     • Kaggle Dataset (1995-2020): https://www.kaggle.com/datasets/kishan305/la-liga-results-1995
     • FBRef (La Liga Seasons History): https://fbref.com/en/comps/12/history/
2. Data Preprocessing:
   - Data is cleaned and merged to create comprehensive datasets.
3. Model Training:
   - Models used: Random Forest Classifier and Regressors
   - Predictions include match outcomes and goal counts.
4. League Simulation:
   - Simulates matches and generates league standings.
5. Final Output:
   - Saves the final league standings as a CSV file.

Benefits and Applications

- Predictive Analysis: Analyze potential outcomes of soccer seasons.
- Performance Tracking: Track team performance across seasons.
- Data Insights: Leverage data-driven insights for better sports analytics.

Project Structure

The project consists of four primary scripts:

1. train_model.py - Trains machine learning models.
2. predict.py - Predicts match outcomes and goals.
3. simulate_match.py - Simulates individual matches.
4. simulate_league.py - Simulates the entire league season.

Conclusion

This project combines machine learning and soccer analytics to simulate a season of La Liga matches. The final league table offers a realistic simulation of potential outcomes based on historical data and predictive modeling.

Generated on: 2025-03-24 23:52:56