Football score prediction

Daniel-Cristian-Marian Țăpuși
Marian-Claudiu Neagu

Motivation

Why Football Predictions?

Help teams make better decisions and give fans a fun way to connect with the game, making matches even more exciting

Project highlights

- Web scraping
- Cleaning and Exploratory data analysis
- Predicting using Deep Learning Model
- Application

Project Plan

1. Build the initial dataset
2. Train language models
3. Build reward dataset (using human feedback / metrics averaging)
4. Train reward model
5. Further fine-tune the initial language model, using Reinforcement Learning

Web Scraping – Extraction

- Use libraries: requests, BeautifulSoup, JSON
- Create dynamic URLs (Leagues & Seasons)
- Find <script> tag with "teamsData"
- Clean & parse JSON string

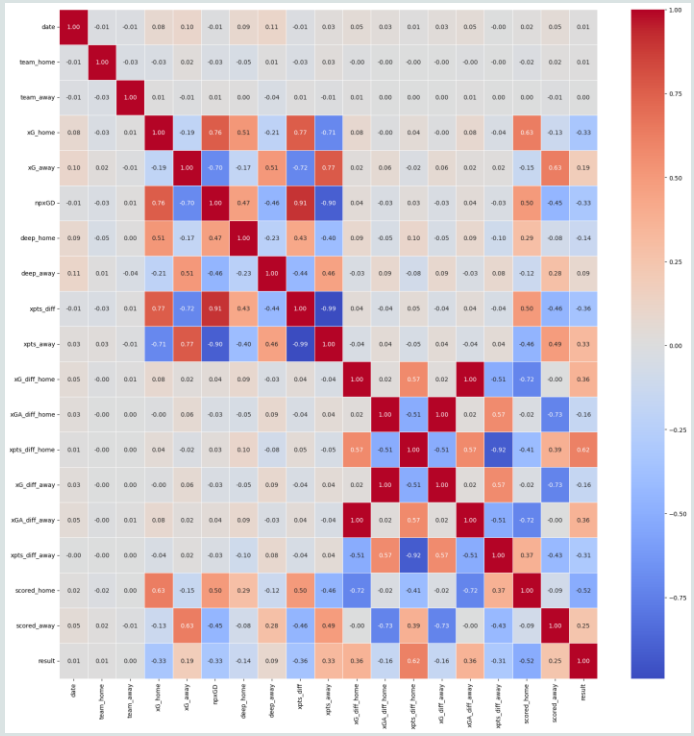
Data Processing and Aggregation

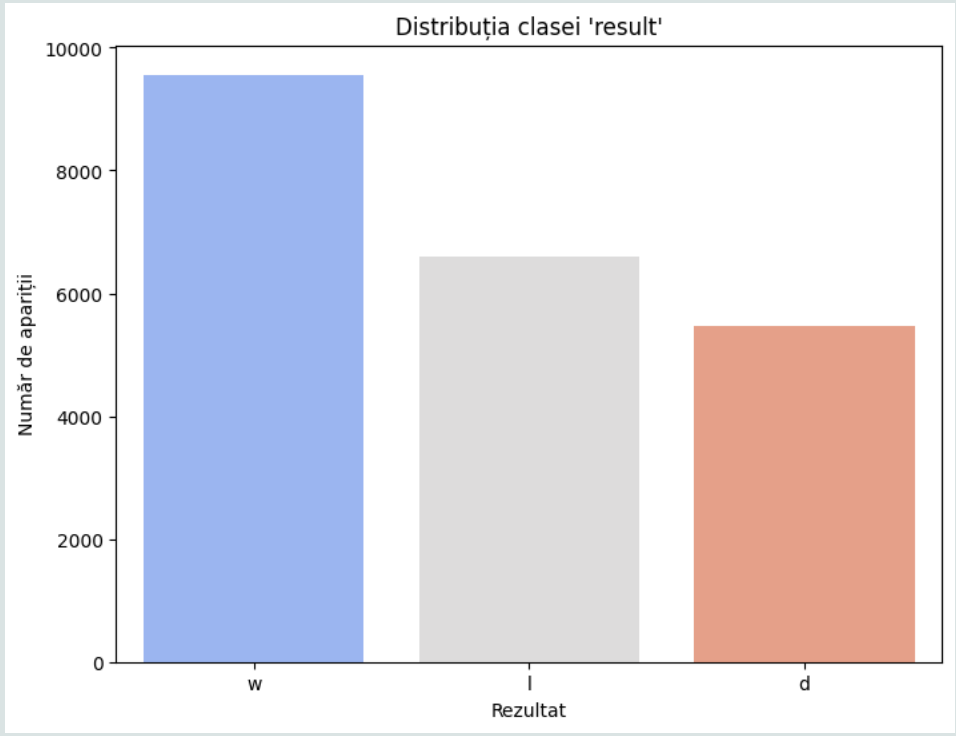
- Convert JSON → Python dictionary → DataFrames
- Compute PPDA metrics (ratio & raw values)
- Merge team, season, and league data
- Export final DataFrame to CSV

Data Cleaning and Exploratory Data Analysis (EDA)

- Transformed raw data into a structured CSV by handling missing values and duplicates
- Visualized key distributions and correlations
- Uncovered insights into team performance and match patterns
- Produced a Confusion Matrix
- Prepared a clean dataset ready for modeling

EDA - plots




Distribuția clasei 'result'
Număr de apariții
10000
8000
6000
4000
2000
0
w
l
d
Rezultat

Model Development - Data Preparation and Feature Engineering

- Data Loading & Preprocessing:
  - Load CSV, convert dates, & compute time weights
  - Create rolling averages (xG, scored, deep, xpts, xG_diff, xGA_diff)
- Team Ratings:
  - Compute offensive & defensive ratings via linear programming (Pulp)
  - Incorporate home advantage

Model Architecture and Custom Loss

- Deep Neural Network:
  - 6 fully connected layers with BatchNorm & Dropout
- Dual output heads:
  - Goals head (predicts home/away goals)
  - Result head with tanh (match outcome)
- Custom Loss Function:
  - SoccerLoss combining weighted MSE for goals & result

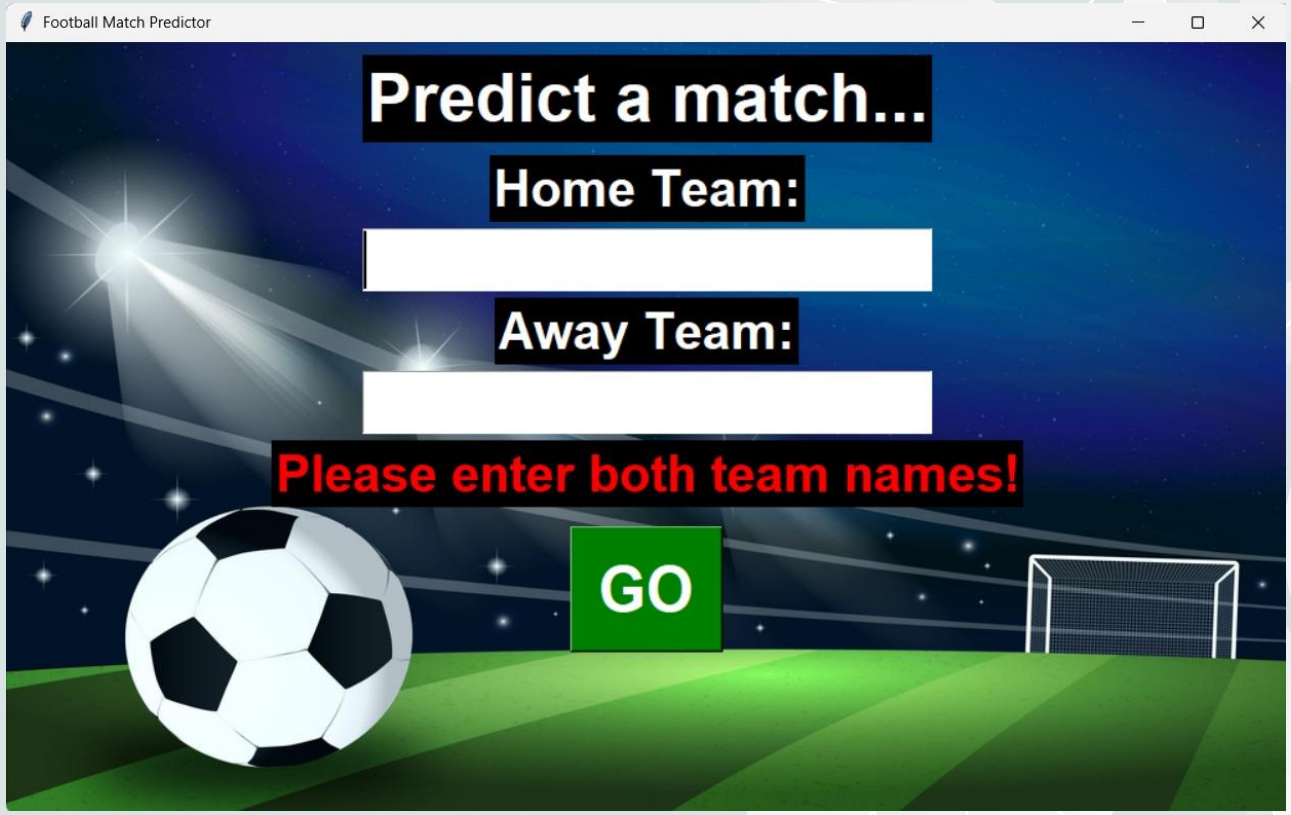
Training, Prediction and Evaluation

- Training Pipeline:
  - Data split, normalization, & construction of custom Dataset/DataLoader
  - Optimizer: AdamW; Scheduler: ReduceLROnPlateau
  - Monitor individual losses & save best model
- Prediction & Evaluation:
  - Generate predictions, compute Poisson-based outcome probabilities
  - Assess performance with sample match accuracy and team ratings

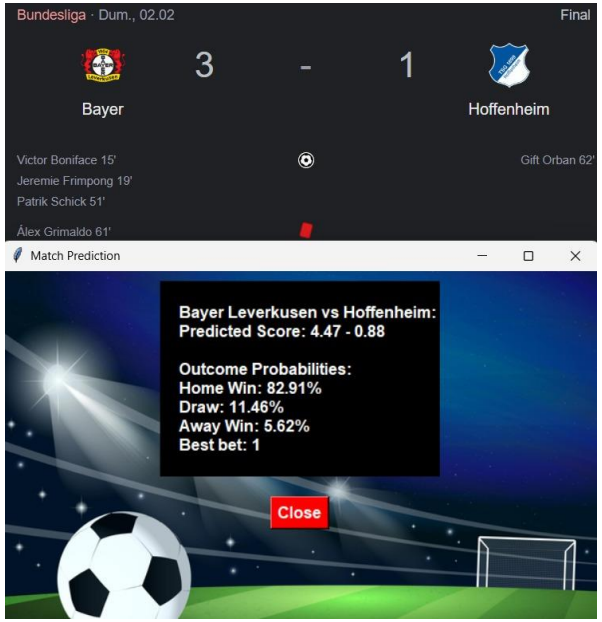
Application and Integration

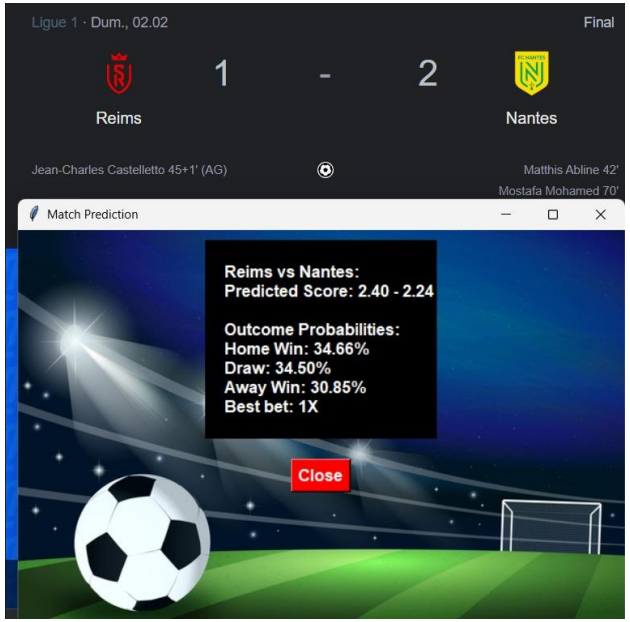
- Tkinter UI collects team names and displays predictions in a popup
- Integrates preprocessed data, team ratings (JSON), and a trained PyTorch model to compute scores, outcome probabilities, and best bet

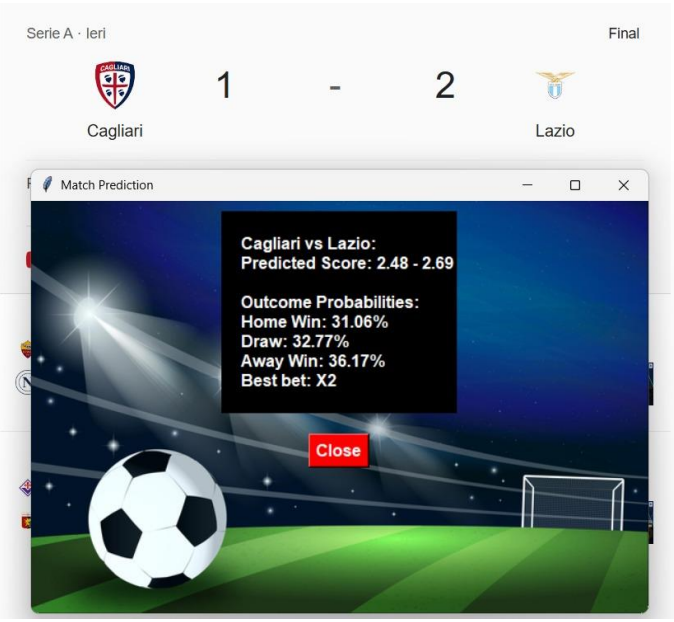
Final application



Examples


Bundesliga · Dum., 02.02
Final
3 - 1
Bayer
Hoffenheim
Victor Boniface 15'
Jeremie Frimpong 19'
Patrik Schick 51'
Álex Grimaldo 61'
Gift Orban 62'
Match Prediction
Bayer Leverkusen vs Hoffenheim:
Predicted Score: 4.47 - 0.88
Outcome Probabilities:
Home Win: 82.91%
Draw: 11.46%
Away Win: 5.62%
Best bet: 1
Close


Ligue 1 · Dum., 02.02
Final
1 - 2
Reims
Nantes
Jean-Charles Castelletto 45+1' (AG)
Matthis Abline 42'
Mostafa Mohamed 70'
Match Prediction
Reims vs Nantes:
Predicted Score: 2.40 - 2.24
Outcome Probabilities:
Home Win: 34.66%
Draw: 34.50%
Away Win: 30.85%
Best bet: 1X
Close


Serie A · Ieri
Final
1 - 2
Cagliari
Lazio
Match Prediction
Cagliari vs Lazio:
Predicted Score: 2.48 - 2.69
Outcome Probabilities:
Home Win: 31.06%
Draw: 32.77%
Away Win: 36.17%
Best bet: X2
Close

Thank you!