EPL Match Predictor

Using Machine Learning to predict Premier League matches

Noth1ng Real
@noth1ng_real on X

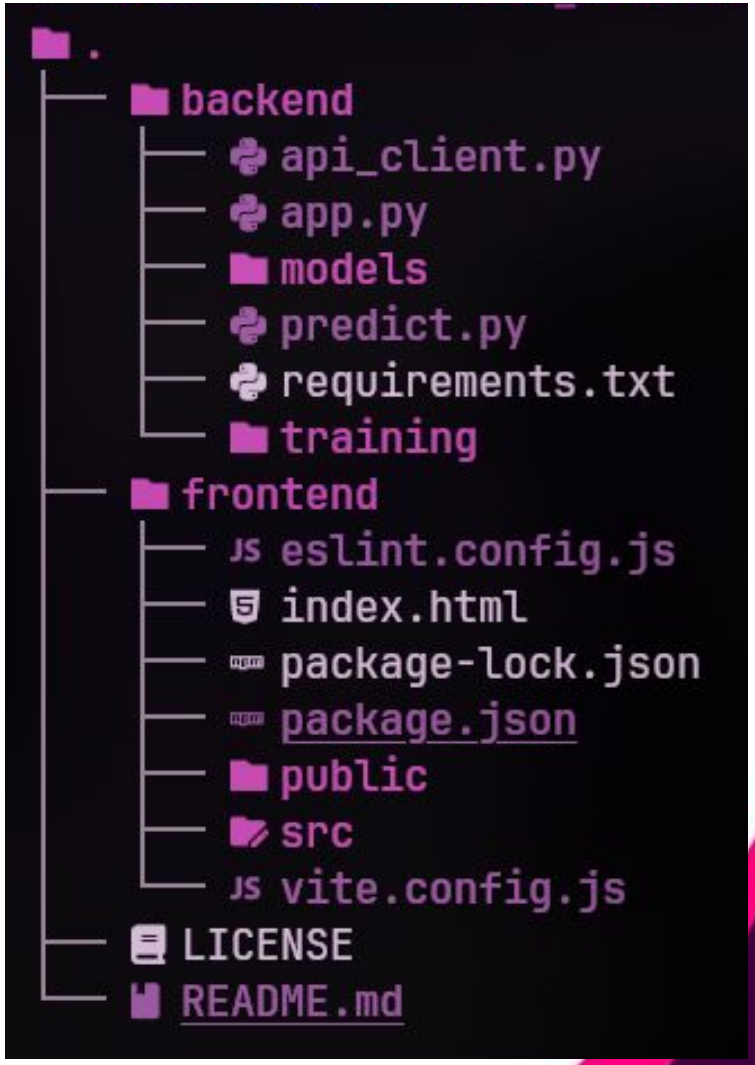
Project Overview

- A full-stack web application that predicts English Premier League (EPL) football match outcomes using machine learning

- Key Components:
  - Machine Learning model trained on historical match data
  - Real-time data integration via football-data.org API
  - Flask backend serving predictions
  - React frontend for user interaction

- Prediction Categories:
  - HOME WIN
  - DRAW
  - AWAY WIN

- Model Accuracy: ~56-57% (significantly better than random guessing at 33%)



Premier League

Problem Statement

- **Question**: Can we predict the outcome of a football match before it happens?

- Why is this hard?
  - Football is inherently unpredictable ("Any given Sunday")
  - Injuries, weather, referee decisions, player form
  - Psychological factors (rivalry matches, relegation battles)
  - Random events (own goals, red cards)

- Why is this valuable?
  - Sports analytics industry worth $3.4 billion (2023)
  - Used by betting companies, fantasy football, team management
  - Demonstrates practical ML application on real-world data

Premier League

Data Description

- Dataset: Premier League Matches (Kaggle)
  - Time period: 2019 - 2024 (5 seasons)
  - Total of 4788 matches (rows) each with 28 columns
- Key Features used (Total of 8):
  - Home team
  - Away team
  - Goals for (home)
  - Goals against (home)
  - Venue
  - Result
- No Data quality issues
- Key Finding: Home advantage is REAL - teams win ~46% at home vs ~27% away

```
❯ wc -l matches.csv
4789 matches.csv
```

#	date	venue	result	gf	ga	opponent	team
0	2020-09-21	Away	W	3	1	Wolves	Manchester City
2	2020-09-27	Home	L	2	5	Leicester City	Manchester City
4	2020-10-03	Away	D	1	1	Leeds United	Manchester City
5	2020-10-17	Home	W	1	0	Arsenal	Manchester City
7	2020-10-24	Away	D	1	1	West Ham	Manchester City
9	2020-10-31	Away	W	1	0	Sheffield Utd	Manchester City
11	2020-11-08	Home	D	1	1	Liverpool	Manchester City
12	2020-11-21	Away	L	0	2	Tottenham	Manchester City

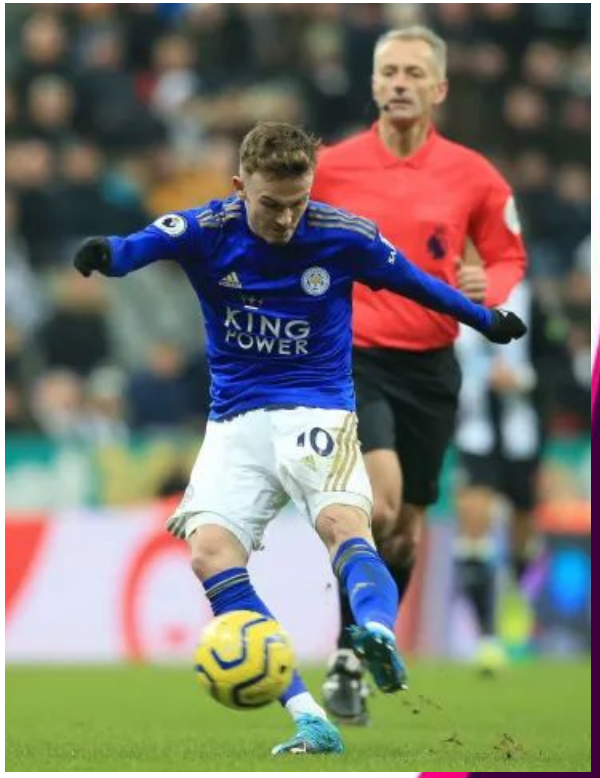
Premier League

Feature Engineering

- **The Challenge**: Raw match results don't directly predict future matches. We need to engineer meaningful features.

Season Performance Metrics (6)

Feature	Description	Why It Matters
Home Goals/Game	Avg goals scored by home team	Offensive strength
Away Goals/Game	Avg goals scored by away team	Offensive strength
Home Conceded/Game	Avg goals conceded by home team	Defensive weakness
Away Conceded/Game	Avg goals conceded by away team	Defensive weakness
Home Win Rate	% of matches won by home team	Overall team quality
Away Win Rate	% of matches won by away team	Overall team quality



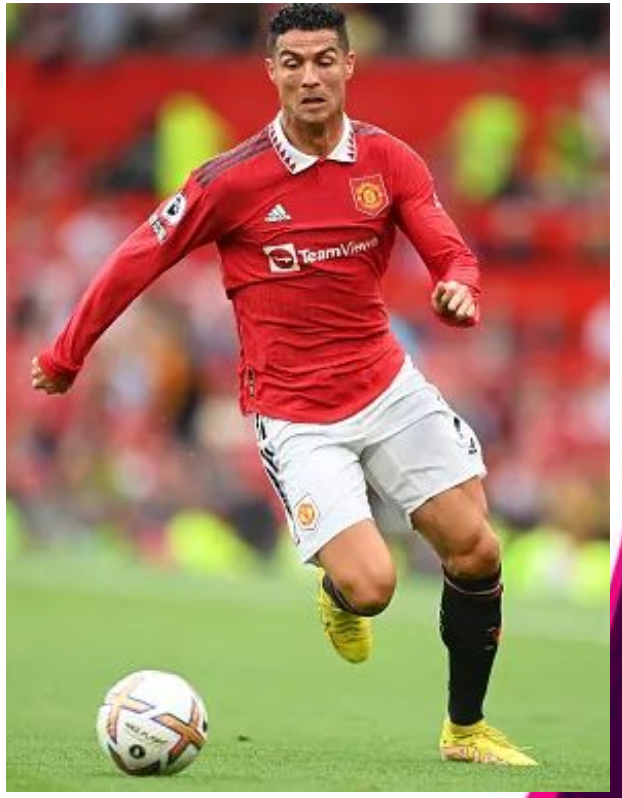
Recent Form(2)

Feature	Description	Why It Matters
Home Recent Form	Points from last 5 games (normalized 0-1)	Current momentum
Away Recent Form	Points from last 5 games (normalized 0-1)	Current momentum

Premier League

Model Selection: Random Forest

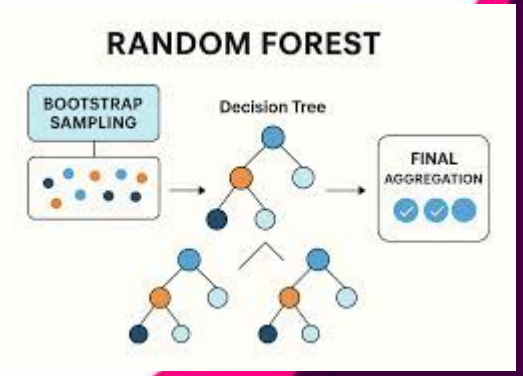
- Random Forest: An ensemble learning method that builds multiple decision trees and combines their predictions



Why Random Forests for this Problem?

- Handles Non-Linear Relationships: Football outcomes aren't linear (2x goals ≠ 2x win probability)
- Resistant to Overfitting: Ensemble of trees reduces variance
- Feature Importance: Can identify which features matter most
- No Feature Scaling Required: Trees don’t care about feature magnitudes
- Probabilistic Output: Gives confidence score, not just predictions

```
RandomForestClassifier(
    n_estimators=100,     # 100 decision trees
    max_depth=10,         # Limit tree depth (prevents overfitting)
    random_state=42       # Reproducibility
)
```



Premier League

Model Evaluation

- Model Accuracy: **56.58%**

Baseline	Accuracy	How
Random Guessing	33.3%	Pick HOME/DRAW/AWAY randomly
Always Predict “HOME”	46.0%	Home teams win ~46% of the time
Our Model	56.6%	Learns team pattern
Professional Analysts	55-60%	Experts at sports betting firms

- Sports is inherently chaotic and unpredictable, matches pro analysts

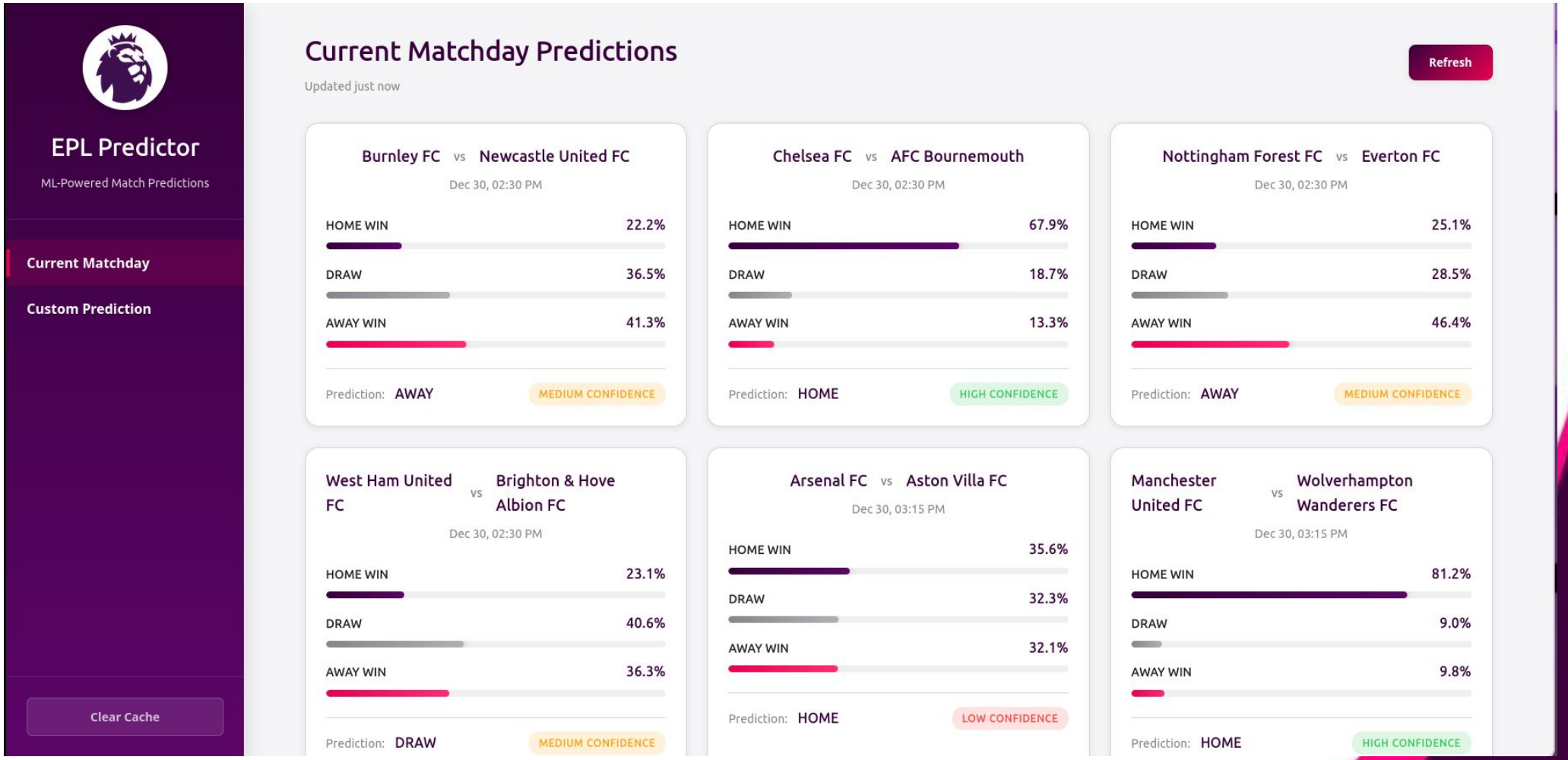
	HOME	DRAW	AWAY
HOME	152	28	45
DRAW	58	32	41
AWAY	49	25	70

- Model learned real patterns (strong on home wins)
- Struggles with draws because they're inherently unpredictable
- Appropriately cautious - doesn't overconfident on close matches
- Honest performance, not overfitting

Premier League

Final Result

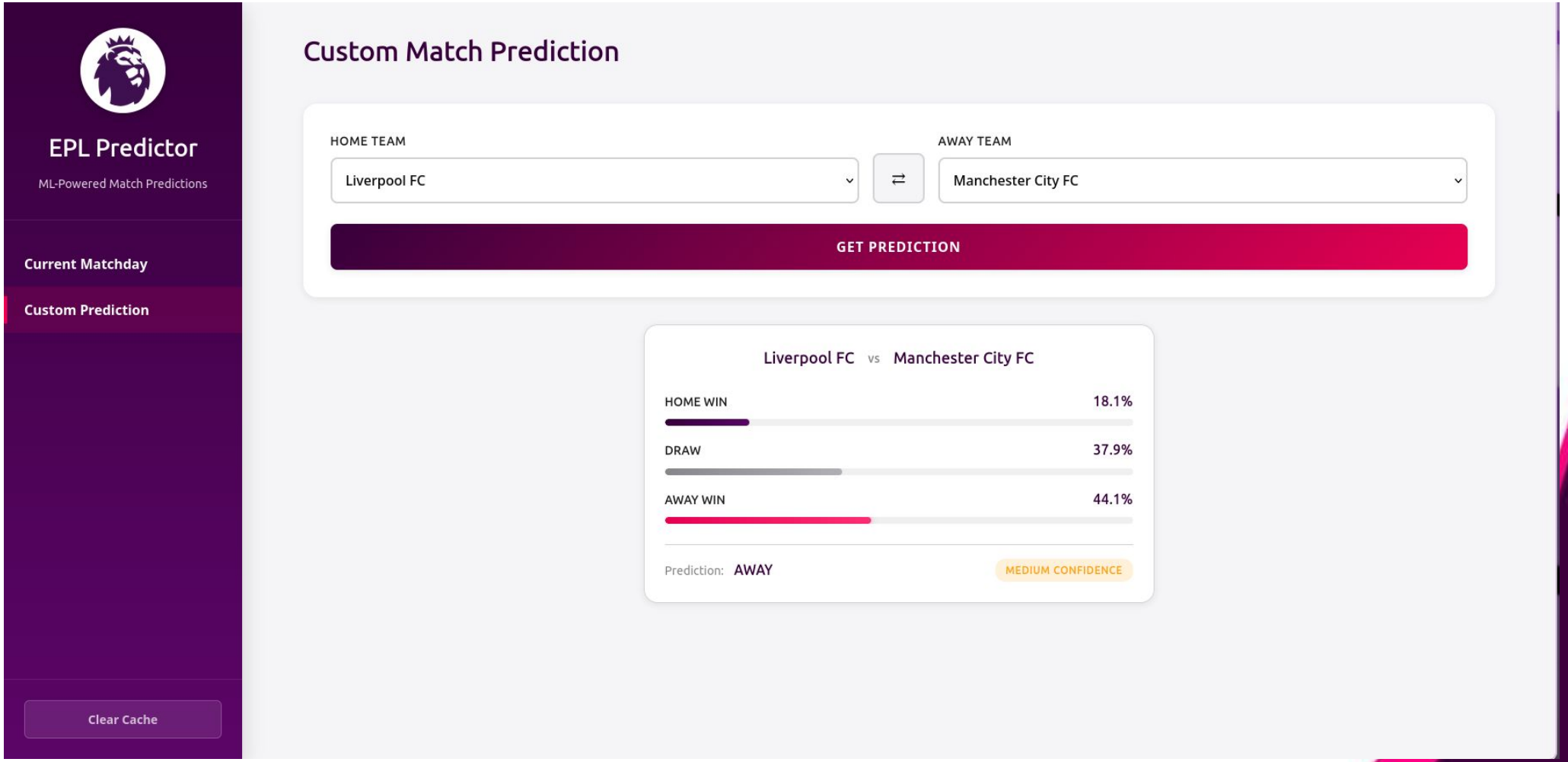
App can predict current matchday results ...



Premier League

Final Result

As well as custom matchups



Premier League