ChatGPT

Datasets for Football Match Outcome Prediction (Win/Draw/Loss)

To build a model for classifying match results (win, draw, or loss), several datasets and data sources are available. Below we outline **freely available** datasets (or sources for creating datasets) from roughly the last 5 years (and beyond, if needed), along with their **reliability**, **update frequency**, **ease of access**, **cost**, and other relevant factors. We also mention similar projects and the data they used. The goal is to potentially combine these sources into a comprehensive dataset for training a prediction model.

1. Kaggle Compiled Datasets

Kaggle hosts user-contributed datasets that aggregate football match data from various sources. These are convenient to download and often free. Two notable ones:

- **Club Football Match Data (2000–2025)** – A **large dataset (~230,000 matches)** spanning 27 leagues worldwide, compiled by a Kaggle user (adamgbor) [1]. This includes match results, basic **match stats** (e.g. goals, cards), and **betting odds**, and even team ratings from Club Elo (a global club Elo rating system) [2] [3].
- *Reliability:* High – data is primarily derived from the well-known Football-Data.co.uk site and ClubElo (both reputable free sources) [2] [3]. Betting odds (mostly Bet365) are included for many matches [4]. The data covers top divisions of major countries and some secondary leagues, so it is fairly comprehensive.
- *Update Frequency:* **One-off compilation** (updated through January 2025) [1]. The dataset isn't updated in real-time on Kaggle, but it captures recent years up to 2024. For future matches, one could manually update it (since the sources are public).
- *Ease of Obtaining:* Easy – downloadable as CSVs via Kaggle or the author's GitHub [5]. No scraping needed; it's ready to use.
- *Cost:* **Free.** (Kaggle account required to download, but no payment.)

- *Notes:* This is one of the **most comprehensive free datasets** available, suitable for machine learning. It spans ~25 years (useful if you later want more than 5 years), and for your focus (last 5 years), you can filter to recent seasons. It contains **match outcome (W/D/L)** plus rich features like odds and possibly Elo ratings, which can help in modeling.

- **European Soccer Database (2008–2016)** – A classic Kaggle dataset created by Hugo Mathien [6], containing detailed data for 25,000+ matches across 11 European countries. It includes results and extensive features: **team lineups, events (goals, fouls, cards)**, player and team attributes (sourced from EA Sports FIFA game stats), and betting odds from up to 10 bookmakers [7].

- *Reliability:* High for its timeframe – this dataset has been widely used in academic and hobby projects. It was built by scraping reliable sources and integrating FIFA game ratings for players, which many projects (e.g. by Kaggle users and researchers) found useful [7].
- *Update Frequency:* **Static/archival** – covers 2008–2016 and is **not updated beyond 2016** [7]. It's great for historical analysis and feature engineering (since it has rich data), but it doesn't include the last 5 years of matches.

- *Ease of Obtaining:* Easy – available on Kaggle (as an SQLite database or CSV export). No scraping necessary.
- *Cost:* **Free.**

- *Notes:* This dataset was used in several past projects. For example, one 2024 project used it to incorporate player attributes and betting odds into match outcome models [7]. If you need detailed player-level features or historical betting odds for older seasons, this is a goldmine. However, you would need to supplement it with newer data for 2017–2025.

- **League-Specific Kaggle Datasets** – Kaggle also has many community datasets focusing on specific leagues or competitions. For example:

- *English Premier League results (1993–2022):* A dataset of all EPL matches from 1993/94 up to 2021/22 [8]. This likely includes scores and basic stats for each match. Reliability is high (sourced from official results or Football-Data), but it may not have the latest seasons and might not update further.
- *Serie A matches (2015–2023):* (From a Kaggle challenge) which provided recent 8 seasons of Italian Serie A with odds for outcome prediction [9].
- *International football results (1872–2023):* A dataset of ~48,000 international match results [10] – useful if you ever model national team games. It's reliably compiled and updated through 2023, but covers national teams (not club leagues) which may be beyond your current scope.

Similar Projects & Their Data: Many Kaggle notebooks and blog posts exist on match outcome prediction – they often cite using the above datasets or combinations:
- Some used the **Hugo Mathien European database** to get match results plus FIFA-based ratings [6] [7].
- Others compiled **Premier League data from Kaggle or Football-Data**. For example, one Kaggle project by user *siddhrath* aggregated EPL matches (early 2000s to recent) into a single "football match dataset" (with a few older seasons filled in separately). This was used in a Kaggle notebook for Premier League outcome prediction, indicating data came mostly from a Kaggle source and likely Football-Data for any gaps [11].
- Advanced projects combined multiple sources: one 2024 Medium blog integrated **transfer market spending data (scraped from Transfermarkt)** with match outcomes from the Kaggle European dataset to study the effect of club finances on results [12] [13]. Others incorporated **EA FIFA ratings, betting odds, and team form** to engineer features (e.g. projects by Kaggle users *Airback*, *Samuel Benichou*, *msoczi* achieved ~53–55% accuracy using such enriched data) [14] [15]. These examples suggest that combining match outcome data with supplemental features (player stats, team ratings, odds, etc.) can improve predictive models – and the datasets above make that feasible.

2. Football-Data.co.uk (Historical Results & Odds)

[16] **Football-Data.co.uk** (also referred to as *Football Data UK*) is one of the most reliable free sources for historical match results **with betting odds**. It provides CSV files per league and season.

- *Description & Coverage:* Football-Data offers **downloadable CSV files** for major leagues in Europe (and some worldwide), typically covering top divisions and some lower divisions. For example, in England it has Premier League, Championship, League One, League Two, and even the Conference/National League [17]. For other countries, it usually includes the top division (and sometimes 2nd division for the biggest nations) [18]. Each file includes every match of that season with columns for full-time score, half-time score, and often match statistics (shots,

corners, cards) and odds from various bookmakers [19] [20] . This means you get not just W/D/L outcome but also rich context (which can be useful features if needed).

- *Reliability:* **Very high.** This site has been around for many years and is widely used by sports modelers and bettors. Data is accurate and consistent (they provide a notes file explaining all columns). In fact, the large Kaggle dataset mentioned above was built "with most of the data derived from Football-Data.co.uk" [2] . The **betting odds** come primarily from known bookmakers (e.g., Bet365, Pinnacle) and are reliable indicators of match expectations [4] .
- *Update Frequency:* **Frequent/ongoing.** Football-Data updates their files **immediately after each matchday** – often daily or weekly during the season. The site shows a "last updated" date (e.g., updated to 20/08/2025 for the current season) [21] , indicating very recent results are included. You can count on it for current-season data as well as final data for past seasons.
- *Ease of Access:* **Very easy.** You can directly download CSV files from their website – no login or API needed [22] . This makes it trivial to acquire data for multiple leagues (though doing it manually for many leagues is a bit laborious; you can script the downloads since the file URLs follow a pattern). The data is well-formatted for analysis.
- *Cost:* **Free.** The site is free to use (they sustain via affiliate links/ads). They explicitly note the data files are free to access [23] .
- *Notable Limitations:* The main limitation is **coverage** of leagues. Football-Data focuses on popular leagues and doesn't have lower-tier or more obscure leagues beyond a point. For example, a user noted it covers only top 2 German leagues and one division in many countries, missing smaller leagues like lower German divisions, many Balkan leagues, lower Nordic leagues, etc [18] . If your project eventually needs very extensive coverage, you'd need additional sources for those. But for **Premier League and other major European leagues (La Liga, Serie A, Bundesliga, Ligue 1, etc.)**, Football-Data's coverage is excellent and up-to-date.

3. Open Data Repositories (DataHub, OpenFootball, etc.)

There are open-source projects aggregating football data, which can be freely used and are sometimes updated by the community or data platforms:

- **DataHub's English Premier League Dataset:** DataHub.io (an open-data platform) hosts an EPL dataset that compiles match data from 1993/94 up to the current season [24] . This is essentially the Football-Data.co.uk EPL files bundled together with daily updates.
- *Reliability:* Very high – it's explicitly *"sourced from the Football Data UK website"* [25] , so the data quality is the same as Football-Data's.
- *Update Frequency:* **Daily.** DataHub's page mentions **daily updates** to include new matches (including mid-week games) [26] . So it stays current through each season.
- *Ease of Access:* Easy – DataHub provides the data in CSV format, organized by season (e.g., `season-2223.csv` for 2022/23) [27] . You can download files individually or possibly use their API. The dataset is under an open license [28] .
- *Cost:* **Free** (open data license).
- *Content:* It includes detailed match stats: full-time and half-time scores, corners, yellow/red cards, etc., for each game [20] . Essentially the same fields as the Football-Data files (which are quite comprehensive for match statistics).

- *Note:* This is a convenient way to get the entire Premier League history in one place. If you plan to focus on Premier League for modeling, this single source could cover ~30 years of data with minimal effort and up-to-date results [24] . (For last 5 years specifically, you could just use the recent seasons from here.)

- **OpenFootball (football.db):** An open-source community project that maintains football results in a structured text format (public domain) [29] . Volunteers contribute data for leagues and tournaments worldwide.

- *Reliability:* Generally good for major competitions. The data is often cross-verified with official results. However, since it's volunteer-driven, some less popular leagues or recent matches might lag or have gaps if not updated by someone. Major leagues and international tournaments are usually well-covered.
- *Update Frequency:* **Active, but varies by league.** The project shows updates for 2025/26 league schedules for big leagues [30] . In-season updates happen, but timeliness can vary – popular leagues are updated quickly, smaller leagues may be slower.
- *Ease of Access:* Moderate. Data is stored in repositories (on GitHub) as human-readable text (YAML/CSV format). For example, each league has files for each season with fixtures and results. You can download or `git clone` these and then either parse them or use provided tools to import into an SQLite database (they offer a script to load the "football.db"). It's not as plug-and-play as a CSV, but it's structured. No need for web scraping since the data is openly given, but some assembly is required.
- *Cost:* **Free** (public domain).
- *Content:* Typically includes dates, team names, scores, and sometimes goal scorers. It does **not include odds or detailed stats** – it's more focused on results and basic info. Good for getting the raw outcomes of matches in many leagues.
- *Use case:* If you want to expand to many leagues and have a consistent format, OpenFootball is a solid base. You could combine it with other data (for example, one might take results from OpenFootball and then merge with odds from elsewhere, etc.). But since you specifically want model features to predict outcomes, you might still lean on Football-Data or Kaggle for richer features. OpenFootball's strength is breadth and openness.

4. Web-Scraping Options

For data not readily available as a bulk dataset, **web-scraping** can be an option. Some well-known sites for scraping football results and stats include:

- **WorldFootball.net:** An extensive archive of football statistics and results. WorldFootball provides match results, lineups, goal scorers, tables, etc., for leagues all over the world (from top divisions to lower tiers). It's known for historical depth and accuracy.
- *Reliability:* Excellent – WorldFootball is regarded as a very accurate source of historical data (many researchers use it). If a match was played, WorldFootball likely has the correct score, and often additional info like who scored, attendance, referee, etc.
- *Update Frequency:* Usually updated **within a day or two** of matches. For major leagues, updates are prompt (often same-day).
- *Ease of Access:* Needs scraping or use of a helper library. The site doesn't offer official CSV downloads, but its HTML is consistent, making it scrape-friendly. In fact, there is an R package `worldfootballR` that directly pulls data from WorldFootball (and FBref) into data frames [31] [32] . If you are comfortable with R or Python scraping, you can automate retrieval of results. Alternatively, one can manually copy tables for small tasks, but for thousands of matches automation is better.
- *Cost:* **Free.**
- *Content:* Very rich. For example, you can get not only scores but also team line-ups, player statistics for the match, etc. However, for a straightforward outcome dataset, you might just scrape results and basic stats (which are readily visible).

- *Note:* WorldFootball's breadth is a big plus – it covers virtually **every country and league** (including those not covered by Football-Data). If your project expands beyond mainstream leagues or if you want to validate data, this is a go-to. Scraping it will require some effort (or using worldfootballR which simplifies the task in R).

- **FBref.com:** A site by Sports-Reference that provides detailed team and player statistics for many leagues (including advanced metrics like xG, etc.). FBref has results and rich stats.

- *Reliability:* High – data is sourced from Opta and other reliable stat providers.
- *Update Frequency:* Quickly after matches (often same day).
- *Ease of Access:* No bulk download, but the site's tables can be scraped. There are tools (like `worldfootballR` and others) that can retrieve FBref data for an entire season or league. You could also manually use their CSV export for certain tables on the site. Effort is required to get it in bulk, though.
- *Cost:* **Free.**
- *Content:* Includes things like possession, shots, expected goals (xG), etc., along with results. These can be great features for analysis, but note they are **post-match stats** – for a predictive model (pre-match), you'd use such stats as historical averages or trends rather than the current match's values (since they're not known before the match).

- *Note:* If you plan to incorporate team performance metrics (e.g., average shots, goals, xG from past games as features), scraping FBref or using their data via an API can be useful. However, for purely the outcome and basic data, it might be overkill compared to Football-Data which already gives a lot.

- **Wikipedia or Official League Websites:** Some projects resort to scraping Wikipedia's season pages or the official league archives for results.

- *Reliability:* Generally good (Wikipedia is usually accurate for final scores; official sites are obviously accurate).
- *Update Frequency:* Wikipedia often updates within minutes/hours after a match (thanks to contributors). Official sites update in real-time or same day.
- *Ease of Access:* Varies. Wikipedia is scrape-friendly (consistent markup for tables), but each season/league is a separate page. Official sites differ widely in structure and may have anti-scraping measures or require navigating page by page.
- *Cost:* Free.
- *Note:* Given the easier options above, scraping Wikipedia or others is a fallback if you need something very specific not covered elsewhere (or if you can't find a ready dataset for a minor league). For mainstream needs, the above sources are preferable and less labor-intensive.

5. Sports Data APIs

Another route is to use APIs provided by sports data services. These can return structured data (JSON/XML) for matches, which you can collect and build your own dataset. Some noteworthy ones:

- **Football-Data.org API:** *(Not to be confused with Football-Data.co.uk)*. This is a free-tier API for football data [33].
- *Reliability:* Good – many developers use it; it's considered a reliable and up-to-date source for scores, fixtures, standings, etc.
- *Update Frequency:* **Real-time.** It provides live scores, or at least very prompt updates for matches as they finish. You can pull results as soon as games are done.

- *Ease of Access:* Moderate – you need to register for a free API key. The free tier has generous usage limits (enough for personal projects) and covers a decent number of competitions [34] . You'd write scripts to hit endpoints (e.g., "all matches in Premier League 2021/22") and collect the data. The API returns data in JSON; you'll need to parse and store it.
- *Cost:* **Free tier** for basic data. They do offer paid plans for expanded access (for example, an add-on for detailed odds requires a paid subscription) [35] . But for just results and standard data, the free tier is usually sufficient, covering dozens of competitions.
- *Coverage:* Very good – major European leagues, many smaller leagues, and international competitions. Check their docs for which leagues are included in free tier (it's quite broad).

- *Use case:* If you prefer not to manually download files and want to automate data collection/ updating within code, this API is convenient. For instance, you could script to fetch the last 5 seasons of Premier League, La Liga, etc., via the API. The downside is that assembling a large historical dataset via API calls can take time and careful handling of pagination/rate limits. In contrast, downloading CSVs from Football-Data might be quicker for bulk data. A nice benefit, however, is that you can easily update your dataset in the future by scheduling API calls for new matches (keeping your dataset current).

- **SportMonks API & Data**: SportMonks is a commercial API that was used in a Kaggle competition for match outcome probabilities [36] .

- *Reliability:* Very high – it's a professional data provider with extensive information (including odds, player stats, etc.).
- *Update Frequency:* Real-time/live.
- *Ease of Access:* Requires an API key and **paid subscription** for full access. They have a free trial but generally, to get a large historical dataset, a paid plan is needed. The Kaggle competition *"Football Match Probability Prediction"* referenced SportMonks data (participants predicted outcomes for ~150k matches using each team's recent 10-match history as features) [36] . SportMonks data is rich but since our focus is on free/accessible data, it may not be ideal unless budget allows.
- *Cost:* **Paid (beyond a limited free trial).**

- *Note:* If you were open to paying for data or wanted an alternative source of odds and detailed stats, SportMonks is an option. But given the free resources above can cover your needs, you might not need this. (Many community projects avoid the cost by using the free sources mentioned and get very good results.)

- **FiveThirtyEight's SPI Dataset:** While not a traditional API, FiveThirtyEight provides a dataset of their Soccer Power Index ratings and match predictions which can serve as a feature source. They publish **SPI ratings and win/draw/lose probability forecasts for matches** in top leagues, historically back to around 2016. This can be obtained via their GitHub or on Kaggle [37] .

- *Reliability:* High – FiveThirtyEight's data science team curates this. The data includes actual match results along with their model's expected probabilities.
- *Update Frequency:* Updated throughout each season (after each matchday, the SPI ratings and upcoming match forecasts refresh). If you use historical files, you'll get data up to whatever date they last updated (often current season).
- *Ease of Access:* Easy – available as CSV (e.g., `spi_matches.csv`) on their GitHub or Kaggle [38] . No need to scrape; just download the file.
- *Cost:* **Free.**

- *Use as Feature:* The SPI dataset could be **combined** with your outcome data to add features like team ratings or even use their win probability as a feature. Essentially, FiveThirtyEight's model distills a lot of info (team strength, prior performance, etc.) into rating numbers, which might boost your model's accuracy if used appropriately. However, be cautious: if you use their win probability directly as a feature, your model might just learn to mimic FiveThirtyEight. Some projects instead use SPI ratings as one input among many. It's an interesting data source to consider if you want to incorporate an external team-strength metric.
- *Example:* A Kaggle dataset "Club Soccer SPI Ratings and Forecasts" provides match-by-match SPI data back to 2016 [39] . This means for each match you'd have columns for home team SPI, away team SPI, and predicted win/draw/lose probabilities, in addition to the actual result. It's a rich augmentation to raw results data.

6. Comparison of Dataset Options

To help you evaluate each source, here's a **summary comparison** of key factors:

- **Reliability:** All listed sources are considered reliable. **Football-Data.co.uk and Kaggle's compiled data** (which is largely derived from it) are highly trusted for accuracy of scores and odds [2] . **WorldFootball/FBref** are extremely accurate for historical data. **OpenFootball** is generally reliable but might require double-checking for any community edits. **APIs (Football-Data.org, SportMonks)** are reliable as they aggregate official info. In short, reliability is high across the board, with minor caveats for community-contributed data.

- **Frequency of Update:**

- For **current ongoing data**, Football-Data.co.uk (and DataHub's EPL set) are updated **daily/weekly** [26] , APIs provide **live updates**, and scraping targets like WorldFootball/FBref get updated within a day of matches. Kaggle static datasets are usually updated only when the author refreshes them (the big 2000–2025 set was updated to Jan 2025, but may not update further automatically). If you need continuous updates, integrating an **API or a scraping routine** will help. For a one-time model training (using last 5 years of historical data), a static dump from Kaggle or Football-Data will suffice.

- **Historical completeness (last 5 years):** All these sources can provide the full set of matches for, say, 2018–2023. Kaggle 2000–2025 covers that easily. Football-Data has season files for all those years. OpenFootball and WorldFootball have them as well. So you can obtain the last 5 years from any combination of these.

- **Ease of Obtaining:** The **easiest** options are those with ready CSVs: **Kaggle datasets, Football-Data.co.uk, DataHub**. These require no coding to acquire the raw files (just download). **APIs** and **scraping** require more work – you'd need to write scripts and handle data parsing. That said, libraries like `worldfootballR` (for R) or `soccerdata` (for Python, as mentioned in a 2024 SportAnalytics blog [40]) can simplify data retrieval from web sources. If you prefer a code-first approach, the API route might integrate nicely with your pipeline. If you prefer to start by exploring data in Excel/Pandas, the CSV route (Kaggle/Football-Data) is quicker.

- **Free vs Paid:** All sources we listed are free except SportMonks (and possibly other commercial APIs). **Football-Data.co.uk, Kaggle datasets, DataHub, OpenFootball, WorldFootball, FBref, FiveThirtyEight** – all free. SportMonks and similar premium APIs or Opta data are paid and likely not needed given the quality of free data available. Stick to free unless you identify a specific gap that only a paid source can fill.

- **Other Metrics (Coverage & Content):** If you aim to include **multiple leagues** beyond the Premier League:

- The **Kaggle Club Football (2000–2025)** set and **OpenFootball/WorldFootball** have the *broadest* coverage (dozens of leagues globally). Football-Data.co.uk covers all major European leagues and some others (like MLS, Brazil, etc.), but not every smaller league [18]. If combining data, you might use Football-Data for all the leagues it offers (with odds and stats), and supplement with another source for any league not covered.

- In terms of **data fields**: Football-Data and the Kaggle 2000–2025 give you outcomes plus goals, odds, and some match stats (shots on target, etc., for many leagues) [20]. The European Soccer DB (2008–16) gives even more (player line-ups, etc.), but for recent years you'd rely on other sources for similar depth. If you need **player-level data or advanced stats**, you'd consider FBref or similar scraping. If you just need team-level features (like recent form, average goals, etc.), you can derive those from the basic match results.

- **Combining Sources:** It's very feasible to **merge data from multiple sources** by matching dates/ teams:

- For example, you could take the Football-Data results (for accurate scores/odds) and merge with team strength metrics from ClubElo or SPI. The Kaggle 2000–2025 set essentially did this for you by including Elo and odds.
- If you want to be **up-to-the-minute**, you might use an API to fetch the latest season's ongoing results and append to your static dataset.
- Many projects create a unique dataset by merging: e.g., one added **TransferMarkt club financial data** to match outcomes to see if spending influences wins [41] [42]. Another added **FIFA player ratings** to team profiles [43] [44]. Such merges require a common key (like team names and date/season). It's work, but the free data above gives you the pieces to do it.

In summary, you have **ample free data options** for training a match outcome model. For last 5 years of Premier League (and other top leagues), **Football-Data.co.uk (via direct download or DataHub) is highly recommended for its reliability and richness** [20]. The **Kaggle "Club Football 2000–2025" dataset** is a great all-in-one source as well, already compiled from similar inputs [2]. You can trust these for accurate results and use them as the backbone of your model's dataset. Then, depending on how complex your model's features get, consider augmenting with additional data (odds, rankings, player stats) from sources like FiveThirtyEight SPI or others. Each dataset above has strengths, and many are complementary – combining them can yield a unique and powerful training dataset tailored to your project's needs.

Sources:

- Kaggle Club Football Match Data (2000–2025) – user adamgbor (data from Football-Data.co.uk and Club Elo, ~230k matches) [2] [3]
- Reddit discussion of free football data (sportsanalytics/datasets) – confirms dataset source and openness [2] [3]
- Football-Data.co.uk (historical results & odds downloads) – England leagues example [21] [19]
- DataHub EPL Dataset (1993–2025) – overview and features (sourced from Football-Data) [25] [20]
- Medium blog (Fion Ouyang, 2024) – describes using Kaggle European Soccer DB (Hugo Mathien) and its content [7] [43]
- SportAnalytics blog (2024) – on accessing free soccer data, mentions worldfootballR and Football-Data.co.uk as resources [31] [32]

- Kaggle/538 SPI data reference – match-by-match SPI ratings & forecasts since 2016 [37]
- Reddit (r/algobetting) on Football-Data.co.uk limitations – covers of leagues in various countries [18]

---

[1] [2] [3] [4] The biggest free & open Football Results & Stats Dataset : r/datasets
https://www.reddit.com/r/datasets/comments/1hqyyzj/the_biggest_free_open_football_results_stats/

[5] The biggest open & free football dataset just got an update! : r/sportsanalytics
https://www.reddit.com/r/sportsanalytics/comments/1j0zx1f/the_biggest_open_free_football_dataset_just_got/

[6] [7] [12] [13] [14] [15] [41] [42] [43] [44] Predicting Soccer Match Outcomes:A Data-Driven Approach | by Fion Ouyang | Medium
https://medium.com/@fion.ouyang/predicting-soccer-match-outcomes-a-data-driven-approach-df6bea2429ec

[8] English Premier League (EPL) Results - Kaggle
https://www.kaggle.com/datasets/irkaal/english-premier-league-results

[9] Serie A Match Outcome Prediction Model - Kaggle
https://www.kaggle.com/code/marcelbiezunski/serie-a-match-outcome-prediction-model

[10] International football results from 1872 to 2025 - Kaggle
https://www.kaggle.com/datasets/martj42/international-football-results-from-1872-to-2017

[11] Football-Premier-League-Match-Outcome-Prediction - Kaggle
https://www.kaggle.com/code/acalles/football-premier-league-match-outcome-prediction

[16] [31] [32] [40] Get Free Data for Powerful Soccer Analytics Projects
https://sportanalytics.com/get-free-data-for-powerful-soccer-analytics-projects/

[17] [19] [21] [22] [23] England Football Results Betting Odds | Premiership Results & Betting Odds
https://www.football-data.co.uk/englandm.php

[18] Football(Soccer) historical data + fixtures + odds through API or datasets : r/algobetting
https://www.reddit.com/r/algobetting/comments/1mcb9hi/footballsoccer_historical_data_fixtures_odds/

[20] [24] [25] [26] [27] [28] Discord
https://datahub.io/core/english-premier-league

[29] [30] Welcome - football.db - Open Football Data
https://openfootball.github.io/

[33] API Reference - Football-Data.org
https://www.football-data.org/documentation/api

[34] End-to-End Data Project About Collecting And Summarizing ... - Reddit
https://www.reddit.com/r/dataengineering/comments/1iziyyr/endtoend_data_project_about_collecting_and/

[35] Pricing - Football-Data.org
https://www.football-data.org/pricing

[36] Football Predictions: Kaggle competition - Sportmonks
https://www.sportmonks.com/blogs/football-predictions-kaggle-competition/

[37] [39] Club Soccer: SPI Ratings and Forecasts - Kaggle
https://www.kaggle.com/datasets/thedevastator/club-soccer-predictions-spi-ratings-and-forecast

[38] Soccer Spi - DataHub.io
https://datahub.io/core/five-thirty-eight-datasets/datasets/soccer-spi