Referee Bias and Racial Discrimination in the English Premier League: A Machine Learning Analysis

Prepared by:

Erwin Medina
March 2025

Submitted To:

Dr. Kenneth Kung
Department of Computer Science
College of Engineering and Computer Science
California State University, Fullerton

CALIFORNIA STATE UNIVERSITY
FULLERTON™

Department of Computer Science

CPSC 597 / 598 PROJECT / THESIS DEFINITION

To The Graduate Student:
1. Complete a project proposal, following the department guidelines
2. Have this form signed by your advisor and reviewer / committee
3. Submit it with the proposal attached to the Department of Computer Science.

[X] Project
[ ] Thesis

Please print or type.

Student Name: **Erwin Medina**   Student ID: **888233475**

Address: **1812 Anacapa, Irvine, CA, 92602**

Home Phone: **909-503-2834**   Work Phone: _____

Email: **erwinmedina@csu.fullerton.edu**   Units: **24** Semester: **Spring**

Are you a Classified Graduated Student?   Is this a group project?

[X] Yes   [ ] Yes
[ ] No   [X] No

Proposal Date: _05/2025_

Tentative Date for Demonstration/
Presentation/Oral Defense: _____

Completion Date: _____

Tentative Title: Referee Bias and Racial Discrimination in the English Premier League: A Machine Learning Analysis

We recommend that this proposal be approved:

Faculty Advisor: _____
             Printed Name   Signature   Date

Faculty Reviewer: _____
             Printed Name   Signature   Date

Faculty Reviewer: _____
             Printed Name   Signature   Date

Table of Contents

1. Introduction

Referee bias in the sport of soccer has been a subject of debate for some time. Bias, in this context, can be defined as having favoritism for a specific team, such as by carding one team more than another throughout multiple games over time, allowing additional time atop of the extra time after the 90 minutes, or simply not calling and overlooking deliberate fouls. During the 2019-2020 Premier League season, a Video Assistant Referee (VAR) was introduced. A VAR is a referee sitting offsite, in front of multiple screens, with the ability to replay fouls, offside calls, penalties, and goals. They have direct communication with the on-field referee and can alert the referee in real time. VAR was designed and implemented to help on-field referees make accurate decisions and maintain the integrity of the game.

Despite this technological advancement, there are limitations placed on this referee to avoid VAR from dictating the flow of the game. VAR is only allowed to make decisions on direct red cards such as denying a goal scoring opportunity (DOGSO), altercations, or dangerous fouls. They also have power on whether a goal is valid or not, along with determining if a player was offside. Despite the sport having five referees (1 on-field, 2 linesman, 1 fourth official, 1 VAR), mistakes continue to occur, and fouls are continuously missed [11]. Fans, analysts, and pundits agree that referees are inconsistent across games when calling fouls or handing out cards. This inconsistency breeds doubt and confusion as to whether the referees fully under the rules of the game or are biased and avoiding making calls due to favoritism.

Favoritism is only half the battle. Within the sport, there exists a tremendous amount of racism by the fanatics [12, 13]. Football players are often criticized, ridiculed, and sent racist comments over social media when they miss crucial penalties, after losing a game, or if they gave a lackluster performance. Players will either report these messages to the authorities, and or post them on their social media by adding a small caption of, "this needs to stop" or "disgusting!". Due to the amount of racism and animosity, the English Premier League (EPL) began a movement called, "No Room for Racism". Both players and referees wear a badge with the quote on their shoulder, and the quote can be seen throughout the game as well. A few actions that the EPL have taken against racism are, investigating over 2500 cases of online discriminatory abuse targeted at players since 2020, providing mental and emotional support for players, and providing an online reporting system on club websites and the EPL website [1]. Even with all the support, advertisements, and repeated messages, racism, unfortunately, continues to plague the sport. But does this racism transcend into the referees as well? Are there any patterns as to which referees are handing out cards to the players of a certain nationality more than others?

2. Problem Domain

Many soccer match predictor tools exist throughout the internet, with some ranging from amateur models that are not well known and hard to understand, to higher end, refined, and well-known models such as FiveThirtyEight's prediction tool. Oftentimes, these predictor tools are hidden behind paywalls because they're valuable monetary tools that can lead to successful bets for those that gamble. Some tools fail to mention how they draw their conclusions, while others like FiveThirtyEight outline exactly how they draw their predictions and the parameters they use [2]. Many of these tools use the same parameters such as how often a team is scoring, shots on target, their form (winning/losing streaks), whether they're away or at home, using previous seasons' data, and much more. All these parameters are crucial for determining how well a team is doing and whether if they can produce a win in their following game. But a parameter that seems to be ignored or unaccounted for is the referee.

Given how vital a referee is in each match, it is difficult to understand the reasoning on why the referee parameter is ignored. A couple of potential conclusions that can be drawn are that it could be possible that a referee makes little to no impact on the outcome of the game relative to all the other parameters, which most would disagree with or have a hard time believing. Alternatively, they do have an impact, but to reduce harassment, threats, or even physical violence against the referees by fans, predictor tools simply fail to mention that as a parameter. Along the same lines, there are no tools that mention whether a referee has a racial bias, presumably for the same reasons. Determining whether a referee has a bias in the sport is crucial for viewers, analysts, betting prediction sites, and most importantly for ethics and impartiality of the game.

Within Europe, there are many rivalries among cities and teams. For example, in England, Liverpool Football Club (FC) and Manchester United FC have a rivalry dating back several decades if not over a hundred years due to their cities' histories. With some referees originating or having grown up in Manchester and refereeing a game of Liverpool vs Manchester United, many fans began to feel unsettled that this would generate a conflict of interest. Are there patterns with less fouls for the Manchester team with a Manchester referee or is it irrelevant?

2.1. Key Issues

2.1.1. Data Gathering

An important issue with determining whether a referee has a bias is gathering key data points. In Table 1A in the appendix, I've outlined which data points I aim to collect, what type of data point they are, and what my reasoning is behind that data point. This

clarity allows readers to determine if my proposed analysis is on the right track or if it requires refining. Machine learning models are only as good as the data they are given, and training a model can be challenging without the correct data. The majority, if not all, of this information can be collected manually, or web scraped, through websites like the Premier League official website, Google, Wikipedia, LiveScore, and much more.

2.1.2. Safety Concerns

Soccer has a tendency of riling individuals up and unfortunately experiences hooliganism and ultra fans. Hooliganism are individuals who cause disruptive, unlawful behavior such as rioting, bullying, and vandalism. While Ultras are fans who are known for their extreme support for their team but have been linked to extremist movements [3]. Although soccer attracts all types of individuals, it should be noted that not all individuals are violent or extreme. However, as mentioned before, when players miss crucial penalties, lose a game, or perform poorly, fans tend to have a negative reaction and send hateful messages through social media. Referees are not exempt from this. An example of this is Anthony Taylor, an English referee, who partook as a referee in the Europa League final, a notable European cup. After questionable on-field decisions, he was heckled and harassed later that day at the airport with his family, along with chairs being thrown at him [4]. Another notable example occurred recently in Spain when a player, Jude Bellingham from Real Madrid, was sent off with a red card for foul language that the referee thought was directed at him but had misunderstood. Jose Munuera, the referee for the game received an extreme level of harassment, death threats, and had his family harassed as well, leading to him shutting down his social media [5]. Thus, we can see that violence and harassment are not solely targeted at players.

Although using a referee as a parameter is important to understand whether there is a bias, racial or not, it is crucial to note that analyzing referee bias is not intended to cause violence, hatred, or anger towards specific referees. Instead, it is intended to be used as a tool to help understand if such a bias exists, and if so, are there ways to prevent, reduce, or eliminate this bias through additional training.

2.1.3. Subjective Decision-Making

Soccer is a contact sport where fouls and injuries can happen within seconds and decisions need to be made on whether that foul was a warning, yellow, or a red. Split second decision making will always be prone to error, and it can be challenging to determine whether the call was biased or not. The referee also experiences social or external pressures from the home crowd leading to home team bias. External factors such as crowd noise, proximity of the crowd, and instant decision making are all challenges

that are incorporated into whether a decision was correct or not. What also makes this topic difficult is that not all referees apply the rules equally. One might consider something being a handball while another might see the same play and determine it is not a handball. However, this paper and project are not determined to find comparisons between referees but instead patterns between referees and teams.

2.1.4. Legal Barriers and Public Perception

Referee bias studies also run the risk of legal pushback if studies single out specific referees or leagues. Referees could sue for defamation if they believe that their reputation is damaged due to a study that suggests they card more Nigerian players than English players, for example. And despite the research being sound and valid, the legal costs could be detrimental to publishing findings. Fédération Internationale de Football Association (FIFA) has historically been reluctant to acknowledge bias in referees or officiating. Although they have been accused of tolerating racial discrimination among referees in the past, they rarely release any disciplinary records. Due to most investigations occurring internally, leagues have little to no incentive to acknowledge that there is any bias in their league. Thus, damaging not only the referee's credibility but also the league's could lead to further legal pushback or defamation lawsuits.

Public perception of referee bias is often dismissed as "cherry picked" data or being "fan-driven narratives". Fans are highly invested in the sport and their team and will accuse the referee of bias anytime a controversial call goes against them. Thus, it is difficult to separate research findings from emotional reactions, which can undermine credibility of the project. Another issue is confirmation bias. If the research finds that there is a significant bias towards Latino players, for example, those who already believe this will support the project, but those that do not or are skeptics would dismiss the findings. And of course, if it shows that there is no evidence of bias, then some may argue that the study or research was flawed and did not use the proper measurements. Public perception of bias, along with legal barriers are plausible reasons for the lack of further research into the topic.

3. Literature Survey

3.1. Survey Paper 1

Referee Bias by T. Dohmen and J. Sauermann [6]

Dohmen and Sauermann go into depth about their findings of referee bias. First, they outline where in the sport they found this bias, and later dive deeper about potential causes on why referee's might have acted that way. They take note on referees having a bias in allowance for time lost after the 45 minutes, also known as stoppage time. This allowance of time is determined by time lost during substitutions, injuries, time wasting, or other causes. And although the amount allotted is at the discretion of the referee, Dohmen and Sauermann mention that more time is often provided to the home team when they are behind. Papers they analyzed mention that stoppage time was approximately 113 seconds longer when the home team was behind one goal, compared to when the home team was ahead by one goal. This evidence suggesting that a referee weighs the social reward from the crowd, thus allowing for more or less time.

Other crucial points that they factor in are biases in other decisions such as goals, penalties, and yellow/red cards. As these decisions impact the game more than providing extra time, referees are known to be more cautious to avoid making the wrong call. Within goals, they determined that there was no evidence that referees award more towards the home team than the visiting teams. In their analysis of penalty kicks, they found that referees were biased in awarding penalty kicks to home teams. With yellow and red cards, they found research that supported that referees tend to favor the home team by punishing their players less and those of the visiting team more strongly. Dohmen and Sauermann mention that the crowd and spectators affect the referee's decisions through social pressure but struggled to find conclusive evidence on how.

A notable item that I found in this paper was that the data analyzed came from the 1990's and early 2000's. Although the sport has not changed, technological advancements in the sport have, such as goal-line technology (checking whether the ball crossed the goal line), VAR, and semi-automated offside cameras. These advancements aid the referees in making more precise and accurate decisions. The laws of the game have also changed such as adjustments to the handball rule, goalkeeper rules for penalties, changes to the number of substitutions, and more. Along the laws of the game, there have also been a new era of referees who might have a different view or opinion on how to manage the game versus 20 years ago. Thus, a revisit of the data and conclusions drawn are important in ensuring the integrity of the game.

3.2. Survey Paper 2

Do Soccer Referees Display Home Team Favoritism? by B. Lucey and D. Power [7]

Lucey and Power delve into determining whether referees provide a home team advantage over the visiting team. They touch on topics and previous research that Dohmen and Sauermann referenced in their paper, such as social pressures from the

crowd, additional time after the 90 minutes, but also how the biases increased by 40% from the beginning to the end of the season. The influence of the crowd tends to have a heavy sway in determining the referee's decision making. A study was done on a sample of referees, half were given audio and video of tackles, while the other half were only given video. What was noticed was that referees with audio were more reluctant to give fouls and were uncertain in their decisions because of the crowd noise. In contrast, the referees with just video were able to classify fouls and make more confident decisions. Although we can see there are external factors that play into the decision making of the referee, it is important to note that my focus is to determine whether there are patterns against certain teams, not simply a home team. In their conclusion, they state that Italian and American referees are not influenced by social environments, but that referees do tend to favor home teams but only in close games. They also outline that their study was only partially done due to examining only one potential area where biases exist, added time.

Although their study is tailored and focused on more of a home team advantage, it is clear and obvious that referees experience some type of bias either through internal or external pressures.

3.3. Survey Paper 3

Favoritism and Referee Bias in European Soccer: Evidence from the Spanish League and the UEFA Champions League by B. Buraimo, R. Simmons, and M. Maciaszczyk [8]

In this paper, they analyze and find evidence of referee bias in favor of home teams within the Union European Football Association (UEFA) and Spain's first division league. Although many papers cover similar topics of home teams getting an advantage based on the proximity of the crowd, or if the home team is trailing by a single goal, this paper is slightly different from the rest. Within their testing, they utilize minute-by-minute analysis to control for within-game events. They also discovered that within the UEFA Champions League, a competition that includes some of the greatest club teams in Europe, the organizers appoint neutral or non-biased referees from countries outside of the represented teams in any of the matches.

They conclude their paper by saying that they hypothesized the Champions League competition would have less bias due to the neutral referees versus Spain's first division league. However, what they discovered is that whether the stadium has a running field track dividing the soccer pitch and the fans plays a significant part in the bias. They mention that seeing a decrease in away team yellow cards and more cards for home teams when there is a track present implies that the referees do carry a certain bias despite having neutral referees. Unfortunately, this fails to touch on the topic of whether

a referee has favoritism or biases for certain teams. Despite this, the paper does include a critical statistical analysis of referee bias in the sport and reinforces the idea of crowd proximity influencing referees.

3.4. Survey Paper 4

An Integrated Conceptual Framework of Decision-Making in Soccer Refereeing by R. Samuel, G. Tenenbaum & Y. Galily [9]

Within this paper, they focus on the decision making of the referee such as any influences, external factors, distance from the foul, positioning of the referee, fatigue, stress, and more. This paper deviates from the over saturated fan proximity research and instead analyzes the complexity of decision making. During the game, referees make numerous repetitive decisions, where many are dependent on the field location of the referee and effective communication of the assistant referees. Although it is evident that the proximity of the referee for a foul impacts the decision making, it is important to note that there exists an optimal distance. Samuel, Tenenbaum, and Galily determined that decision to be approximately 11-15 meters (35-50ft). Any closer and the referee can potentially miss important details. Likewise, any further produces the same result.

Another analysis they focused on was the height of the referee. Their findings concluded that shorter referees tended to produce more yellow cards. But for red cards, it depended on the division. Lower division and short referees had more red cards, but higher divisions and taller referees had the same result. A factor not yet touched on is the decision making of the referee on where to run and position him or herself. A poorly placed referee increases their chances of making a decision error as they are not able to see the important details. But this factor was noted to be attributed to fitness level and physiological fatigue. If positioned correctly, they then must anticipate, through experience and prior knowledge, where to move to next based on how the ball and players are moving. And should the call they make be unpopular, the referee must remain calm and in control to continue with their effective decision making.

They finalize the paper by concluding that there are many critical factors that determine a referee's decision making, and not every factor can be analyzed in one paper. However, they outline a few steps to shift the focus from referee biases to improving successful referee performances. Unfortunately, there is not a mention of whether a referee has a particular team or racial bias. Nonetheless, it sheds light on external factors that were not mentioned before in other research, such as referee height, proximity (optimal distance) to the foul, and mental fatigue. All of which are important factors that can be detrimental on the reasoning behind why a card or foul is given.

3.5. Survey Paper 5

Referee Bias Contributes to Home Advantage in English Premiership Football by R. Boyko, A. Boyko, & M. Boyko [10]

Boyko et al. focus their paper on not only the home advantage bias but whether individual referees vary in their home bias or whether biased decisions contribute to overall home advantage. They examine over 5200 English Premier League matches with over 50 referees and found that home bias differs between referees. They outline that there are different measures of home advantage, such as goal differentials between home and away teams, yellow cards, and penalty differentials. Boyko et al. determined the variability between referees implies that referees are responsible for the observed home advantage but also suggest that the home advantage is dependent on subjective decisions. They then suggest that further research needs to be done on crowd noise, and the referees psychological and behavioral responses to biased crowds.

The unique approach of expanding on the home advantage bias is key for my research topic. Boyko et al. not only approached the problem as an existing issue but expanded on the research already done and produced results of variability amongst referees. Yet, the question lingers on whether these referees have a particular bias for certain teams, or if any racial bias was determined by analyzing the players that were carded. Although their research serves as a good indicator for a bias existing, and varying among referees, can we delve deeper into finding results for team biases and racial discrimination in the sport.

4. Project Objectives and Significance

This project aims to determine whether referee bias exists toward specific teams or players in the English Premier League (EPL), with a particular emphasis on potential racial bias. To achieve this, the first objective is to collect and clean approximately 10 seasons worth of match data, which includes referee information, player nationalities, fouls, yellow/red cards, match outcomes, added time, and much more. The data points to be collected can be found in Table 1A in the appendix section of this paper. The sources of where the data will come from vary from the Premier League's official website, Wikipedia, and LiveScore.

The second objective is to design and train a machine learning model that can analyze this data and identify statistically significant patterns in referee decision-making. This includes detecting whether specific referees consistently penalize certain nationalities or favor certain clubs. Once validated, the third objective is to develop a simple and interactive user interface where users can explore historical referee behavior,

simulate hypothetical matchups, and view predicted outcomes such as likely winners or number of cards issued based on the model's learned patterns.

Identifying and evaluating referee bias is essential to ensuring fairness in the sport. When rules are consistently upheld, it provides confidence to fans, analysts, and betting platforms that the outcomes are legitimate and unbiased. Transparent officiating builds trust in the integrity of the game. Most importantly, this project also addresses the ethical concern of potential racial profiling, reinforcing the need for equality and accountability. Beyond its impact on sports, the project highlights the power of artificial intelligence and machine learning to uncover insights that traditional analysis might overlook or have not captured yet.

5. Project Activities

The project will be broken up into six distinct phases, each designed to build on the previous step to achieve the final objective: identifying patterns of referee bias through ML and presenting the findings in an interactive user-facing tool. The phases cover the full pipeline from data gathering to model training and final deployment. Each phase is described in detail below.

5.1. Phase 1 – Data Collection

Within this phase, data will be collected and compiled using Python's BeautifulSoup web scraping tool or manually depending on the website's design and readability. The data that will be collected can be found in Table A1 in the appendix and ranges from team names, referees, yellow/red cards, player names, and more. Manually acquired data will be saved into a CSV Excel document, while data acquired by BeautifulSoup will utilize Pandas to convert tables of information into a CSV. This approach is done to consolidate and track everything in one place, avoid scattered files, and reduce inconsistent formatting. With all the data centralized in one location, it is easier to jump into Phase 2.

5.2. Phase 2 – Data Cleaning and Processing

This phase is designed to clean up any inconsistencies, anomalies, absent data, but also normalize information such as player names, club names, and encode categorical features such as nationalities. A few examples of normalizing data can be club names such as "Man United" vs "Manchester United", converting timestamps of yellow cards or extra time from "89.5" to 89th minute + 30 seconds, or converting player nationalities into regional codes for encoding categorical values. The overarching goal of this phase is to prepare the dataset for the machine learning model and visualization that will occur in the following phases. By ensuring that all the data is consistent and normalized, I can move toward the next phase.

5.3. Phase 3 – Exploratory Data Analysis (EDA)

Although I will be building our machine learning model with the data I collected along with creating a full stack user interface, it's always best to pinpoint areas or trends of significant importance prior to the machine learning phase. In this phase, I can create visualizations and possibly build or adjust a hypothesis to ensure that time is not being wasted and also highlighting standout points in my data. For example, we can analyze the data by creating bar charts of cards per nationality / team / referee, card distribution by game state (whether a team is losing vs winning), or box plots of added time vs match outcomes. These examples provide further insight on what to look out for in the further stages but also identifies any bias candidates. This helps us guide the model design, the feature engineering, and later how we interpret the results. By the end of this stage, I should have some evidence to suggest that there are specific data points worth looking at and others that show little to no trends. It would also allow me to refine or strengthen my hypotheses, such as "Referee X gives more red cards to African players" or "Referee Y adds more extra time when Home Team A is losing". After we refine our hypothesis, we're ready for Phase 4.

5.4. Phase 4 – Machine Learning Modeling

In this stage, we will apply supervised machine learning techniques to draw meaningful conclusions, such as using regression or classification models. Our objective here is to build and train predictive models that can detect patterns of favoritism, bias, or possibly racial disparity in referee decisions by using historical match data. We want to identify statistically significant features associated with referee's in-game decisions regarding yellow/red cards and added time. The model will ideally help quantify if our hypothesis of referee behaviors matches or aligns with consistent trends, and whether those trends are linked to team affiliation or player nationality. The model that most aligns with our objective would be a classification approach, such as if a player will receive a card in a particular match, or if the likelihood of the card being yellow or red.

Another important aspect of the machine learning model phase is the feature engineering. Building meaningful input features from our data is critical to the output of our model. Examples of input features could be player nationalities, team names, match score at card time, whether a team is home or away, referee name / city / nationality, and much more. Table A1 in the appendix has the input features that will be utilized for the machine learning model.

After we've defined our problem, outlined our objective, built our input features, then we're ready for selecting the model that we want to utilize for our approach. Sticking strictly with one model might not be a wise choice as different models can produce slightly different results. Examples of different models can be a logistic regression, a random

forest, or naive-bayes model. Thus, trying different models can help us determine which one is more accurate or effective for the data that was provided.

Once we've tried a few different models, we can aim to train and validate / evaluate the results. Ideally, we want to provide the model with 70-75% of the data for training, while using 25-30% for testing purposes. Within this portion of the stage, we should be determining the accuracy, precision, recall, and confusion matrix for our results to ensure the model is behaving as expected. Our goal after this is to understand which features are influencing the model's predictions. In doing so, we can determine whether a referee is biased (or not) or if some teams are being penalized more than others.

5.5. Phase 5 – Results and Analysis

The metrics outputted from phase 4 allow us to draw conclusions and meaningful insights. Here we can highlight important trends such as certain referees being statistical outliers, disproportionate yellow/red cards towards certain nationalities, favoritism during close games, and more. Although this phase blends in with the previous, it does allow us to have a period of reflection or retrospection on the outcome of the project. An example of this can be whether the project experienced limitations such as too small of a sample size, or if there was a model bias, and whether we should be revising the data and redoing the training. Tackling these questions with ample time can potentially produce a stronger model in the future, or a quick rework depending on time constraints.

5.6. Phase 6 – Interactive Tool

This section depends on whether Phase 5 detects a pattern of referee bias, racial or otherwise. With the assumption that it does find trends, it would be best to illustrate these trends to a user. By combining machine learning, data science, and full stack engineering, we can create a tool that can be used by viewers, analysts, and anyone interested simply by using a standard laptop environment. The interactive web-based tool will ideally have a simple design that is easy to understand for those inexperienced with the sport. Using Streamlit or Flask (connected with Python) will allow for a seamless transfer of data from the machine learning model to the user interface. In the user interface, a user will be able to select from a dropdown to select referees and match up two different teams. Once done, it'll display the referee profile along with their card history. After the user is ready to move forward, it will then predict a winner along with the expected card counts. This will finalize the project and allow users to see whether the referee chosen has a bias towards the teams selected.

For example, say the user selects "Arsenal" vs "Liverpool" and then selects a referee of "Howard Webb". If this referee has shown bias towards Arsenal in the past and the data suggests that Arsenal wins their games under this referee often, then the predicted output would suggest that Arsenal will beat Liverpool, and we should expect 7

yellow cards. There are many additional parameters that go into winning games such as player fitness, team cohesion, player statistics, etc, which is not what this tool is aiming to achieve. However, it would allow users to interactively see whether having a specific referee assigned to their game would impact their team and thus make an educated guess on which team would win.

The racial bias portion of the project will most likely be omitted from the interactive tool as it might not as accurate or possibly deliver the wrong message. The interactive tool will not predict which players receive yellow or red cards as that would require a continuously updated player database to track team rosters, thus making the racial bias a bit obsolete for the interactive tool.

5.7. Software Requirements Specification

5.7.1. Functional Requirements

- The system shall scrape and preprocess English Premier League (EPL) match data during Phase 1.
- The system shall export match data as CSV files for use in Phase 2.
- The system shall allow users to:
  - Select referees and teams from dropdown menus (Phase 6).
  - Display historical data related to the selected referee and team.
  - Run a predictive model to estimate the number of yellow/red cards and match outcome.

5.7.2. Non-Functional Requirements

- The system shall return predictive results within 5–10 seconds of user input.
- The interface shall be designed for readability and accessibility for non-technical users.
- The system shall be mobile-friendly and responsive to different screen sizes.
- The system assumes that users have basic familiarity with soccer terminology (e.g., yellow/red cards, match outcome).

5.7.3. Constraints

- The system depends on publicly available or manually gathered match and referee data.
- Proprietary data sources, such as internal referee reports or official league datasets, are not accessible.
- Web scraping may be limited by:
  - The HTML structure of the target websites.
  - Anti-scraping measures or rate limits implemented by website owners.

- The accuracy of the machine learning model is constrained by the quality and completeness of the available data.

5.7.4. Assumptions

- All collected data is assumed to be accurate and reliable.
- Users are expected to understand basic soccer-related concepts.
- The interactive tool is intended solely for demonstration and educational purposes, not for production deployment or real-world decision-making.

6. Environment

The development environment for this project will mainly consist of open-source tools and commonly available hardware and software resources. The objective of this section is to ensure that the project can be run and executed on any standard personal machine without requiring specialized resources or hardware (e.g. high-end graphics processing units (GPU) or central processing units (CPU)). Outlined below will be the overall structure of the environment.

- **Operating System:** macOS, Windows
- **Programming Language:** Python 3.10+
- **Integrated Development Environment (IDE) / Editor**: Visual Studio Code
- **Primary Libraries:**
    - Pandas and Numpy for manipulating and processing data.
    - BeautifulSoup4 for web scraping
    - Matplotlib, plotly for visualization and EDA purposes.
    - Scikit-learn, xgboost for the machine learning portion.
    - Streamlit or Flask for building the interactive front-end interface.
- **Data Format:** CSV files will be used for input and output, allowing for easy editing, version control, and readability.
- **Hardware Requirements**: Project will be developed and tested on a MacBook Air with approximately 8GB of Random Access Memory (RAM) and a standard CPU.
- **Version Control:** GitHub will be primarily used for source code tracking and backup.

- **Hosting:** Streamlit

With the above environment, we can ensure that all components of the project, from web scraping to model training, visualization, and deployment, can be run efficiently on a local machine without specialized hardware or external dependencies.

7. Project Results

This project will produce several deliverables which will demonstrate the full lifecycle from data gathering to deployment. The first deliverable will be a clean, well-structured dataset of EPL match data spanning across 10 seasons. Once completed, the next deliverable will be a trained machine learning model that can identify bias patterns in referee decisions (or lack thereof). An example of this would be the machine learning model outputting ("referee name", "is_bias", ["supported teams"]) → "Howard Webb – True – Arsenal". This is the primary deliverable that I aim to produce through this project and hope to deliver to an advisor and the Computer Science (CS) department. However, to add more value to the project, I aim to create a web-based interactive tool where users can select referees and teams, view historical data, simulate a matchup, and see a predicted outcome. The submission of my final presentation and the interactive tool will signal the completion of this project.

8. Project Schedule

This project is expected to span approximately the entire duration of the fall semester of 2025. Below is a table of the timeline.

Phase	Task	# of Hours	Week(s)
Phase 1 – Data Collection	Manual + automated scraping, data exports	25	Weeks 1-2
Phase 2 – Data Cleaning	Normalizing, encoding, missing value checks	15	Week 3
Phase 3 – EDA	Visualizations, adjusting hypotheses	20	Weeks 4-5
Phase 4 – Modeling	Feature engineering, training, evaluation	30	Weeks 6-9
Phase 5 – Results Analysis	Analyze trends, interpret ML results	20	Weeks 10-11
Phase 6 – UI Development	Build and test web tool with Streamlit/Flask	30	Weeks 12-14
Documentation + Final Presentation	Code comments, github polishing, slide prep, functionality check	15	Week 15
Total		**155**	

This timeline allows for overlap between the phases. For example, insights from the EDA phase may directly help with the feature engineering, and the model results may influence how the UI is structured or presented. However, it is expected that there will be minor delays in the modeling and UI development, depending on their success. Thus, I've provided additional time during those weeks to ensure the project is on track.

9. References

[1] "How the premier league continues to fight racism," www.premierleague.com, Apr. 05, 2024. https://www.premierleague.com/news/3951098

[2] "How Our Club Predictions Work" FiveThirtyEight.com https://fivethirtyeight.com/methodology/how- our-club-soccer-predictions-work/ [Accessed Feb. 24, 2025].

[3] W. Haanstra F. Keijzer, "Learning from adjacent fields: the relation between extremism and hooli- ganism", Oct 2018. [Online]. Available: https://home-affairs.ec.europa.eu/document/download/cf539884- 5f8f-4df3-a623-d1434959c622 en. [Accessed: Feb. 24, 2025].

[4] "Roma fans condemned for harassing Europa League final referee Anthony Taylor at airport" ESPN.com. https://www.espn.com/soccer/story/ /id/37776087/roma-fans-condemned-harassment-europa-league- final-referee-anthony-taylor. [Accessed Feb. 24, 2025].

[5] F. Kallas. "Spanish referees condemn abuse of official who gave Bellingham red card'. Reuters.com. https://www.reuters.com/sports/soccer/spain-federation-says-referees-sickened-by-abuse- following-bellingham-red-card-2025-02-18/. [Accessed Feb. 24, 2025].

[6] T. Dohmen and J. Sauermann, "Referee Bias", Discussion Paper Series, Institute for the Study of Labor (IZA), no. 8857, Feb. 2015. [Online]. Available: https://docs.iza.org/dp8857.pdf [Accessed: Feb. 24, 2025].

[7] B. Lucey and D. Power, "Do Soccer Referees Display Home Bias?", SSRN Electronic Journal, Vol. 552223, Jul. 2009.

[8] B. Buraimo, R. Simmons, M. Maciaszczyk, "Favoritism and Referee Bias in European Soccer: Evidence from the Spanish League and the UEFA Champions League", Contemporary Economic Policy, Vol. 30, no. 3, pp. 329-343, Dec. 2011.

[9] R. Samuel, G. Tenenbaum, and Y. Galily, "An Integrated Conceptual Framework of Decision-Making in Soccer Refereeing", International Journal of Sport and Exercise Psychology, Vol. 19, no. 5, pp. 738-760, Apr. 2020.

[10] R. Boyko, A. Boyko, and M. Boyko, "Referee Bias Contributes to Home Advantage in English Premier- ship Football", Journal of Sports Sciences, Vol. 25, no. 11, pp.1185-1194, Jul. 2007.

[11] S. Stone, "Premier League: 13 VAR mistakes in total in season so far, say chiefs," BBC Sport, Feb. 04, 2025. Available: https://www.bbc.com/sport/football/articles/cd9qvdg57n0o

[12] "Fans sentenced to prison for racist insults directed at soccer star Vinícius Júnior in first-of-its-kind conviction - CBS News," www.cbsnews.com, Jun. 10, 2024. https://www.cbsnews.com/news/vinicius-junior-soccer-fans-sentenced-to-prison-racist-insults-spain/

[13] F. Müller, L. van Zoonen, and L. de Roode, "We can't `Just do it' alone! An analysis of Nike's (potential) contributions to anti-racism in soccer," Media, Culture & Society, vol. 30, no. 1, pp. 23–39, Jan. 2008, doi: https://doi.org/10.1177/0163443707084348.

10. Appendix

10.1. Table A1

Data Points	Data Type	Reason:
Home Team Name	String	Good to know
Away Team Name	String	Good to know
Home Team Score	Int	To determine if bias helped home team
Away Team Score	Int	To determine if bias helped away team
Home Team # of Penalties	Int	To determine if referee gave home team (or favorited team) more penalties on average.
Away Team # of Penalties	Int	To determine if referee gave away team (or favorited team) more penalties on average.
Home Team TimeStamp of Goals	List of Floats/Ints	Are there more yellows/red after scoring late to slow game down.
Away Team TimeStamp of Goals	List of Floats/Ints	Are there more yellows/red after scoring late to slow game down.
Name of Stadium	String	To help determine if there's a bias of a particular team. Maybe this helps with grouping?
Name of City for Stadium	String	To help determine if there's a bias for the city [e.g. Manchester vs Liverpool].
Added Time after 90'	Int	To determine if losing team had extra amount of time compared to normal to equalize or win
Time Game Ended	Int/Float	To determine if game ended when it should [extra time on top of extra time for losing team bias]
# of Home Team Yellow Cards	Int	Good to know
# of Away Team Yellow Cards	Int	Good to know

Yellow Card: - Player Name - Player Nationality - TimeStamp - Foul Type - Explanation	List of Objects - String - String - Int/Float - String - String	Looking for patterns on racial bias. Looking for patterns of when, why [more yellow cards when x team is losing?] to see general bias
# of Home Team Red Cards	Int	Good to know
# of Away Team Red Cards	Int	Good to know
Red Card: - Player Name - Player Nationality - TimeStamp - Foul Type - Explanation - Direct Red - Second Yellow	List of Objects - String - String - Int/Float - String - String - Boolean (T/F) - Boolean (T/F)	Looking for patterns on racial bias. Looking for patterns of when, why [more red cards when x team is losing?] to see general bias Want to know if ref hands out more direct reds vs certain teams, vs second yellows.
HOME - Starting XI + Bench: - {Squad Name, Nationality}	List of Objects - Object with 2 strings	Use it to compute ratio of nationalities vs yellow/red cards given to home team.
AWAY - Starting XI + Bench: - {Squad Name, Nationality}	List of Objects - Object with 2 strings	Use it to compute ratio of nationalities vs yellow/red cards given to home team.
Referee Name	String	Good to know.
Referee Nationality	String	Good to know.
Referee City	String	Good to know.
Date	DateTime	To establish a timeline.
Matchday	Int	To determine if there was more bias towards the end of the season.