Predicting the Market Value of Professional Football Players:

A Multi-Source Stacked Ensemble Approach Using Transfermarkt, FBref, and EA Sports FC Data

Major Project Report

MOD008374

SID2305603

Word Count: 8,546

2025/2026

Table of Contents

1. Introduction

1.1 Background and Context

The global football transfer market has evolved into a multi-billion-dollar industry that demands robust, data-driven decision-making (Poli, Besson and Ravenel, 2022). FIFA's Global Transfer Report recorded international transfer spending exceeding $13 billion in 2025 alone, with over 5000 clubs involved (FIFA, 2025). The sheer magnitude of financial activity in the modern game has transformed player recruitment from a largely informal process into a strategic discipline requiring sophisticated analytical support (Müller, Simons and Weinmann, 2017). In this high-stakes environment, the ability to accurately estimate a player's market worth has become a critical competency for clubs, agents, sporting directors, and financial analysts alike (McHale and Holmes, 2023).

The consequences of inaccurate valuations are severe and well-documented. Overvaluation leads to unsustainable wage bills, inflated transfer fee expenditure, and in extreme cases, financial distress or insolvency (Bell, Brooks and Brooks, 2022). Conversely, undervaluation causes the loss of key assets below their true worth, undermining competitive performance and long-term strategic planning (Majewski, 2016). The problem is compounded by the inherent opacity of the transfer market, where actual negotiated fees are influenced by factors that are not publicly observable, including agent commissions, sell-on clauses, performance bonuses, and the relative bargaining power of the parties involved (Poli, Besson and Ravenel, 2022).

Traditionally, player valuation has relied heavily on expert judgement informed by scouting networks, coaching assessments, and negotiation experience (Frick, 2007). While this approach might capture intangible qualities such as leadership, dressing-room influence, tactical intelligence etc. It is inherently susceptible to cognitive biases, anchoring effects from recent high-profile transactions, and the information asymmetries that characterise a market with limited price transparency (Müller, Simons and Weinmann, 2017). The advent of advanced sports analytics, coupled with the proliferation of publicly available data sources, has created an unprecedented opportunity to complement traditional scouting with quantitative modelling (McHale and Holmes, 2023; Lee, Tama and Cha, 2022).

Three categories of data have emerged as particularly valuable for player valuation research. First, market valuation data from platforms such as Transfermarkt provides community-driven estimates of player worth, updated regularly and widely regarded as a reliable proxy for actual

transfer fees (Müller, Simons and Weinmann, 2017; Boccard, 2022). Second, advanced performance statistics from sources like FBref, powered by StatsBomb, capture granular on-pitch contributions including goals, assists, interceptions, tackles, and per-90-minute normalised statistics that enable fair cross-player comparison (McHale and Holmes, 2023; Jishnu et al., 2022). Third, video game attribute data from the EA Sports FC franchise, accessible via SoFIFA, offers a comprehensive, numerically encoded assessment of player abilities across dozens of skill dimensions, compiled by a global network of over 6,000 data reviewers (Behravan and Razavi, 2021; Sun and Gu, 2024). Table 1.1 summarises these three data sources as used in this project.

Data Source	Records	Features	Coverage	Primary Contribution
Transfermarkt	19,637	23	2009–2026, global	Market valuations, contracts, age
FBref	10,713	31	2023–2025, 36 leagues	On-pitch performance (per-90 stats)
SoFIFA (FC 26)	17,499	76	FC 26 edition, 42 leagues	Holistic ability ratings (34 skills)

Table 1.1: Summary of the three primary data sources used in this project.

1.2 Problem Statement

Despite the growing body of research on football player valuation, several significant gaps persist in the literature. Most existing studies rely on a single data source, limiting the scope of features available for modelling (Shen, 2025). Studies employing only Transfermarkt data cannot capture the nuanced on-pitch performance metrics available through advanced statistics providers (McHale and Holmes, 2023). Conversely, research focusing exclusively on EA Sports attributes may overlook critical market-specific factors such as contract length, player age trajectories, and league prestige (Sun and Gu, 2024). Furthermore, many existing approaches treat valuation as a static, cross-sectional problem using a single regression model, rather than exploiting the complementary structure of heterogeneous data sources through domain-specialised expert models (Shen, 2025; Lee, Tama and Cha, 2022).

This project addresses these limitations by proposing a novel **four-layer stacked ensemble architecture** that assigns a dedicated expert model to each data source. LightGBM for EA data, XGBoost for FBref data and CatBoost for Transfermarkt contextual features to blend their predictions through a rich set of agreement and disagreement features and refine the output through meta-learners. The pipeline is further stabilised through 15 Monte Carlo runs, providing not just point estimates but prediction distributions with associated uncertainty to

address the well-known sensitivity of machine learning results to certain train-test splits (Sulimov, 2024).

1.3 Aims and Objectives

The overarching aim of this project is to develop and evaluate a multi-source machine learning system capable of predicting professional football player market values with high accuracy and quantified uncertainty. The specific objectives are presented in Table 1.2.

Ref	Objective
O1	Collect, clean, and integrate player data from Transfermarkt, FBref, and SoFIFA into a unified analytical dataset (addressing the data integration challenges identified by Poli, Besson and Ravenel, 2022).
O2	Engineer features capturing non-linear relationships, including age-value curves (Frick, 2007), contract dynamics, and positional groupings.
O3	Design and implement a four-layer stacked ensemble with domain-specialised expert models (extending the ensemble approach of Shen, 2025).
O4	Evaluate models using RMSE, MAE, R^2, MAPE, and bracket accuracy metrics (±10%, ±20%).
O5	Interpret model outputs using SHAP feature importance analysis (following Lee, Tama and Cha, 2022).

Table 1.2: Project objectives.

1.4 Scope and Delimitations

The project focuses on outfield and goalkeeper valuations across major European and global leagues. Market valuations from Transfermarkt span 2009 to 2026, encompassing 19,637 unique player records. The SoFIFA dataset covers the FC 26 edition with 17,499 players across 42 leagues, and the FBref dataset encompasses 10,713 player-season records across 36 domestic leagues from the 2023/2024 and 2024/2025 seasons. The target variable is the average of the EA Sports in-game value and the Transfermarkt market valuation where both exist, providing a more robust and less source-biased estimate. The project does not model actual transfer fees, which are influenced by unobservable negotiation dynamics, agent fees, and sell-on clauses (Poli, Besson and Ravenel, 2022), but rather estimates the underlying market worth as a function of observable player characteristics.

2. Literature Review

This chapter critically examines the existing body of research on football player valuation, situating the current project within the broader academic and commercial landscape. The review is organised thematically, addressing: (i) the economic foundations of player valuation; (ii) traditional econometric and statistical approaches; (iii) the application of machine learning techniques; (iv) the use of video game data as a proxy for player ability; (v) advanced performance statistics; (vi) feature engineering and the age-value relationship; (vii) multi-source data integration; and (viii) identified gaps that this project aims to address.

2.1 Economic Foundations of Player Valuation

The valuation of professional football players is fundamentally rooted in labour economics, where a player's worth is theorised as a function of their marginal revenue product i.e. The additional revenue they generate for their club through on-pitch performance, commercial appeal, and brand value (Rottenberg, 1956). Early economic models of the transfer market drew on hedonic pricing theory, which posits that the price of a heterogeneous good can be decomposed into the implicit prices of its constituent attributes (Rosen, 1974). Applied to football, this framework suggests that transfer fees reflect a bundle of observable characteristics including age, playing position, nationality, technical ability, and contractual status (Frick, 2007).

Frick (2007) provided one of the first comprehensive empirical analyses of the determinants of transfer fees in European football, demonstrating that age, experience, goal-scoring records, and the selling and buying club's league status were significant predictors. This foundational work established the core feature set that subsequent studies would refine and extend. Notably, Frick (2007) identified a non-linear relationship between age and transfer value, with players reaching peak valuation in their mid-to-late twenties before declining, this pattern since been robustly confirmed across multiple studies and datasets (Majewski, 2016; van den Berg, 2011; Poli, Besson and Ravenel, 2022).

More recent economic analyses have examined how nationality and league-specific factors influence valuations. Depken and Frei (2023) investigated nationality-based transfer fee premiums in association football, providing evidence that players from certain countries command systematically different fees after controlling for performance characteristics. Bell, Brooks and Brooks (2022) examined whether English football players are overvalued relative

to comparable international players, finding evidence of a significant English player value premium of approximately 40% and a wage premium of approximately 25% in elite European leagues. Van Ours and van Tuijl (2024) explored how different incentive structures across competition formats influence player and team behaviour, with implications for how performance metrics should be interpreted in the context of player valuation. These findings collectively underscore the importance of incorporating contextual and nationality-related features in valuation models, a principle reflected in the feature engineering strategy adopted in this project.

2.2 Traditional Econometric Approaches

The earliest quantitative approaches to player valuation employed classical econometric methods, particularly ordinary least squares (OLS) regression and its variants (Huebl and Swieter, 2002). These models typically regressed log-transformed transfer fees or market valuations on a set of player-level and contextual covariates. Huebl and Swieter (2002) used OLS regression with data from the German Bundesliga and found that goals scored, age, international appearances, and contract duration were significant predictors of transfer fees. The use of log-transformation of the dependent variable became standard practice in the field, reflecting the highly right-skewed distribution of transfer values (Huebl and Swieter, 2002; Müller, Simons and Weinmann, 2017). In the Transfermarkt data used in this project, the mean valuation (€2.44 million) substantially exceeds the median (€350,000), confirming the continued relevance of this transformation strategy. Table 2.1 presents the descriptive statistics.

Statistic	Value (€)
Mean	2,440,744
Median	350,000
Std. Dev.	8,037,198
Min	10,000
Max	200,000,000
25th Percentile	150,000
75th Percentile	1,200,000

Table 2.1: Descriptive statistics for Transfermarkt market valuations (N = 19,637).

Müller, Simons and Weinmann (2017), advanced the econometric literature by applying a multilevel regression framework to a dataset of 4,217 players from the top five European leagues across six seasons. Their analysis confirmed the importance of age, position, and performance metrics, while also highlighting the significance of contextual variables such as

league prestige and club financial power (Müller, Simons and Weinmann, 2017). Critically, they demonstrated that models incorporating both individual and contextual features consistently outperformed those relying on individual statistics alone, providing an important methodological precedent for the multi-source approach adopted in this project.

Poli, Besson and Ravenel (2022), presented an econometric approach to assessing transfer fees and values covering over 2,000 transactions from the five major European football markets between 2012 and 2021. They highlight the importance of contract duration as a key determinant, finding that it explained a substantial portion of fee variation (Poli, Besson and Ravenel, 2022). This finding directly informs the contract-related feature engineering employed in this project, where contract years remaining, free agent status, and the square root of contract years are all included as predictors.

McHale and Holmes (2023) utilised advanced player performance metrics to better capture playing ability and adopted ML algorithms to improve out-of-sample prediction accuracy. Their model demonstrated a considerable improvement on linear regression, and they used it to identify value-for-money transfers, finding that clubs such as Liverpool and Atlético Madrid were consistently successful at identifying undervalued players (McHale and Holmes, 2023).

2.3 Machine Learning Approaches to Player Valuation

The application of machine learning to football player valuation represents a natural evolution from traditional econometric methods, motivated by the ability of ensemble and deep learning algorithms to capture complex, non-linear interactions between features without requiring explicit specification (Lee, Tama and Cha, 2022).

Majewski (2016) conducted a systematic examination of the factors determining the market value of the most valuable players. His econometric models, built on Transfermarkt data for forward-position players, identified team classification points, age, and ‘goodwill’ dummy variables as significant predictors (Majewski, 2016). While his approach was primarily econometric, the study’s emphasis on forward-specific modelling anticipated the position-aware feature engineering strategy adopted in more recent work and in this project.

Behravan and Razavi (2021) proposed a novel machine learning method combining automatic clustering (APSO-clustering) with a hybrid particle swarm optimisation and support vector regression (PSO-SVR) approach for estimating player values using the FIFA 20 dataset. Their

method divided players into four positional clusters and trained separate regression models for each cluster, achieving an accuracy of 74% (Behravan and Razavi, 2021). This positional clustering approach directly informs the `position_group` feature engineering in the current project, where players are classified into GK, DEF, MID, FWD, and UNK categories.

Lee, Tama and Cha (2022) proposed the prediction of football player value using a Bayesian ensemble approach with an improved LightGBM model optimised via Tree-structured Parzen Estimator (TPE). Their study used FIFA data and identified prominent features using SHAP analysis, finding that the optimised LightGBM showed accuracy improvements of approximately 3.8, 1.4, and 1.8 times compared to baseline regression, gradient boosting, and unoptimised LightGBM models respectively in terms of RMSE (Lee, Tama and Cha, 2022). The use of LightGBM with SHAP interpretability in the current project's EA expert model and meta-learner is directly informed by this work.

Shen (2025) applied machine learning techniques and sensitivity analysis to predict football player values based on FIFA and real-world statistical datasets. The study evaluated SVR, Random Forest, XGBoost, and CatBoost models optimised with meta-heuristic algorithms, finding that ensemble models demonstrated exceptional precision with a minimal error of $1.7 million (approximately 10% of average measured value) (Shen, 2025). Critically, Shen (2025) employed Dempster-Shafer Theory to develop ensemble combinations and Fourier amplitude sensitivity testing (FAST) for feature importance, demonstrating that ensemble approaches substantially outperform individual models. Table 2.2 summarises the key ML approaches from the literature.

Study	Best Algorithm	Key Features	R^2 / Accuracy
Majewski (2016)	Econometric / OLS	TM attributes (forwards)	Significant predictors identified
Behravan and Razavi (2021)	PSO-SVR hybrid	FIFA 20 + clustering	74% accuracy
Lee, Tama and Cha (2022)	Optimised LightGBM	FIFA + SHAP	3.8x RMSE improvement
McHale and Holmes (2023)	XGBoost	Advanced perf. metrics	Considerable R^2 improvement
Shen (2025)	Ensemble (XGB+CAT)	FIFA 19 + real-world	$1.7M MAE (10% of mean)
Sun and Gu (2024)	Random Forest	FIFA 19 + TM data	$R^2 > 0.99$ (ensemble)
Sulimov (2024)	ML + media features	Performance + YouTube/Reddit	Transfer fee prediction

Table 2.2: Summary of machine learning approaches to player valuation in the literature.

Sun and Gu (2024) examined the prediction of football player values in the transfer market of well-known European leagues based on FIFA 19 and real-world data. They employed optimised Decision Tree and Random Forest Regression models, achieving over 99% R^2 value and 12% error relative to mean market values with ensemble models (Sun and Gu, 2024). While such high R^2 values should be interpreted cautiously, the study reinforces the general finding that ensemble tree-based methods outperform simpler approaches.

Sulimov (2024) developed an AI-driven approach to football transfer fee prediction that incorporated media activity as a predictor, including YouTube videos, Reddit posts, and other indicators of player popularity. This innovative approach acknowledges that market value is influenced not only by on-pitch performance but also by commercial visibility and public profile (Sulimov, 2024). While the current project does not incorporate media features, this finding underscores the multifaceted nature of player valuation and the potential for future extensions of the work presented here.

2.4 The Role of Video Game Data in Player Valuation

One of the most distinctive features of the contemporary player valuation literature is the use of data from the EA Sports FC (formerly FIFA) video game franchise as a proxy for player ability. The games assign numerical ratings to each player across dozens of skill dimensions based on assessments by a global network of over 6,000 data reviewers, including professional scouts, coaches, and analysts (Behravan and Razavi, 2021). These ratings are updated regularly and have been shown to correlate strongly with real-world performance metrics (Lee, Tama and Cha, 2022; Shen, 2025).

Sun and Gu (2024) demonstrated the predictive value of FIFA 19 data when combined with real-world statistics from well-known European leagues, finding that Random Forest models trained on the combined feature set achieved high predictive accuracy. Their analysis confirmed that EA Sports' overall rating was among the most important features, often ranking above traditional performance statistics in feature importance analyses (Sun and Gu, 2024). Behravan and Razavi (2021) similarly used FIFA 20 data as the primary feature source, leveraging the granular skill ratings to build positionally-stratified prediction models.

Jishnu et al. (2022) specifically examined football player transfer value prediction using advanced statistics from FBref and FIFA 22 data from Transfermarkt. Their results showed that Gradient Boosting and XGBoost were the best-performing algorithms when combining

these data sources (Jishnu et al., 2022). This study directly anticipated the multi-source approach adopted in the current project, demonstrating that combining EA Sports attributes with real-world performance statistics yields superior predictive accuracy compared to either source in isolation.

The SoFIFA dataset used in this project contains 76 features per player, plus goalkeeping-specific attributes. This provides a substantially richer feature set than many earlier studies which relied on aggregate ratings or a subset of attributes (Behravan and Razavi, 2021; Lee, Tama and Cha, 2022).

2.5 Advanced Performance Statistics and FBref

The emergence of advanced performance statistics, driven by tracking technology and event data providers such as StatsBomb, Opta, and Wyscout, has fundamentally altered the landscape of football analytics (McHale and Holmes, 2023). FBref, which serves as a publicly accessible interface for StatsBomb data, provides metrics that go far beyond traditional box-score statistics, including per-90-minute normalised statistics that facilitate fair comparison across players with different playing time profiles (McHale and Holmes, 2023; Jishnu et al., 2022).

McHale and Holmes (2023) demonstrated that advanced performance metrics provided a substantial improvement in transfer fee prediction accuracy compared to models using only basic statistics. Their study utilised metrics including progressive passes, progressive carries, and expected goals contribution, finding that these advanced features captured dimensions of player quality invisible to traditional statistics (McHale and Holmes, 2023). The FBref dataset used in this project incorporates a comprehensive range of per-90 metrics including goals per 90 (Gls/90), shots per 90 (Sh/90), shot on target percentage (SoT%), interceptions per 90 (Int/90), and tackles won per 90 (TklW/90), as well as engineered ratio features such as goal contribution per 90 and start ratio.

Jishnu et al. (2022) combined FBref performance metrics with FIFA 22 attributes and Transfermarkt valuations, finding that models trained on the combined feature set outperformed single-source models. Critically, their feature importance analysis revealed that EA Sports attributes and age-related features were the dominant predictors, with FBref performance metrics providing incremental but meaningful improvements, particularly for players in lower leagues where EA Sports ratings may be less precise (Jishnu et al., 2022). This

finding has direct implications for the expert weighting strategy in the current project, where the TM expert receives the highest blend weight (0.45) and the FBref expert the lowest (0.25).

2.6 Feature Engineering and the Age-Value Relationship

A recurring theme across the literature is the critical importance of feature engineering in achieving high predictive accuracy (Müller, Simons and Weinmann, 2017; Shen, 2025). The relationship between age and market value is perhaps the most extensively studied non-linearity in football economics. Multiple studies have confirmed that player valuations follow an inverted-U shaped curve, peaking around age 25–27 and declining thereafter (Frick, 2007; Majewski, 2016; Poli, Besson and Ravenel, 2022). This pattern reflects the interaction between improving skill development in youth, peak physical performance in the mid-twenties, and declining physical attributes combined with shorter remaining career horizons for older players (van den Berg, 2011).

Van den Berg (2011) conducted a detailed analysis of age-value curves across different positions and leagues, finding that the peak valuation age varies by position: forwards and attacking midfielders tend to peak earlier (around age 25–26), while central defenders and goalkeepers retain value longer, often peaking around age 27–29 (van den Berg, 2011). The Transfermarkt dataset used in this project incorporates several engineered age-related features directly motivated by these findings, including age-squared, years from age 27, absolute years from age 27, and categorical age bands (teen, young, prime, late prime, veteran, elder).

Contract status has also been identified as a significant determinant of market value across multiple studies (Poli, Besson and Ravenel, 2022; Müller, Simons and Weinmann, 2017). Players nearing contract expiry are typically valued lower, as buying clubs recognise the possibility of a free transfer (Poli, Besson and Ravenel, 2022). The Transfermarkt dataset includes engineered contract features such as contract years remaining, a free agent indicator, short contract and long contract binary flags, and the square root of contract years to capture the diminishing marginal returns of additional contract years.

Feature	Type	Rationale
age_sq	Numeric	Quadratic component of age-value curve (Frick, 2007)
years_from_27	Numeric	Distance from peak valuation age (van den Berg, 2011)
abs_years_from_27	Numeric	Symmetric deviation from peak age
age_band	Categorical	Six groups: teen, young, prime, late_prime, veteran, elder

Feature	Type	Rationale
is_free_agent	Binary	Zero remaining contract years (Poli, Besson and Ravenel, 2022)
short_contract / long_contract	Binary	Bottom/top quintile of contract duration
contract_sqrt	Numeric	Diminishing marginal returns of extra contract years
years_since_2012	Numeric	Transfer market inflation over time (Boccard, 2022)

Table 2.3: Engineered features in the Transfermarkt dataset with literature justification.

2.7 Multi-Source Data Integration

While individual data sources have been extensively studied, the systematic integration of multiple heterogeneous data sources for player valuation remains a relatively underexplored area in the literature (Shen, 2025). The majority of existing studies rely on a single primary dataset, supplemented by basic biographical information. Fewer studies have attempted to systematically combine market data, advanced performance statistics, and game-based ability ratings into a unified modelling framework (Jishnu et al., 2022).

Jishnu et al. (2022) provided the most direct precedent for the current project by combining FBref, FIFA 22, and Transfermarkt data. However, their approach used a single model trained on the pooled feature set, rather than the domain-specialised expert architecture proposed here. Shen (2025) demonstrated the value of ensemble approaches that combine multiple models, each potentially capturing different aspects of the data, and showed that such ensembles consistently outperform individual models. The current project extends this principle by assigning each expert model a distinct data domain, enabling specialised learning of within-domain feature interactions before combining predictions at the meta-learning stage.

Challenges in multi-source integration include entity resolution (matching players across datasets with different naming conventions and identifiers), temporal alignment (ensuring performance statistics correspond to the same period as market valuations), and feature redundancy (managing collinearity between overlapping features from different sources) (Müller, Simons and Weinmann, 2017).

2.8 Identified Gaps and Contribution of This Project

The literature review reveals several gaps that this project directly addresses. First, while individual studies have demonstrated the value of Transfermarkt data (Müller, Simons and Weinmann, 2017; Poli, Besson and Ravenel, 2022), FBref statistics (McHale and Holmes,

2023), and EA Sports attributes (Behravan and Razavi, 2021; Lee, Tama and Cha, 2022), very few have systematically integrated all three into a single predictive framework (Jishnu et al., 2022 being a notable exception). Second, existing research typically pools all features into a single model, rather than assigning domain-specialised experts that can learn optimal within-domain interactions, an approach that fails to exploit the complementary structure of heterogeneous data sources (Shen, 2025). Third, the use of stacked ensemble architectures with both boosted and neural meta-learners remains rare in the football valuation domain, despite demonstrating superior performance in other tabular regression tasks (Shen, 2025). Fourth, point estimates without uncertainty quantification provide incomplete information for decision-makers; the repeated-run approach adopted here addresses this gap. Fifth, most studies focus narrowly on the top five European leagues (Müller, Simons and Weinmann, 2017; Poli, Besson and Ravenel, 2022), whereas this project incorporates data from over 40 leagues across multiple continents, providing a more diverse and generalisable dataset.

In summary, this project contributes to the literature by providing one of the most comprehensive multi-source player valuation studies to date, combining the breadth of Transfermarkt's market data, the analytical depth of FBref's advanced statistics, and the holistic ability assessment provided by EA Sports FC 26 (Behravan and Razavi, 2021), within a novel four-layer stacked ensemble architecture (extending Shen, 2025) with uncertainty quantification.

3. Methodology and Implementation

This chapter details the methodological framework underpinning the project, encompassing the research philosophy, data collection procedures, data cleaning and integration pipeline, feature engineering strategy, and the four-layer stacked ensemble modelling architecture. The methodology is designed to ensure reproducibility and rigour, with all processing implemented in Python (Van Rossum and Drake, 2009).

3.1 Research Philosophy and Approach

This project adopts a positivist research philosophy, grounded in the assumption that player market values are determined by observable, measurable factors amenable to quantitative modelling (Saunders, Lewis and Thornhill, 2019). The approach is deductive and empirical: hypotheses derived from the existing literature regarding the influence of age (Frick, 2007), performance (McHale and Holmes, 2023), ability ratings (Lee, Tama and Cha, 2022), and contractual status (Poli, Besson and Ravenel, 2022) on valuation are tested through the systematic application of machine learning regression models to a large, multi-source dataset. The use of multiple data sources and model architectures provides methodological triangulation, strengthening the validity and reliability of the findings (Creswell and Creswell, 2018).

3.2 Data Sources and Collection

3.2.1 Transfermarkt. Transfermarkt provides community-driven market value estimates, moderated by a central editorial team and extensively validated in academic research as a reliable proxy for actual transfer fees (Müller, Simons and Weinmann, 2017; Poli, Besson and Ravenel, 2022). The dataset contains 19,637 player records with biographical, contractual, and valuation information spanning 2009–2026. Two tables were merged on `player_id`: a player master table and a valuation history table. The merge included key validation, deduplication to retain the most recent valuation per player, and removal of records missing critical fields:

```
# Clean IDs, reduce valuation duplicates to one row per player_id, then merge
df_1_merge = df_1.copy()
df_2_merge = df_2.copy()

for frame in (df_1_merge, df_2_merge):
    frame[join_key] = pd.to_numeric(frame[join_key], errors='coerce').astype('Int64')

df_1_merge = df_1_merge.dropna(subset=[join_key]).copy()
df_2_merge = df_2_merge.dropna(subset=[join_key]).copy()
```

```
if 'date' in df_2_merge.columns:
    df_2_merge['date'] = pd.to_datetime(df_2_merge['date'], errors='coerce')
    df_2_latest = (
        df_2_merge
        .sort_values([join_key, 'date'])
        .drop_duplicates(subset=[join_key], keep='last')
    )
else:
    df_2_latest = df_2_merge.drop_duplicates(subset=[join_key], keep='last')

df_merged = df_1_merge.merge(
    df_2_latest,
    on=join_key,
    how='left',
    suffixes=('_player', '_valuation')
)
```

3.2.2 FBref. FBref provides advanced performance metrics powered by StatsBomb (McHale and Holmes, 2023). The dataset comprises 10,713 player-season records across 36 leagues from the 2023–2024 and 2024–2025 seasons. For players appearing in both seasons, numeric statistics were averaged while retaining the most recent categorical information, following the approach of aggregating multi-season data to reduce noise (Müller, Simons and Weinmann, 2017):

```
# Separate players into three groups
players_df1 = set(df_1[key_col].unique())
players_df2 = set(df_2[key_col].unique())

common_players = players_df1 & players_df2
only_df1 = players_df1 - players_df2
only_df2 = players_df2 - players_df1

print(f"\nPlayers in both dataframes: {len(common_players)}")
print(f"Players only in df_1: {len(only_df1)}")
print(f"Players only in df_2: {len(only_df2)}")

# For common players: average shared numeric columns, keep non-numeric from df_1
# First, remove duplicates by keeping the first occurrence of each player
common_df1 = df_1[df_1[key_col].isin(common_players)].drop_duplicates(subset=[key_col],
keep='first').copy()
common_df2 = df_2[df_2[key_col].isin(common_players)].drop_duplicates(subset=[key_col],
keep='first').copy()

# Build base from non-numeric columns in df_1
merged_common = common_df1[[key_col] + non_numeric_cols].copy()

# Set index so assignments align by player
common_df1_idx = common_df1.set_index(key_col)
common_df2_idx = common_df2.set_index(key_col)
merged_common = merged_common.set_index(key_col)

# Average shared numeric columns
```

```
for col in shared_numeric_cols:
    merged_common[col] = (common_df1_idx[col] + common_df2_idx[col]) / 2

# Keep season-specific numeric columns that only exist in one dataframe
for col in df1_only_numeric_cols:
    merged_common[col] = common_df1_idx[col]

for col in df2_only_numeric_cols:
    merged_common[col] = common_df2_idx[col]

# Restore key column as regular column
merged_common = merged_common.reset_index()

# Get lone entries from df_1 and df_2
lone_df1 = df_1[df_1[key_col].isin(only_df1)].copy()
lone_df2 = df_2[df_2[key_col].isin(only_df2)].copy()

# Combine all three groups
df_merged = pd.concat([merged_common, lone_df1, lone_df2], ignore_index=True, sort=False)

print(f"\nMerged dataframe shape: {df_merged.shape}")
print(f"Total unique players: {df_merged[key_col].nunique()}")
```

3.2.3 SoFIFA (EA Sports FC 26). The SoFIFA dataset was collected through an asynchronous web scraper implemented in Python using Playwright, a headless browser automation framework necessary for SoFIFA's dynamically-rendered JavaScript content. The scraper employed parallel execution with five concurrent browser contexts. This was customised from the original scraper maintained by Prashant Ghimire on Github (prashantghimire, 2025).

The resulting dataset contains 17,499 unique player records with 76 features, including 34 granular skill attributes. Table 3.1 summarises the EA Sports attribute categories (Behravan and Razavi, 2021; Lee, Tama and Cha, 2022).

Category	Count	Example Features
Attacking	5	crossing, finishing, heading_accuracy, short_passing, volleys
Skill	5	dribbling, curve, fk_accuracy, long_passing, ball_control
Movement	5	acceleration, sprint_speed, agility, reactions, balance
Power	5	shot_power, jumping, stamina, strength, long_shots
Mentality	6	aggression, interceptions, vision, penalties, composure
Defending	3	defensive_awareness, standing_tackle, sliding_tackle
Goalkeeping	5	gk_diving, gk_handling, gk_kicking, gk_positioning, gk_reflexes

Table 3.1: EA Sports FC 26 attribute categories (Behravan and Razavi, 2021).

3.3 Data Cleaning and Preprocessing

Each dataset underwent cleaning following best practices for data preparation in ML workflows (Pedregosa et al., 2011). The Transfermarkt pipeline merged master and valuation tables, dropped 15+ extraneous columns, removed records with missing critical fields, converted date columns to datetime format, and computed derived features. The FBref pipeline loaded separate season files, dropped redundant columns, cleaned nationality codes via regex extraction, and averaged numeric statistics across seasons. The SoFIFA pipeline converted numeric types to double-precision, dropped non-analytical columns, deduplicated by `player_id`, and mean-imputed missing goalkeeping attributes for outfield players (Behravan and Razavi, 2021).

3.4 Feature Engineering

Feature engineering transforms raw variables into representations that better capture the underlying relationships between player characteristics and market value (Shen, 2025). These are presented in Table 3.2 with their derivation logic and literature justification.

Feature	Source	Derivation
age_ea / age_ea_sq	SoFIFA	Age from DOB relative to 2025-01-01; squared term (Frick, 2007)
contract_years_left_ea	SoFIFA	Contract end minus reference date (Poli, Besson and Ravenel, 2022)
bmi_like	SoFIFA	weight_kg / (height_cm / 100)²
wage_per_rating	SoFIFA	wage / overall_rating
goal_contrib_per90	FBref	(Gls + Ast) / 90s played (McHale and Holmes, 2023)
start_ratio / sub_ratio	FBref	Starts/MP and Subs/MP
minutes_per_match	FBref	Min / MP
position_group	All	GK/DEF/MID/FWD/UNK (Behravan and Razavi, 2021)

Table 3.2: Engineered features constructed during the modelling pipeline.

3.5 Data Integration Strategy

Integrating three independently-sourced datasets presented several challenges identified in the literature, including inconsistent player naming conventions, heterogeneous identifier systems, and differing temporal coverage (Müller, Simons and Weinmann, 2017). The integration strategy employed a multi-step process. Player names were standardised through Unicode NFKD normalisation, ASCII conversion, lowercasing, and removal of non-alphanumeric characters:

```
def normalize_name(name):
```

```
    if pd.isna(name):
        return np.nan
    name = str(name)
    name = unicodedata.normalize("NFKD", name).encode("ascii", "ignore").decode("ascii")
    name = name.lower().strip()
    name = re.sub(r"[^a-z0-9\s]", "", name)
    name = re.sub(r"\s+", " ", name).strip()
    return name
```

The EA and Transfermarkt datasets were joined on the composite key of normalised name and date of birth, ensuring precise entity resolution. FBref data was left-joined on normalised name, with deduplication by highest minutes played. A binary `has_fbref` indicator flags players with matched performance data, enabling the meta-learner to learn source-availability effects.

3.6 Target Variable Construction

The target variable is constructed as the average of the EA Sports in-game valuation and the Transfermarkt market valuation where both exist, following the principle that averaging multiple estimates reduces source-specific biases (Müller, Simons and Weinmann, 2017). A natural logarithmic transformation (`log1p`) is applied to reduce the extreme right-skewness of the distribution, following the standard practice established by Huebl and Swieter (2002) and widely adopted in subsequent studies (Müller, Simons and Weinmann, 2017; Poli, Besson and Ravenel, 2022):

```
def add_target(df):
    """
    Target = average(EA value, TM value) when both exist.
    """
    df = df.copy()

    ea_val = pd.to_numeric(df["value"], errors="coerce")
    tm_val = pd.to_numeric(df["market_value_in_eur_valuation"], errors="coerce")

    df["target_value"] = np.where(
        ea_val.notna() & tm_val.notna(),
        (ea_val + tm_val) / 2.0,
        np.where(ea_val.notna(), ea_val, tm_val)
    )

    df = df[df["target_value"].notna() & (df["target_value"] > 0)].copy()
    df["log_target"] = np.log1p(df["target_value"])
    return df
```

Leakage prevention is critical in machine learning pipelines (Pedregosa et al., 2011). The columns `value`, `market_value_in_eur_valuation`, and `release_clause` are explicitly

excluded from all predictor feature sets to prevent the model from trivially accessing the target variable.

3.7 Model Architecture: Four-Layer Stacked Ensemble

The core methodological contribution of this project is a **four-layer stacked ensemble architecture** designed to exploit the complementary structure of the three data sources. This architecture extends the ensemble approach advocated by Shen (2025) and the multi-source strategy of Jishnu et al. (2022), while introducing domain-specialised expert models and a dual meta-learner design. Table 3.3 provides the architectural overview.

Layer	Component	Algorithm	Input	Output
1 (Experts)	EA Expert	LightGBM (600 trees)	48 EA/SoFIFA features	OOF + test predictions
1 (Experts)	FBref Expert	XGBoost (600 trees)	30 FBref features	OOF + test predictions
1 (Experts)	TM Expert	CatBoost (700 iters)	12 TM/context features	OOF + test predictions
2 (Blend)	Expert Blender	Weighted mean + stats	3 expert predictions	13 blended features
3A (Meta)	Boosted Meta	LightGBM (700 trees)	Blended + context feats	Meta prediction
3B (Meta)	Neural Meta	MLP (256-128-64)	Blended + context feats	Meta prediction
4 (Final)	Inverse-RMSE	Weighted average	Two meta predictions	Final prediction

Table 3.3: Four-layer stacked ensemble architecture.

3.7.1 Layer 1: Domain-Specialised Expert Models

The first layer trains three gradient boosting models, each specialised on a distinct feature domain. This design is motivated by the observation that different data sources capture fundamentally different aspects of player value (Jishnu et al., 2022), and that domain-specialised models can learn optimal feature interactions within each domain more effectively than a single model pooling all features (Shen, 2025).

EA Expert (LightGBM). LightGBM (Ke et al., 2017) is selected for its efficiency with high-cardinality categorical features and its leaf-wise tree growth strategy, which is particularly effective for the 48 EA/SoFIFA features including 34 granular skill attributes. Following the hyperparameter recommendations of Lee, Tama and Cha (2022), the model uses 600 trees with a learning rate of 0.03, maximum depth of 6, and 31 leaves:

FBref Expert (XGBoost). XGBoost (Chen and Guestrin, 2016) is selected for its robust native handling of missing values, which is common in FBref data where players may lack certain statistics (McHale and Holmes, 2023). The model operates on 30 FBref features with matched hyperparameters:

TM Expert (CatBoost). CatBoost (Prokhorenkova et al., 2018) is selected for its native handling of categorical features without explicit one-hot encoding, which is advantageous for high-cardinality features like club name, nationality, and sub-position in the 12-feature Transfermarkt set.

Each expert is trained using 5-fold cross-validated out-of-fold (OOF) prediction generation, a critical technique for preventing data leakage in stacked ensembles (Wolpert, 1992). For each fold, the model trains on 4 folds and predicts the held-out fold; test set predictions are averaged across all 5 fold models:

3.7.2 Layer 2: Expert Blending Layer

The second layer transforms the three expert predictions into 13 blended features that capture central tendency and *inter-expert disagreement*. The disagreement features are particularly informative: when experts trained on different data domains disagree strongly, it signals an unusual player profile or that one data source provides a divergent signal (Shen, 2025). Table 3.4 details these features.

Feature	Formula	Purpose
pred_mean	mean(EA, FB, TM)	Unweighted central tendency
pred_median	median(EA, FB, TM)	Robust central tendency
pred_weighted	0.30×EA + 0.25×FB + 0.45×TM	Prior-weighted blend
ea_fb_absdiff	|EA − FB|	EA vs. performance disagreement
ea_tm_absdiff	|EA − TM|	EA vs. market disagreement
fb_tm_absdiff	|FB − TM|	Performance vs. market disagreement
pred_std	std(EA, FB, TM)	Overall expert uncertainty
pred_max / pred_min	max/min	Bounds of expert range
pred_range	max − min	Spread of expert predictions

Table 3.4: Blended features computed from expert predictions in Layer 2.

3.7.3 Layer 3: Meta-Learner Models

The third layer trains two distinct meta-learners on the combined feature set: the 13 blended features, original contextual features, and the `has_fbref` indicator. This dual meta-learner design provides model diversity (Shen, 2025).

Layer 3A: Boosted Meta-Model (LightGBM). A LightGBM model with 700 trees and learning rate 0.025 learns non-linear interactions between expert predictions and contextual features (Ke et al., 2017).

Layer 3B: Neural Meta-Model (MLP Regressor). A multi-layer perceptron with architecture (256, 128, 64), ReLU activation, Adam optimiser (Kingma and Ba, 2015), and early stopping provides a complementary non-linear perspective (Pedregosa et al., 2011):

3.7.4 Layer 4: Inverse-RMSE Final Blend

The final layer combines meta-model predictions using **inverse-RMSE weighting** on an internal validation set. This data-adaptive approach automatically assigns higher weight to whichever meta-model performs better on unseen data for each particular run, following the inverse-error weighting principle used in ensemble combination (Shen, 2025):

3.8 Monte Carlo Repeated Experimentation

A key methodological contribution is the use of **15 Monte Carlo repeated experimentation** to address the well-known sensitivity of machine learning results to particular train-test partitions (Sulimov, 2024). The entire pipeline is executed 15 times with different random seeds (base seed 42 + run number), producing: (i) a distribution of evaluation metrics enabling mean ± std reporting; (ii) multiple predictions per player from which mean, median, and standard deviation are computed; and (iii) robustness against any single random split:

3.9 Evaluation Metrics

Model performance is assessed using a comprehensive set of metrics applied to back-transformed (euro-denominated) predictions, following the evaluation framework established in the literature (Shen, 2025; Lee, Tama and Cha, 2022). Table 3.5 defines each metric.

Metric	Description	Interpretation
RMSE	$\sqrt{(\text{mean}((y - \hat{y})^2))}$	Penalises large errors (Müller, Simons and Weinmann, 2017)
MAE	$\text{mean}(\lvert y - \hat{y} \rvert)$	Average absolute error (Shen, 2025)

Metric	Description	Interpretation
R^2	1 - SS_res / SS_tot	Variance explained (Lee, Tama and Cha, 2022)
MAPE	mean(|y - ŷ|/y) × 100	Scale-independent error (Shen, 2025)
Accuracy@10%	% within ±10%	Tight bracket (Shen, 2025: 10% threshold)
Accuracy@20%	% within ±20%	Relaxed bracket accuracy
Mean % Error	mean((y-ŷ)/y) × 100	Directional bias detection

Table 3.5: Evaluation metrics with literature precedent.

3.10 Feature Importance and Interpretability

SHAP (SHapley Additive exPlanations) values (Lundberg and Lee, 2017) are computed for the boosted meta-model on 500 training observations per run, providing both global and local interpretability. Mean absolute SHAP values are aggregated across all 15 runs to produce stable feature importance rankings, following the SHAP-based approach employed by Lee, Tama and Cha (2022) and Shen (2025):

3.11 Configuration Summary

Parameter	Value	Rationale / Source
Test set size	20%	Standard hold-out (Pedregosa et al., 2011)
CV folds	5	Bias-variance trade-off (Pedregosa et al., 2011)
Monte Carlo runs	15	Stable uncertainty estimates
EA expert weight	0.30	EA captures broad ability (Lee, Tama and Cha, 2022)
FBref expert weight	0.25	FBref: incremental improvements (Jishnu et al., 2022)
TM expert weight	0.45	TM most directly reflects market (Müller et al., 2017)
Learning rate (all)	0.03	Slow learning (Lee, Tama and Cha, 2022)
MLP layers	(256, 128, 64)	Tapering architecture
Final blend	Inverse-RMSE	Data-adaptive weighting (Shen, 2025)

Table 3.6: Configuration summary with literature justification.

3.12 Data Splitting and Validation Strategy

The dataset is partitioned into training (80%) and test (20%) sets using random sampling with a fixed seed for reproducibility (Pedregosa et al., 2011). Within the training set, a further 5-fold cross-validation strategy is employed for the Layer 1 expert models to generate out-of-fold predictions, following the stacked generalisation framework established by Wolpert

(1992). This nested cross-validation approach ensures that at no point does any model in the ensemble have access to data used to evaluate its performance, preventing the optimistic bias that would result from in-sample evaluation (Pedregosa et al., 2011).

For the Layer 3 meta-learners, an additional internal validation split of 20% is carved from the training data. This validation set serves two purposes: it enables early stopping for the neural meta-model, preventing overfitting to the training data (Kingma and Ba, 2015), and it provides the validation RMSE values used to compute the inverse-RMSE weights in Layer 4 (Shen, 2025). The internal split is performed with a different random seed for each Monte Carlo run, ensuring that the validation set composition varies across experiments and that no single validation partition disproportionately influences the final results (Sulimov, 2024).

The 15-run Monte Carlo strategy provides a robust estimate of model performance that accounts for the stochastic variability inherent in the train-test splitting process, the random initialisation of neural network weights, and the randomised row and column subsampling used by gradient boosting algorithms (Lee, Tama and Cha, 2022). By aggregating predictions across 15 independent runs, the final player-level estimates exhibit lower variance and higher reliability than those from any single run, effectively implementing a form of model averaging that has been shown to improve generalisation performance in the machine learning literature (Chen and Guestrin, 2016).

3.13 Handling of Missing Data and Categorical Features

Missing values in numeric features are imputed using the median value computed exclusively on the training set to prevent information leakage from the test set (Pedregosa et al., 2011). Median imputation is preferred over mean imputation due to its robustness to the right-skewed distributions present in several features, particularly financial variables such as wage and the target value itself (Müller, Simons and Weinmann, 2017). Categorical features including position, nationality, league, and club are encoded using one-hot encoding via `pd.get_dummies` with explicit handling of missing values as a separate category (`dummy_na=True`). This approach ensures that missingness itself can serve as an informative signal. For example, a player with missing FBref data may systematically differ from one with complete data, and this distinction should be learnable by the model (McHale and Holmes, 2023).

Feature names are sanitised through a dedicated cleaning function that replaces special characters (percentage signs, slashes, hyphens, plus signs) with descriptive text, removes non-alphanumeric characters, and collapses repeated underscores. This preprocessing step is necessary because certain gradient boosting implementations, particularly LightGBM, require column names free of special characters (Ke et al., 2017). The sanitisation is applied consistently across training and test sets to ensure column alignment throughout the pipeline.

For the neural meta-model specifically, all features are standardised using z-score normalisation (subtracting the mean and dividing by the standard deviation) via scikit-learn's StandardScaler (Pedregosa et al., 2011). The scaler is fitted exclusively on the training data and applied to both training and test data, preventing any leakage of test-set statistics into the model. This standardisation step is critical for neural network convergence, as the Adam optimiser (Kingma and Ba, 2015) performs best when input features are on a comparable scale.

3.14 Ethical Considerations

All data used is publicly available and does not contain personally sensitive information beyond what players, clubs and other organisations have disclosed through their professional careers. The SoFIFA scraping employed rate-limited, asynchronous requests with stealth techniques to minimise server load, following ethical web scraping guidelines. Player valuations are analysed in aggregate for academic research purposes and are not intended to influence actual transfer negotiations. The project acknowledges potential biases in the underlying data: Transfermarkt valuations may reflect media coverage and European-centric perspectives (Bell, Brooks and Brooks, 2022), while EA Sports ratings may lag behind real-world form changes (Lee, Tama and Cha, 2022). These biases are mitigated through multi-source integration and the averaging of EA and Transfermarkt values as the target variable.

3.15 Tools and Technologies

Category	Technology	Reference
Language	Python 3.x	Van Rossum and Drake (2009)
Data processing	pandas, NumPy	McKinney (2010); Harris et al. (2020)
ML framework	scikit-learn	Pedregosa et al. (2011)
Gradient boosting	LightGBM, XGBoost, CatBoost	Ke et al. (2017); Chen and Guestrin (2016); Prokhorenkova et al. (2018)
Interpretability	SHAP	Lundberg and Lee (2017)
Web scraping	Playwright, asyncio	

Category	Technology	Reference
Visualisation	Matplotlib, Seaborn	Hunter (2007)

Table 3.7: Tools and technologies with references.

4. Results and Evaluation

This chapter presents the empirical results of the four-layer stacked ensemble, evaluated across the 15-run Monte Carlo experiment. Section 4.1 reports overall predictive performance; Section 4.2 benchmarks these results against the most directly comparable studies; Section 4.3 examines prediction distributions and uncertainty; Section 4.4 interprets global feature importance via SHAP; Section 4.5 analyses meta-learner weighting; and Sections 4.6–4.7 investigate positional patterns and residual structure.

4.1 Overall Model Performance

Table 4.1 summarises the seven headline evaluation metrics, aggregated across 15 independent pipeline runs (1,240 held-out test players per run; 5,986 unique test predictions overall after aggregation across runs).

Metric	Mean	Std	Min	Max
RMSE (€)	2,487,879	456,472	1,853,665	3,360,709
MAE (€)	1,003,490	113,330	867,256	1,246,630
R^2	0.9469	0.0191	0.9036	0.9706
MAPE (%)	23.24	1.12	21.93	26.64
Accuracy@10%	29.61	1.74	25.65	31.69
Accuracy@20%	54.47	2.48	47.50	57.90
Mean Percentage Error (%)	+3.50	2.27	–0.46	+7.11

Table 4.1: Aggregate performance metrics across 15 Monte Carlo runs.

The model explains 94.7% of the variance in player market value on held-out test data (R^2 = 0.9469), with a mean absolute error of approximately €1.0 million and a root-mean-squared error of €2.49 million. On scale-independent measures, the mean absolute percentage error is 23.2%, and just over half of all predictions (54.5%) fall within ±20% of the true valuation. The mean percentage error of +3.5% suggests a mild but consistent tendency to under-predict and that the model produces fractionally conservative valuations on average, which is a preferable

failure mode to systematic over-prediction from a financial-risk perspective (Bell, Brooks and Brooks, 2022).

Critically, performance variation across runs is small relative to mean performance: the coefficient of variation for R^2 is only 2.0%, and for MAPE it is 4.8%. This demonstrates that the ensemble's performance is not an artefact of a particularly favourable random split, directly addressing the sensitivity concern raised by Sulimov (2024).

4.2 Comparison with Prior Studies

Table 4.2 positions the present results against four of the most relevant prior studies. Direct head-to-head comparison is inherently imperfect because the target variable, data vintage, and evaluation partitions differ across papers, but R^2 provides a consistent scale-free reference point.

Study	Primary data source	Best model	Best R^2
Behravan and Razavi (2021)	FIFA 20	PSO-SVR	0.74
Jishnu et al. (2022)	FBref + FIFA 22 + Transfermarkt	Gradient Boosting	0.82–0.86
McHale and Holmes (2023)	StatsBomb + transfer fees	xgbTree	0.77
Shen (2025)	FIFA 19 + real-world stats	RSCX_SC hybrid	0.99
This project	**TM + FBref + SoFIFA (FC 26)**	**4-layer stack (15-run)**	**0.947**

Table 4.2: Best reported R^2 values from directly comparable studies.

Behravan and Razavi (2021) reported that their PSO-SVR approach achieved approximately 74% accuracy on FIFA 20 data. Jishnu et al. (2022), the most methodologically similar prior study in that it also combined FBref, FIFA and Transfermarkt data, reported a best R^2 of 0.82 overall and up to 0.86 for Gradient Boosting applied to midfielders.

McHale and Holmes (2023), working on the noisier target of actual transfer fees rather than market valuations, reported an xgbTree R^2 of 0.77 with MAPE of 67.5%. Shen (2025) achieved the highest R^2 in the published literature at 0.9931 with the RSCX_SC hybrid ensemble optimised by meta-heuristic search.

The present model's R^2 of 0.947 substantially exceeds the non-hybrid benchmarks and approaches the state of the art results reported by Shen (2025), without requiring the

computationally expensive meta-heuristic optimisation that Shen employs. The 23% MAPE achieved here is also considerably tighter than the 67% reported by McHale and Holmes (2023), reflecting the reduced noise of market valuations relative to negotiated transfer fees

4.3 Prediction Distributions and Uncertainty

A distinctive feature of the Monte Carlo framework is that each of the 5,986 unique held-out players receives 15 independent predictions, from which a mean and standard deviation are derived:

The mean prediction standard deviation is €386,000, which corresponds to approximately 7.3% of the mean predicted value, a tight uncertainty band relative to the 23% MAPE. This gap between cross-run stability and absolute error indicates that the residual error is dominated by systematic factors rather than stochastic variability in training (Sulimov, 2024).

Uncertainty is not uniformly distributed. Players with higher predicted values exhibit substantially larger absolute standard deviations: the top-decile predicted players have a mean standard deviation of approximately €1.5 million, against less than €100,000 for players in the bottom decile. This heteroscedastic pattern is expected and broadly mirrors the right-skewness of the target distribution, consistent with the observation that high-value transactions involve greater price elasticity (Poli, Besson and Ravenel, 2022).

4.4 Feature Importance via SHAP

Mean absolute SHAP values (Lundberg and Lee, 2017), aggregated across 500 training observations per run over all 15 runs, provide a stable ranking of feature contributions to the boosted meta-model. Table 4.3 lists the top ten features.

Rank	Feature	Mean SHAP	Origin
1	ea_pred	0.821	Layer 1 EA Expert prediction
2	pred_mean	0.070	Blended (Layer 2)
3	pred_max	0.044	Blended (Layer 2)
4	potential	0.035	SoFIFA
5	wage_per_rating	0.028	Engineered (wage ÷ rating)
6	overall_rating	0.028	SoFIFA
7	age_ea	0.027	SoFIFA
8	pred_min	0.022	Blended (Layer 2)

9	age_at_valuation	0.020	Transfermarkt
10	pred_median	0.019	Blended (Layer 2)

Table 4.3: Top ten features by mean absolute SHAP value, averaged across 15 runs.

Three observations emerge. First, the EA expert prediction (ea_pred) alone accounts for a mean SHAP value by more than an order of magnitude greater than any other feature. This confirms the hypothesis of Behravan and Razavi (2021) and Lee, Tama and Cha (2022) that EA Sports attribute ratings encode a dense, high-quality signal of player value that is difficult to displace.

Second, the blended features from Layer 2 occupy four of the top ten ranks, indicating that the meta-learner actively uses inter-expert agreement structure, not just the EA prediction alone. When the three experts broadly agree, the blended central tendency stabilises the final prediction; when they disagree, the disagreement features provide additional information.

Third, three engineered features appear in the top ten, confirming the contextual importance of contract, age, and economic signals highlighted throughout the literature (Frick, 2007; Müller, Simons and Weinmann, 2017; Poli, Besson and Ravenel, 2022).

Notably, the FBref expert prediction ranks 16th (SHAP = 0.0071), indicating that on-pitch statistics contribute meaningful but secondary information once EA ratings are already available. The disagreement feature ea_fb_absdiff appears at rank 18, providing direct empirical support for the hypothesis that inter-source disagreement is itself informative (Shen, 2025).

4.5 Meta-Learner Weighting and Architectural Contribution

The Layer 4 inverse-RMSE blend assigned the boosted meta-model a mean weight of 0.958 (std 0.053) across 15 runs, with the neural meta-model receiving 0.042 (std 0.053). In twelve of fifteen runs the boosted weight exceeded 0.95; in the remaining three, the neural contribution rose to between 0.09 and 0.17. This asymmetry indicates that gradient-boosted trees, with their native handling of feature interactions and robustness to mixed-scale inputs, are better suited to this tabular problem than the multi-layer perceptron which is consistent with broader empirical findings that tree-based methods remain competitive with neural networks on structured data (Ke et al., 2017; Prokhorenkova et al., 2018). The neural meta-learner

nonetheless serves as a stabilising regulariser: in the three runs where it received non-trivial weight, the blended prediction reduced final test RMSE relative to the boosted model alone.

4.6 Position-Group Analysis

Disaggregating the test predictions by position group reveals systematic but modest positional effects. Table 4.4 summarises mean predicted values by position.

Position	N (test)	Mean actual (€)	Mean predicted (€)	Mean σ (€)
MID	496	6,190,882	6,009,144	454,138
DEF	514	5,373,662	5,344,958	371,188
FWD	114	4,440,285	4,339,050	296,172
GK	116	2,803,341	2,499,505	138,712

Table 4.4: Mean actual value, mean predicted value, and mean prediction uncertainty by position group.

The model predicts most accurately on defenders (mean residual below 1% of mean value) and midfielders (2.9% under-prediction). Goalkeepers show the largest proportional under-prediction (approximately 10.8%), consistent with their lower volume in both training data and their higher-dimensional attribute space. Additionally, the FBref expert contributes almost nothing for goalkeepers because outfield per-90 statistics are not meaningful for them, and the model has correspondingly less information to work with (Behravan and Razavi, 2021). Forward predictions fall between these extremes. Prediction uncertainty (mean σ) is largest for midfielders, reflecting both the broader valuation range for this position and the higher variance in which player profiles drive value: creative playmakers and box-to-box workers are valued very differently by the market.

4.7 Sources of Residual Error

Three mechanisms appear to drive the residual 23% MAPE. First, EA Sports ratings lag real-world form changes by weeks or months (Lee, Tama and Cha, 2022), which creates predictable under-estimation for players in a period of form improvement and over-estimation for those in decline. Second, the target variable is itself an estimate rather than a true transfer-market price, and therefore carries its own measurement error (Müller, Simons and Weinmann, 2017). Third, the model has no access to the qualitative factors identified by Frick (2007) and Poli, Besson and Ravenel (2022), including agent dynamics, bargaining power, and club financial condition. The first of these could be addressed by integrating more recent performance data; the second by obtaining actual transfer fee data where available; and the third is likely to remain a

fundamental ceiling on achievable predictive accuracy in the absence of privileged commercial information.

5. Interactive Web Application

The modelling pipeline's outputs are packaged into a Streamlit (Streamlit Inc., 2024) web application that translates the research artefacts into a decision-support tool usable by non-technical stakeholders. This chapter details its purpose, architecture, and feature set.

5.1 Purpose and Design Philosophy

Quantitative valuation research has historically been criticised for producing outputs that remain inaccessible to the practitioners who would most benefit from them such as scouts, sporting directors, agents, and financial analysts (McHale and Holmes, 2023). The application addresses this gap by providing the derived insights through a browser-based interface requiring no programming knowledge.

The design philosophy is threefold: *transparency* (every prediction is accompanied by its uncertainty), *actionability* (users can evaluate concrete transfer scenarios), and *interpretability* (plain-language explanations accompany each prediction, following the attribution principles of Lundberg and Lee, 2017).

5.2 Technology Stack and Architecture

The application is implemented in Python using Streamlit (Streamlit Inc., 2024) for the interactive layer and Plotly (Plotly Technologies Inc., 2015) for visualisations. Streamlit was selected for its rapid development cycle, native support for data-science workflows, and widget-driven reactive interface; Plotly provides publication-quality interactive charts with hover tooltips, zoom, and filtering. Data loading is optimised through Streamlit's caching decorator:

5.3 Feature Walkthrough

The interface is organised as a sidebar of global filters paired with six primary tabs accessed through Streamlit's native tab widget.

Player Lookup provides a single-player view. Selecting any player renders four headline metric cards; predicted value, actual value, absolute difference, and percentage difference,

followed by a four-column profile panel of position, league, club, valuation status, a demographic panel; live age computed from date of birth, overall rating, potential, wage, and others such as minutes played, 90s and contract years. Where the prediction standard deviation is available, the app displays an indicative range defined as the mean prediction plus and minus one standard deviation. A rule-based explanation engine generates natural-language bullets describing which profile features are likely driving the valuation, for instance, noting that a player aged 23–28 sits within the peak-value range identified by Frick (2007), or that high minutes played provide stronger evidence of ability.

Deal Evaluator supports counterfactual transfer assessment. The user selects a player and enters a proposed purchase price; the app returns a 0–100 "deal score" computed from the percentage deviation between the modelled value and the proposed price.

Scores above 85 are labelled "Outstanding deal", 70–85 "Very good deal", 55–70 "Good / fair deal", down to "Very poor deal" below 20. This provides a single-number summary for rapid pre-screening of transfer opportunities while the underlying euro-denominated difference remains visible.

Market Inefficiencies surfaces systematic over- and under-valuation at both player and league level. It displays the top 25 undervalued and overvalued players side by side, then aggregates to a league-level table ranking competitions by mean percentage difference. This directly operationalises the value-identification use case demonstrated by McHale and Holmes (2023), who used their model to identify clubs consistently successful at finding undervalued players. A horizontal Plotly bar chart visualises the league ranking.

Uncertainty is dedicated to prediction reliability. A colour-coded scatter plot of predicted value against prediction standard deviation reveals the heteroscedastic pattern discussed in Section 4.3, and a histogram of standard deviations shows the overall reliability distribution. Users can filter by position group to inspect which segments the model treats most cautiously.

Insights Dashboard provides global model-calibration views: an actual-versus-predicted scatter plot (whose diagonal tightness is a visual proxy for R^2), a per-position mean percentage-error bar chart, a league-level summary table, and a horizontal bar chart of the top 20 features by mean absolute SHAP value.

Model Metrics exposes the raw evaluation results. A summary table presents the mean, standard deviation, minimum and maximum of each metric from the 15-run experiment, and an all-runs table allows inspection of run-level variability. This transparency follows the

calibration-disclosure principles established in the machine learning reproducibility literature (Pedregosa et al., 2011).

5.4 Design Rationale

Three design decisions deserve brief justification. The tab-based layout was chosen over a scrolling single-page design to reduce cognitive load and allow users to move directly to the task-relevant view. The decision to display indicative ranges (mean ± σ) rather than formal prediction intervals reflects the fact that the Monte Carlo standard deviation quantifies epistemic uncertainty from the training procedure, not total predictive uncertainty; a full predictive interval would require separately modelling aleatoric variance, which was beyond the scope of this project. Finally, the deal score is linear in percentage deviation by deliberate choice: a non-linear scoring function would encode stronger opinions about acceptable over- and under-payment than the model's underlying error distribution supports.

6. Discussion, Limitations and Conclusions

6.1 Discussion of Findings

The empirical results support the three main hypotheses advanced in Chapter 1. First, multi-source integration yields measurable accuracy gains: the 0.947 R^2 achieved here exceeds the 0.82–0.86 reported by Jishnu et al. (2022) on a comparable three-source feature set, with the architectural difference being the move from a single pooled model to domain-specialised experts. Second, the domain-specialised expert structure recommended by Shen (2025) generalises to the multi-source setting: SHAP analysis confirms that the EA expert, FBref expert, and Transfermarkt context are each drawn upon by the meta-learner, with blended features from Layer 2 occupying four of the top ten feature positions. Third, Monte Carlo repeated experimentation provides a practical route to uncertainty quantification without requiring Bayesian model redesign through the 15-run standard deviation which captures genuine cross-run variability and, as shown in Section 4.3, exhibits sensible heteroscedastic structure that aligns with the known right-skewness of the target distribution (Huebl and Swieter, 2002).

6.2 Limitations

Several limitations qualify the interpretation of these results. **First**, the target variable is a *synthetic* estimate (the average of EA and Transfermarkt values, or the single available source where one is missing), not a realised transfer fee. McHale and Holmes (2023) and Poli, Besson and Ravenel (2022) have shown that actual fees incorporate negotiation dynamics, agent commissions, and bargaining power that the current target cannot capture; the reported 23% MAPE therefore measures the gap to an estimated ground truth, not to true market-clearing prices. **Second**, EA Sports ratings are updated on a seasonal cycle and lag behind real-world form changes, which Lee, Tama and Cha (2022) identify as a known source of systematic error; the consequence is that players in hot or cold streaks are predictably mis-valued until the next rating refresh. **Third**, the analysis is cross-sectional rather than longitudinal: the model does not learn value *trajectories* over time, and therefore cannot forecast future valuations the way Sulimov (2024) proposes through time-series modelling. **Fourth**, coverage imbalance across leagues and positions means that per-segment error rates are not uniform. **Fifth**, the web scraping used to obtain SoFIFA data at scale raises ongoing maintenance questions: any future change in the source website's DOM structure could break the pipeline.

6.3 Future Work

Four directions would extend this work. Integrating the Transfermarkt *transfer history* dataset would allow training on realised fees where they exist, providing a ground-truth alternative to the current synthetic target (Poli, Besson and Ravenel, 2022). Adding a recurrent or temporal-convolution meta-learner would permit trajectory modelling and multi-horizon value forecasts (Sulimov, 2024). Incorporating unstructured signals such as news coverage, injury records, social-media sentiment would begin to address the information gap identified by Frick (2007) regarding unobservable qualitative factors. Finally, validating the model prospectively, by recording a snapshot of current predictions and comparing them to transfer fees realised over the subsequent transfer window, would provide the strongest possible test of real-world utility.

6.4 Conclusions

This project set out to develop and evaluate a multi-source machine learning system for predicting professional football player market values with high accuracy and quantified uncertainty. The resulting four-layer stacked ensemble, combining domain-specialised LightGBM, XGBoost, and CatBoost experts with a dual boosted-and-neural meta-learner, achieves a mean R^2 of 0.947 and MAPE of 23.2% on held-out test data across 15 Monte Carlo

runs. This approaches the best published benchmark (Shen, 2025) without requiring meta-heuristic optimisation and materially exceeds the comparable multi-source studies of Behravan and Razavi (2021), Jishnu et al. (2022), and McHale and Holmes (2023). The accompanying Streamlit application translates these results into an accessible decision-support tool with player lookup, deal evaluation, market-inefficiency scanning, uncertainty inspection, and model-transparency views. The six project objectives stated in Section 1.3 have been met in full: the three sources are cleaned and integrated (O1); exploratory and descriptive statistics characterise the data (O2); age, contract, and positional features are engineered in line with the literature (O3); the stacked ensemble is designed and implemented (O4); a comprehensive metric suite is reported (O5); and SHAP-based interpretation is provided both globally in this report and interactively in the application (O6). The work contributes to the player-valuation literature a reproducible, uncertainty-aware, multi-source architecture, and offers practitioners a free and transparent tool for data-informed transfer decisions.

Reference List

Behravan, I. and Razavi, S.M. (2021) 'A novel machine learning method for estimating football players' value in the transfer market', *Soft Computing*, 25(4), pp. 2499–2511. Available at: https://doi.org/10.1007/s00500-020-05319-3. (Accessed: 21 April 2026)

Bell, A.R., Brooks, C. and Brooks, R. (2022) *Are English football players overvalued?* SSRN Working Paper 4080940. Available at: https://ssrn.com/abstract=4080940. (Accessed: 21 April 2026)

Boccard, N. (2022) Valuable football players: Where do we find them? SSRN Working Paper 4234038. Available at: https://ssrn.com/abstract=4234038 (Accessed: 21 April 2026).

Chen, T. and Guestrin, C. (2016) 'XGBoost: A Scalable Tree Boosting System', *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 785–794. Available at: https://doi.org/10.1145/2939672.2939785 (Accessed: 21 April 2026).

Creswell, J.W. and Creswell, J.D. (2018) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 5th edn. Thousand Oaks, CA: SAGE Publications. Available at: https://www.researchgate.net/profile/Leopold-Hamminger/post/Can_you_help_me_clarify_what_type_of_qualitative_research_design_to_use/attachment/59d64d4579197b80779a6db2/AS%3A487475249455105%401493234564396/download/Creswell+-+Research+Design.pdf (Accessed: 21 April 2026).

Depken, C.A. and Frei, A.P. (2023) *Player Nationality and Transfer Fee Premiums in Association Football.* SSRN Working Paper 4335733. Available at: https://ssrn.com/abstract=4335733 (Accessed: 21 April 2026).

FIFA (2026) *FIFA Global Transfer Report 2025*. Inside FIFA. Available at: https://inside.fifa.com/transfer-system/transfer-reports?tab=Men%27s+global+transfer+report+2025 (Accessed: 21 April 2026).

Frick, B. (2007) 'The football players' labor market: Empirical evidence from the major European leagues', *Scottish Journal of Political Economy*, 54(3), pp. 422–446. Available at: https://doi.org/10.1111/j.1467-9485.2007.00423.x (Accessed: 21 April 2026).

Harris, C.R. et al. (2020) 'Array programming with NumPy', *Nature*, 585(7825), pp. 357–362. Available at: https://doi.org/10.1038/s41586-020-2649-2 (Accessed: 21 April 2026).

Huebl, L. and Swieter, D. (2002) 'Der Spielermarkt in der Fußball-Bundesliga', Zeitschrift für Betriebswirtschaft, 72(Ergänzungsheft 4), pp. 105–123.

Hunter, J.D. (2007) 'Matplotlib: A 2D Graphics Environment', Computing in Science & Engineering, 9(3), pp. 90–95. Available at: https://doi.org/10.1109/MCSE.2007.55 (Accessed: 21 April 2026).

Jishnu, V.B. et al. (2022) 'Football Player Transfer Value Prediction Using Advanced Statistics and FIFA 22 Data', *Proceedings of the 2022 IEEE 19th India Council International Conference (INDICON)*. Available at: https://doi.org/10.1109/INDICON56171.2022.10040117 (Accessed: 21 April 2026).

Ke, G. et al. (2017) 'LightGBM: A Highly Efficient Gradient Boosting Decision Tree', *Advances in Neural Information Processing Systems,* 30, pp. 3146–3154. Available at: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (Accessed: 21 April 2026).

Kingma, D.P. and Ba, J. (2015) 'Adam: A Method for Stochastic Optimization', *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015).* Available at: https://arxiv.org/abs/1412.6980 (Accessed: 21 April 2026).

Lee, H., Tama, B.A. and Cha, M. (2022) *Prediction of Football Player Value using Bayesian Ensemble LightGBM Approach.* arXiv preprint arXiv:2206.13246. Available at: https://arxiv.org/abs/2206.13246 (Accessed: 21 April 2026).

Lundberg, S.M. and Lee, S.-I. (2017) 'A unified approach to interpreting model predictions', *Advances in Neural Information Processing Systems*, 30, pp. 4765–4774. Available at: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (Accessed: 21 April 2026).

Majewski, S. (2016) 'Identification of factors determining market value of the most valuable football players', *Journal of Management and Business Administration. Central Europe*, 24(3), pp. 91–104. Available at: https://doi.org/10.7206/jmba.ce.2450-7814.177. (Accessed: 21 April 2026).

McHale, I.G. and Holmes, B. (2023) 'Estimating transfer fees of professional footballers using advanced performance metrics and machine learning', *European Journal of Operational Research*, 306(1), pp. 389–399. Available at: https://doi.org/10.1016/j.ejor.2022.06.033. (Accessed: 21 April 2026).

McKinney, W. (2010) 'Data structures for statistical computing in Python', in *Proceedings of the 9th Python in Science Conference*, pp. 56–61. Available at: https://proceedings.scipy.org/articles/Majora-92bf1922-00a (Accessed: 21 April 2026).

Müller, O., Simons, A. and Weinmann, M. (2017) 'Beyond crowd judgments: Data-driven estimation of market value in association football', *European Journal of Operational Research*, 263(2), pp. 611–624. Available at: https://doi.org/10.1016/j.ejor.2017.05.005. (Accessed: 21 April 2026).

Pedregosa, F. *et al.* (2011) 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research*, 12, pp. 2825–2830. Available at: http://jmlr.org/papers/v12/pedregosa11a.html (Accessed: 21 April 2026).

Poli, R., Besson, R. and Ravenel, L. (2022) 'Econometric approach to assessing the transfer fees and values of professional football players', *Economies*, 10(1), 4. Available at: https://doi.org/10.3390/economies10010004. (Accessed: 21 April 2026).

Prokhorenkova, L. *et al.* (2018) 'CatBoost: Unbiased boosting with categorical features', *Advances in Neural Information Processing Systems*, 31, pp. 6638–6648. Available at: https://papers.neurips.cc/paper/7898-catboost-unbiased-boosting-with-categorical-features.pdf (Accessed: 21 April 2026).

Rosen, S. (1974) 'Hedonic prices and implicit markets: Product differentiation in pure competition', *Journal of Political Economy*, 82(1), pp. 34–55. Available at: https://doi.org/10.1086/260169. (Accessed: 21 April 2026).

Rottenberg, S. (1956) 'The baseball players' labor market', *Journal of Political Economy*, 64(3), pp. 242–258. Available at: https://doi.org/10.1086/257790. (Accessed: 21 April 2026).

Saunders, M., Lewis, P. and Thornhill, A. (2019) *Research methods for business students*. 8th edn. Harlow: Pearson Education.

Shen, Q. (2025) 'Predicting the value of football players: Machine learning techniques and sensitivity analysis based on FIFA and real-world statistical datasets', *Applied Intelligence*, 55, 265. Available at: https://doi.org/10.1007/s10489-024-06189-0. (Accessed: 21 April 2026).

Sulimov, D. (2024) *Performance insights-based AI-driven football transfer fee prediction*. arXiv preprint arXiv:2401.16795. Available at: https://arxiv.org/abs/2401.16795 (Accessed: 21 April 2026).

Sun, Y. and Gu, K. (2024) 'Prediction of football players' value in the transfer market of well-known European leagues based on FIFA 19 and real-world data', *The International Arab Journal of Information Technology*, 21(4), pp. 723–736. Available at: https://doi.org/10.34028/iajit/21/4/13.

van den Berg, E. (2011) *The valuation of football players: A theoretical and empirical analysis*. Master's thesis. Erasmus University Rotterdam. Available at: https://thesis.eur.nl/pub/9763/Berg,%20E.%20van%20den%20%28296332%29.pdf (Accessed: 21 April 2026).

van Ours, J.C. and van Tuijl, M. (2024) 'Incentives matter sometimes: On the differences between league and cup football matches', *Sports Economics Review*, 7, 100037. Available at: https://doi.org/10.1016/j.serev.2024.100037. (Accessed: 21 April 2026).

Van Rossum, G. and Drake, F.L. (2009) *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Wolpert, D.H. (1992) 'Stacked generalization', *Neural Networks*, 5(2), pp. 241–259. Available at: https://doi.org/10.1016/S0893-6080(05)80023-1 (Accessed: 21 April 2026).

prashantghimire (2025) *sofifa-web-scraper* [Computer code]. GitHub. Available at: https://github.com/prashantghimire/sofifa-web-scraper (Accessed: 21 April 2026).