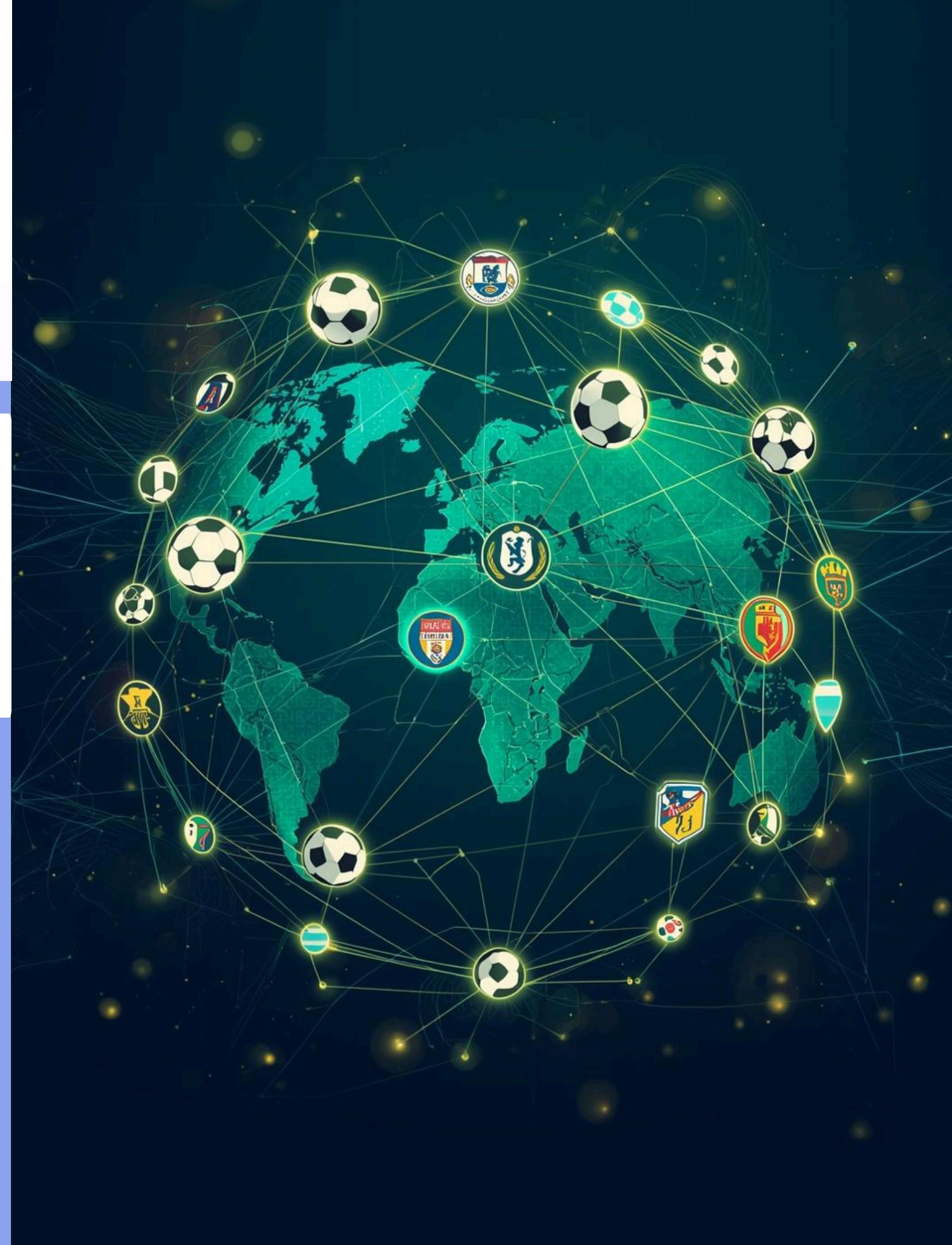




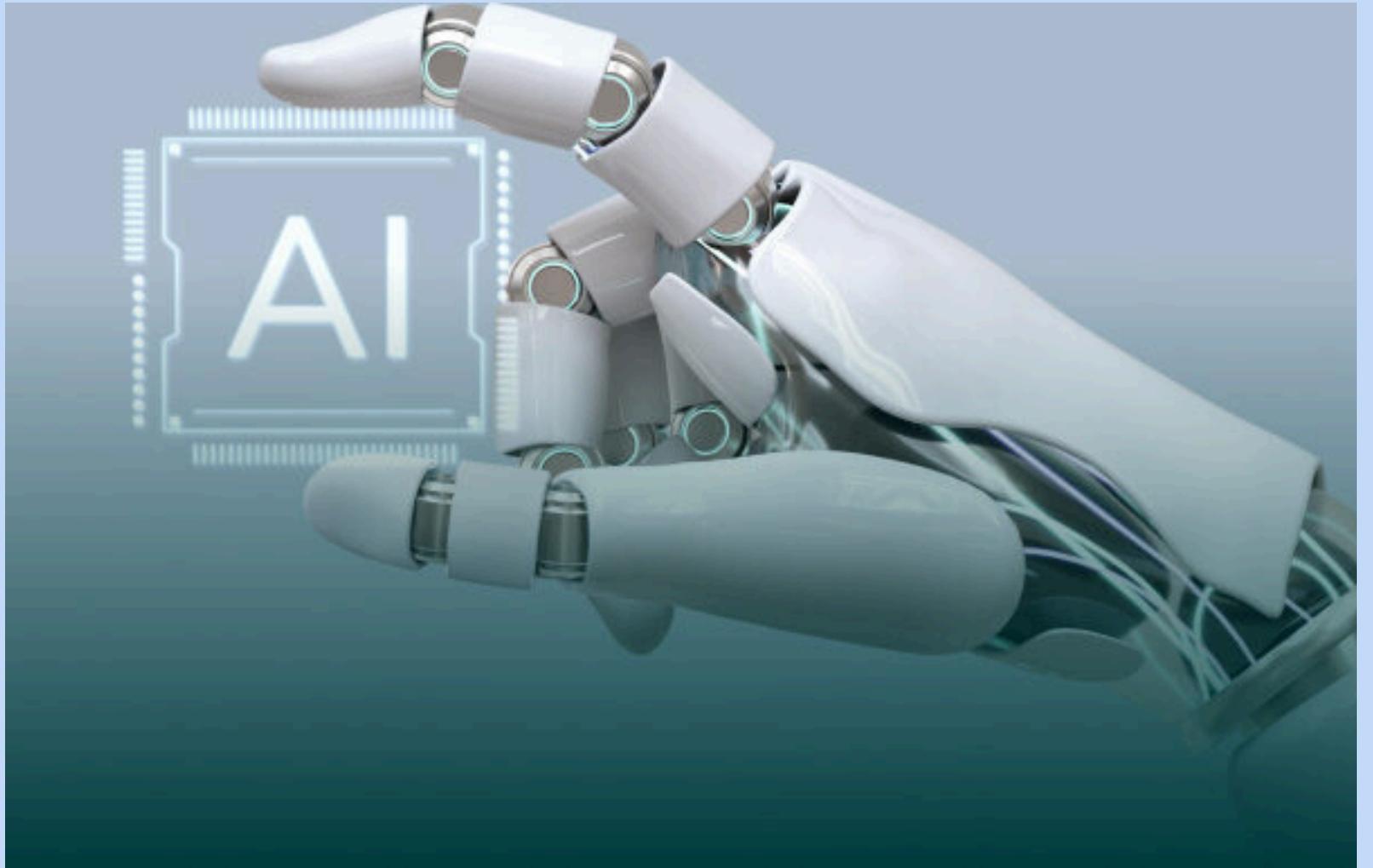
Dynamic Football Player Transfer Value Prediction

Presented By :
Karan Agrawal
Dhanshri Supratkar
Renu Kumari Prajapati
Shrabani Mondal
Tejaramasai Matta



Overview

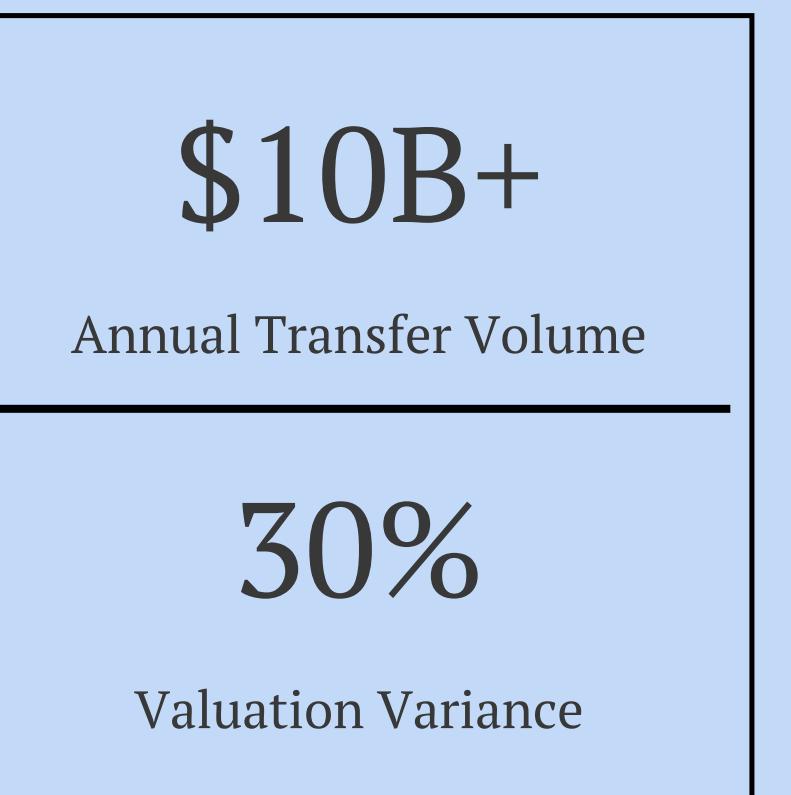
- ▶ Introduction
- ▶ Business Problem
- ▶ System Architecture
- ▶ Data Collection & Preparation
- ▶ Working of model
- ▶ Results and discussions
- ▶ Limitations and future work
- ▶ References

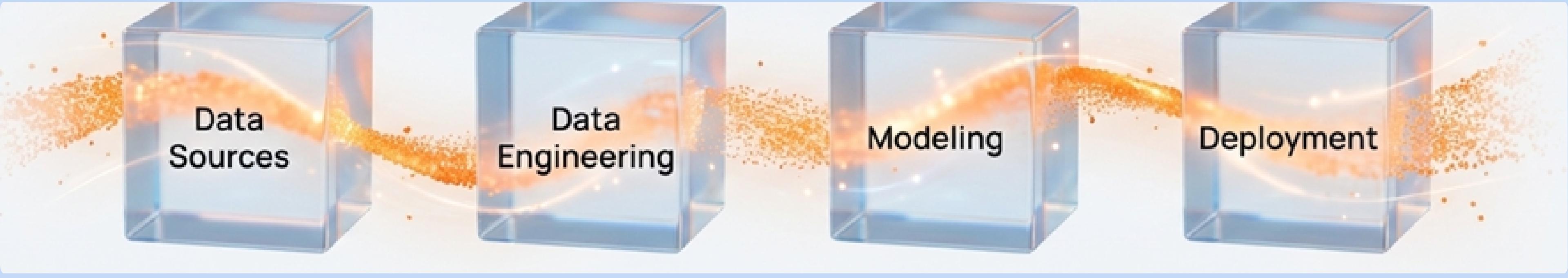


Introduction

Understanding the Dynamics of Football's Lucrative Financial Landscape

- The football transfer market is a **multi-billion dollar industry**, driven by player valuations that have significant financial implications.
 - Inaccurate assessments can lead to substantial financial risk for clubs.
 - As the market expands, understanding its complexities becomes essential for informed decision-making and strategic investments in talent acquisition.





Our Solution: An End-to-End ML System



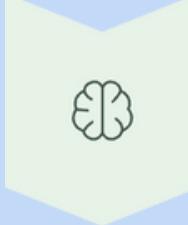
Data Sources

Multi-source integration aggregating statistics from Transfermarkt, FBref, and official league databases. Comprehensive coverage of player performance, transfer history, and market trends.



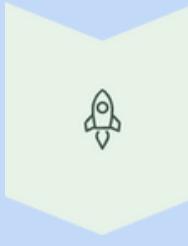
Data Engineering

Robust cleaning, transformation, and feature creation pipelines. Automated quality control ensuring data consistency and reliability across all sources.



Modeling

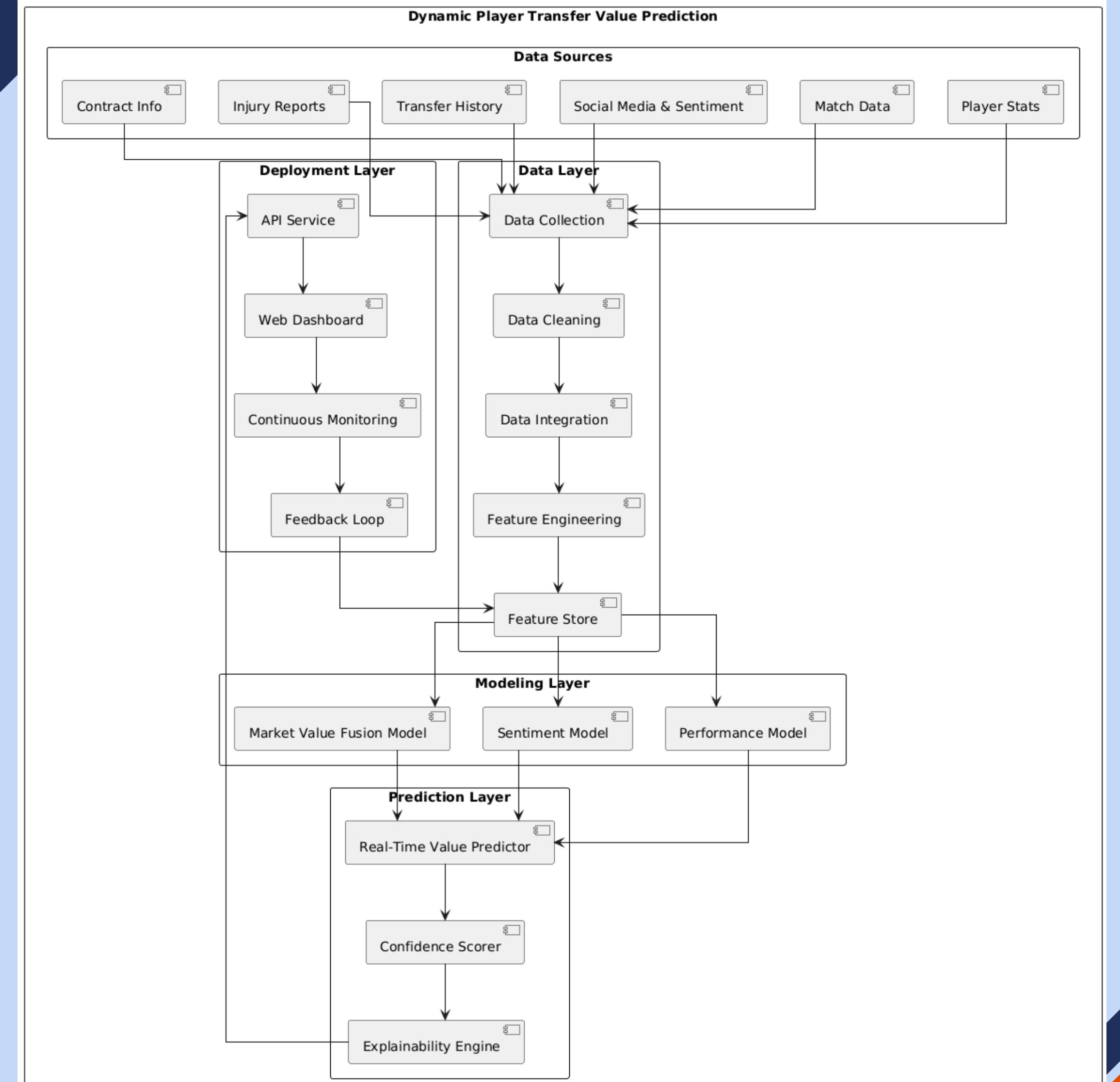
Advanced machine learning algorithms with built-in explainability. Ensemble methods combining multiple models for superior prediction accuracy and confidence intervals.



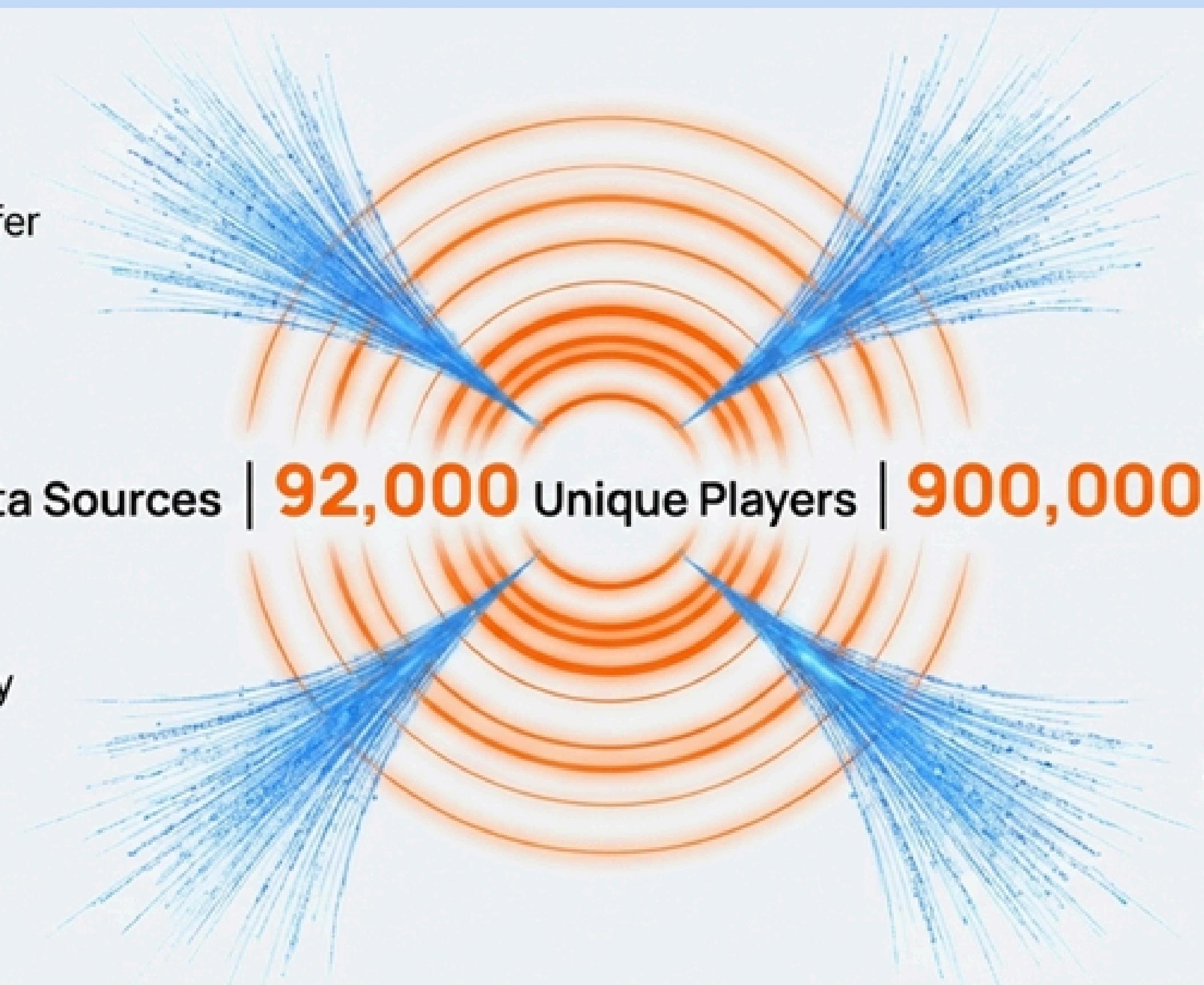
Deployment

Production-grade web application enabling instant valuations. Real-time inference with interactive visualizations and comprehensive player insights.

Methodology



DATA COLLECTION



Career: Complete transfer history, contract status, agent details.

Performance: On-field stats, team context, competition level.

11 Independent Data Sources | **92,000** Unique Players | **900,000+** Raw Event Records

Physical: Detailed injury history, days missed, recurrence.

Sentiment: Public perception, social media tone, brand value.

The Data Engineering Challenge

Raw Data State

Noisy, duplicated records with inconsistent formats. Some columns showed up to 100% missing values.

Clean Dataset

A reliable, analysis-ready dataset optimized for machine learning.



Transformation

Player-level aggregation, date standardization, and systematic removal of non-informative fields.

Pipeline Automation

Reusable, automated data pipeline enabling repeatable and scalable data processing.

Before: Fragmented data across multiple schemas and formats

Process: Automated validation and quality control pipelines

After: Unified, consistent data infrastructure

FEATURE ENGINEERING

Translating Data into Value Drivers

Raw statistics don't directly predict market value. We engineered features that capture the underlying economic and performance dynamics that drive player valuation.



Per-90 Normalization

Standardized metrics accounting for playing time, enabling fair comparisons across players with different minutes.



Career Dynamics

Age curves, contract expiration timing, and transfer frequency as market signals.



Rolling Averages

Captured recent form and momentum through time-windowed performance indicators.



Injury Risk

Quantified injury history as an economic depreciation factor affecting long-term value.

Key Aspects of Feature Engineering

Performance Metrics

Evaluating player output through key stats

Age Dynamics

Understanding age impact on valuations

Transfer Trends

Analyzing market fluctuations for insights

Injury Risk

Assessing health's impact on value

Sentiment Analysis

Capturing public opinion on players

Advanced Feature Engineering

1. Handling High-Cardinality Data

The Challenge: Players, Clubs, and Leagues create thousands of categories.

The Solution: Cross-Validated Target Encoding.

The Impact: Converts complex text (e.g., "Real Madrid") into a single predictive numerical signal without data leakage.

2. Fixing Skewed Distributions

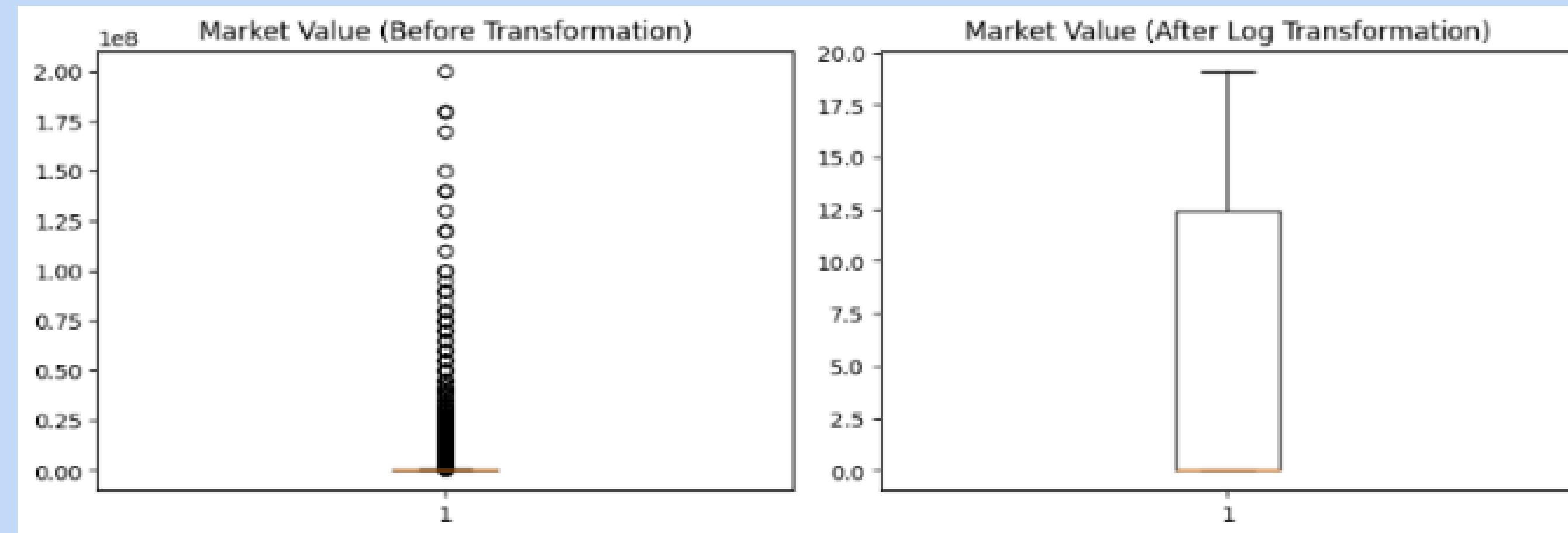
The Challenge: Extreme outliers in minutes played and transfer fees.

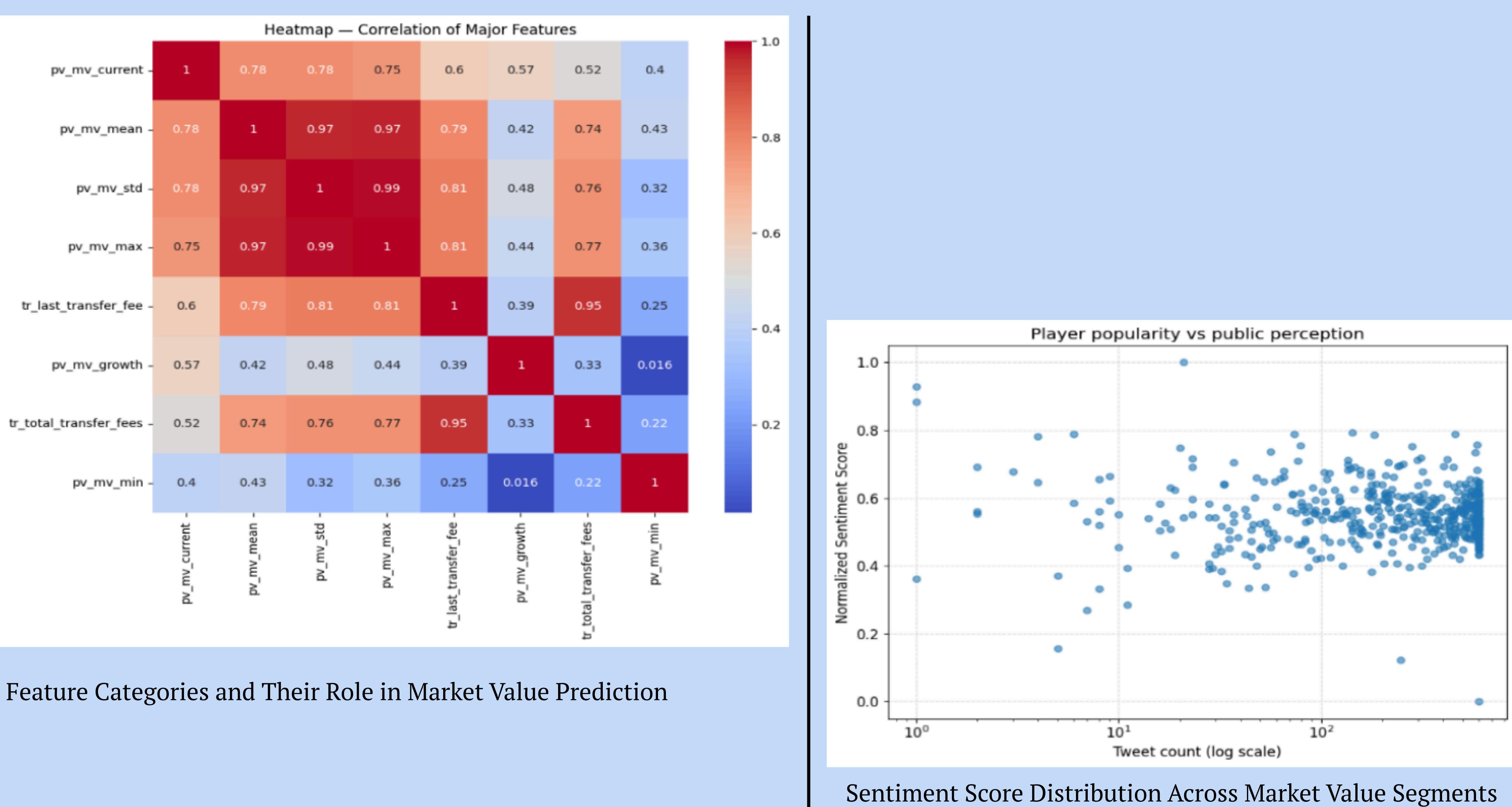
The Solution: Log Transformation (\log_{10}).

The Impact: Normalizes "superstar" data points, ensuring the model learns from the whole league, not just the top 1%.

Symmetric Output :

Normal-like distribution ready for modeling





Working of model

Progressive Evaluation Framework

Linear Regression

Simple baseline—too restrictive for complex, non-linear relationships

Lasso Regression

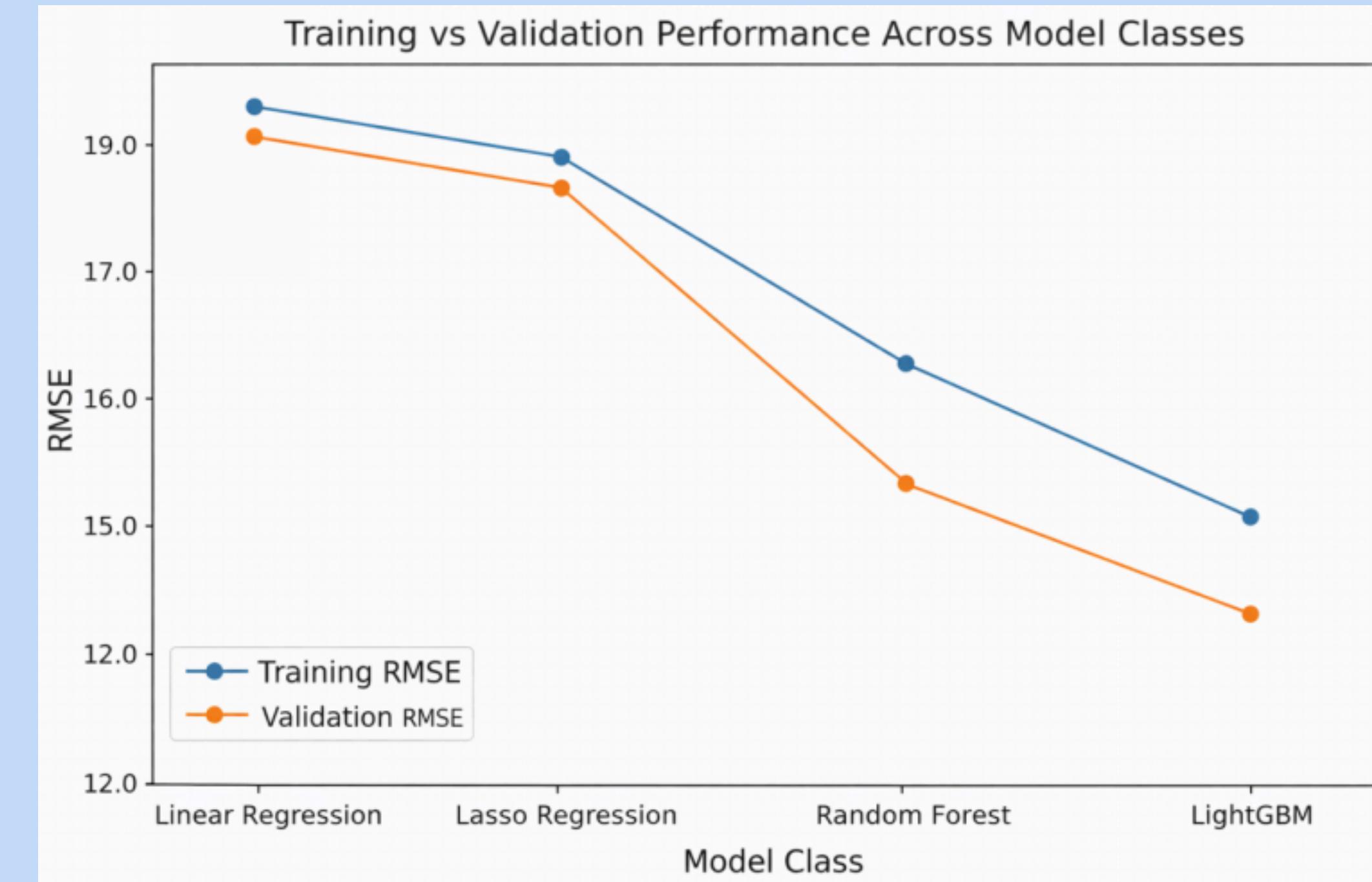
Added regularization but still limited by linearity assumptions

Random Forest

Strong performance but exhibited instability and longer training times.

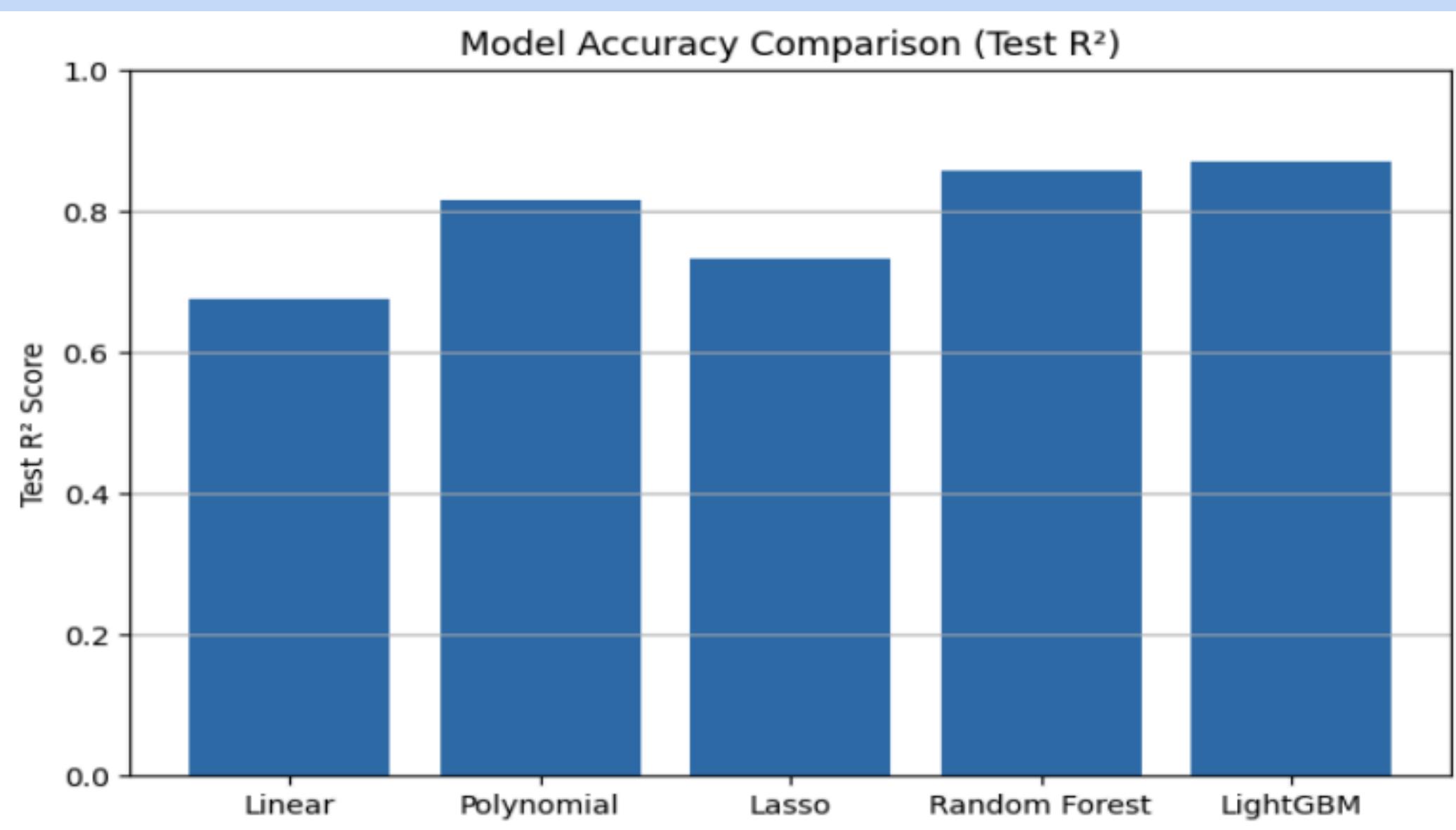
LightGBM

Winner: Optimal balance of accuracy, stability and computational efficiency.



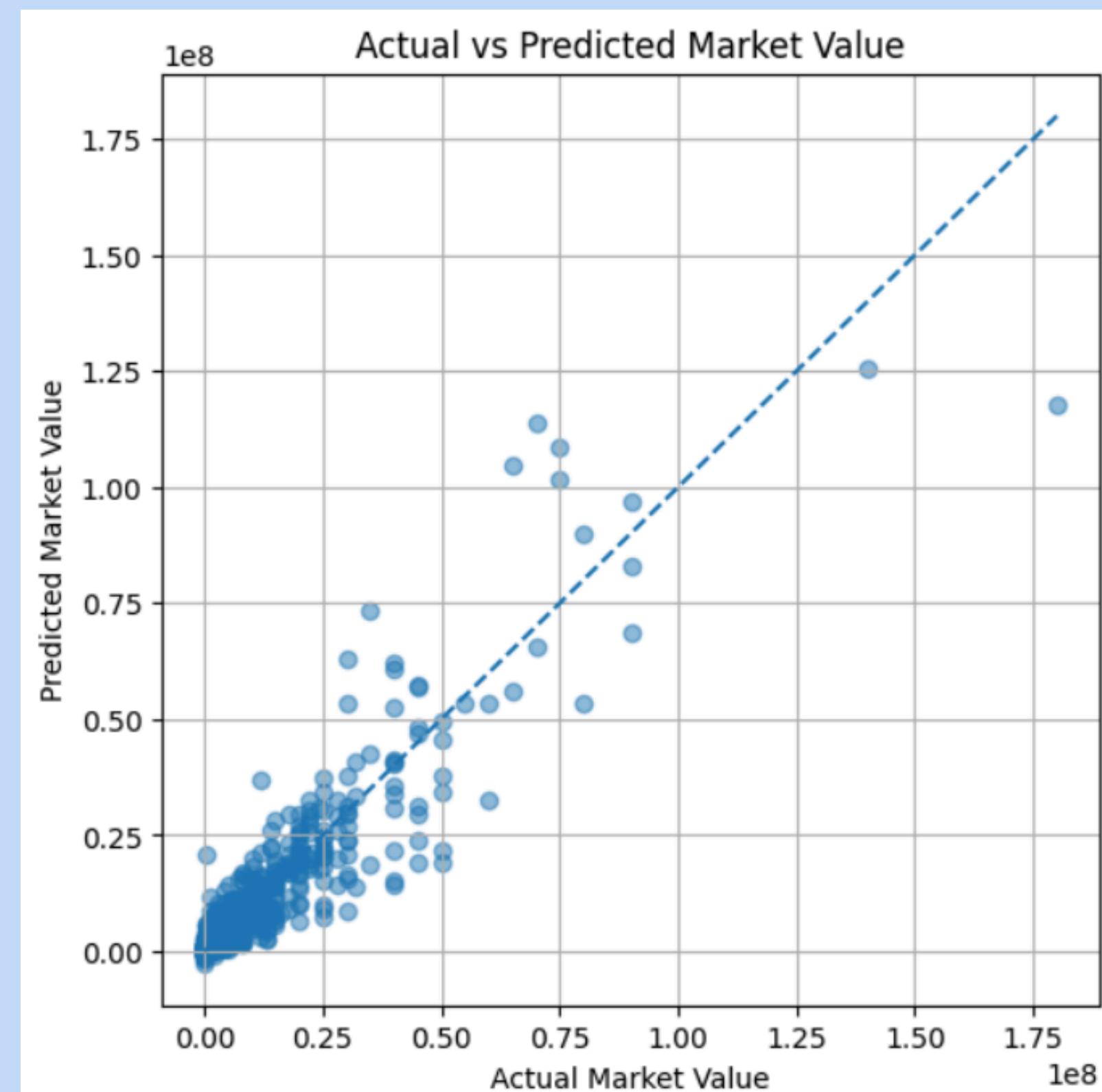
LightGBM's gradient boosting framework delivered superior generalization with faster inference—critical for production deployment at scale.

Results and Discussions



Model Accuracy Comparison

Actual vs Predicted Market Value
(Final Model)



 Player Inputs

Age

29

[-](#) [+](#)

Total Goal Assists

13

[-](#) [+](#)

Assists (Last 10 Matches)

8

[-](#) [+](#)

Total Injuries

3

[-](#) [+](#)

Years Since Last Transfer

1.50

[-](#) [+](#)

Last Transfer Fee (€)

980000.00

[-](#) [+](#)

Market Value % from Peak

0.70

[-](#) [+](#)

Football Player Market Value Prediction

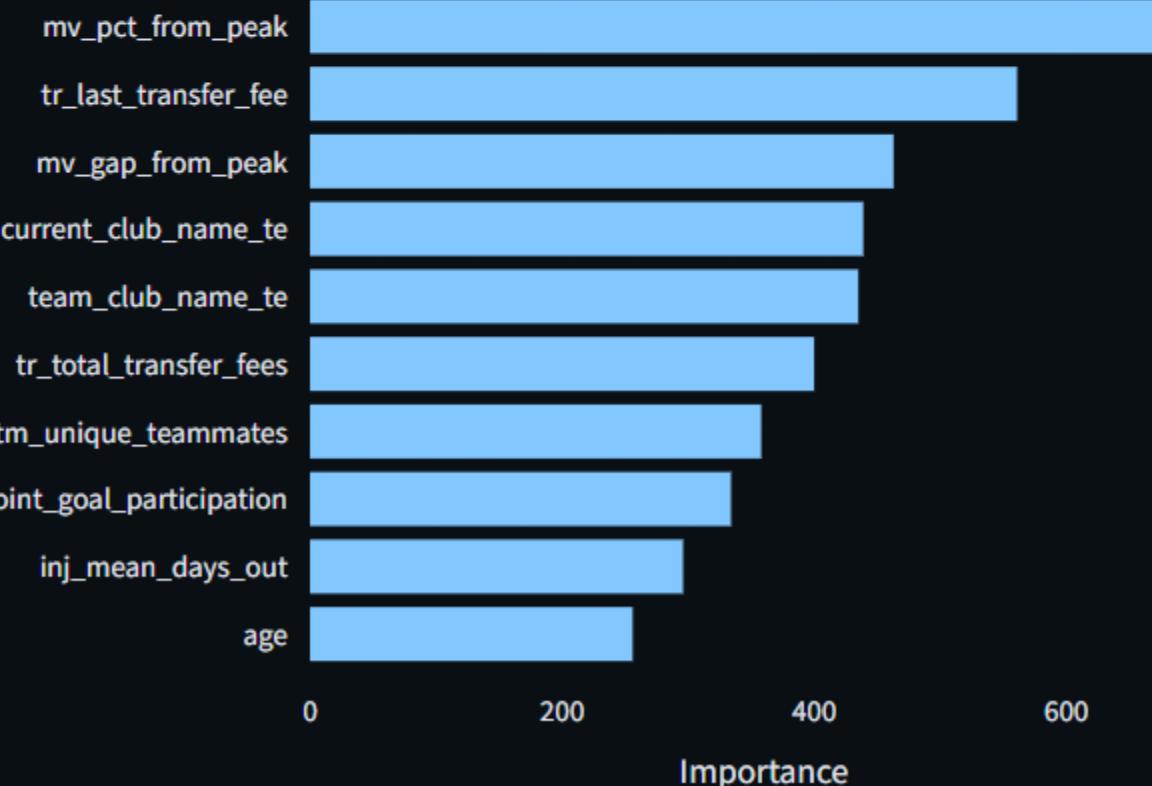
Predict player market value using a LightGBM-based machine learning model trained on performance, transfer, injury, and market data.



Feature Importance (Model Level)

Top Features Driving Market Value

Feature



Predicted Market Value

[Predict Market Value](#)

Estimated Market Value: €1,376,809

Estimated Market Value (€)

Limitations & Future Scope

Current Limitations

Data sparsity: Predictions are less certain for players with limited historical data.

Sentiment noise: Social media signals are weaker for low-visibility players

Market shocks: Sudden transfers or major injuries may not be reflected immediately.

Future Scope

Planned Enhancements:

- Confidence scoring: Attach uncertainty estimates to predictions
- Advanced NLP: Improve sentiment analysis using richer text models
- Scouting integration: Incorporate qualitative scouting reports and tracking data

Vision :

Evolve the system into a comprehensive, decision-intelligence platform for football analytics.



References

- Transfermarkt GmbH. Player Market Value, Transfer History, and Contract Data.
- FIFA & Domestic League Match Data Providers. Player and Team Performance Statistics.
- Pedregosa et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research.
- Ke, G. et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. NeurIPS.
- Lundberg, S. M., & Lee, S. I. A Unified Approach to Interpreting Model Predictions (SHAP). NeurIPS.
- Twitter / X API Documentation. Public Social Media Data for Sentiment Analysis.
- Streamlit Documentation. Interactive Web Application Deployment for ML Models.

THANK YOU!

