



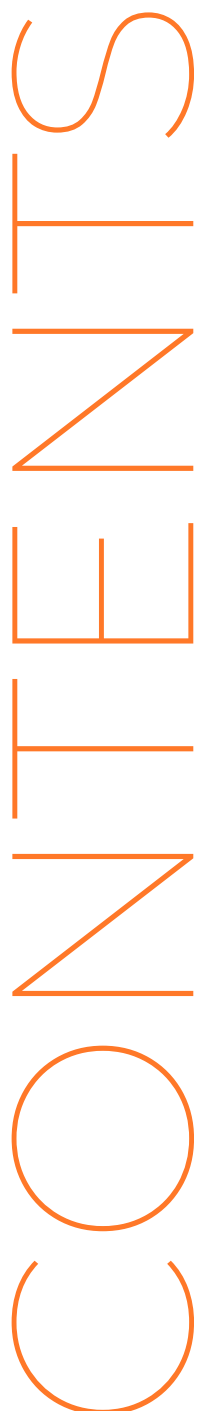
INFOSYS
SPRINGBOARD

AI Transfer IQ Project Report

Prepared by
Jeevan George John

Presented To
Mentor, Infosys

Table of Contents



02.

MILESTONE 1

04.

MILESTONE 2

06.

MILESTONE 3

08.

MILESTONE 4

10.

MILESTONE 5

12.

MILESTONE 6

14.

MILESTONE 7

16.

NEXT STEPS

17.

CONCLUSION

Milestone 1: Data Collection and Initial Exploration (Week 1)

Objective

The objective of this milestone was to establish a strong and diverse data foundation required for estimating football player market value in a realistic and business-aligned manner.

Description

Player market value is influenced by multiple factors beyond simple performance metrics. In this phase, data was collected from different sources to reflect real-world valuation logic used by clubs, scouts, and analysts. This included player performance data, market value benchmarks, public sentiment, and injury-related information. Initial exploration was conducted to understand data distribution, missing values, and overall data structure. This step ensured early identification of data gaps and inconsistencies before further processing.

Outcome

Availability of raw datasets from multiple sources
Initial understanding of player value distribution
Clear insight into key valuation drivers

master_dataset_with_ids - Saved to this PC																			
File Home Insert Draw Page Layout Formulas Data Review View Help Tell me what you want to do																			
Clipboard Font Alignment Number Styles Cells Editing																			
Comments Share																			
player_unique_id																			
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Player Unique	Player	player_key	Rank	Age	Nationality	Club	Market Value	Value Cr	Rank_perf	Age_perf	Personality	Matches	Goals	Own Goals	Assists	Yellow Card	Red Card	Cards	
1	Lamine Yamal	1	18	Spain	FC Barcelona	€1,600 Cr	1600	1	18	Spain	11	6	0	6	2	0			
2	Jude Bellingham	2	22	England	Real Madrid	€1,440 Cr	1440	2	22	England	11	3	0	1	1	0			
3	Erling Haaland	3	25	Norway	Manchester City	€1,440 Cr	1440	3	25	Norway	16	20	0	1	0	0			
4	Kylian Mbappé	4	26	France	Real Madrid	€1,440 Cr	1440	4	26	France	16	18	0	2	2	0			
5	Vinicius Jr	5	25	Brazil	Real Madrid	€1,200 Cr	1200	5	25	Brazil	18	6	0	5	3	0			
6	Pedri	6	22	Spain	FC Barcelona	€1,200 Cr	1200	6	22	Spain	13	2	0	2	0	1	0		
7	Jamal Musajamal	7	22	Germany	Bayern M.	€1,120 Cr	1120	7	22	Germany	0	0	0	0	0	0			
8	Bukayo Saka	8	24	England	Arsenal FC	€1,120 Cr	1120	8	24	England	15	7	0	0	0	0			
9	Alexander Isak	9	26	Sweden	Liverpool	€1,120 Cr	1120	9	26	Sweden	8	1	0	1	0	0			
10	Florian Grillitsch	10	22	Germany	Liverpool	€1,040 Cr	1040	10	22	Germany	16	0	0	3	2	0			
11	Michael Olise	11	23	France	Bayern M.	€1,040 Cr	1040	11	23	France	17	7	0	7	5	0			
12	Federico Vicedini	12	27	Uruguay	Real Madrid	€1,040 Cr	1040	12	27	Uruguay	15	0	0	4	1	0			
13	Cole Palmer	13	23	England	Chelsea FC	€960 Cr	960	13	23	England	4	2	0	0	0	0			
14	Declan Rice	14	26	England	Arsenal FC	€960 Cr	960	14	26	England	18	2	0	5	2	0			
15	Moisés Caicedo	15	24	Ecuador	Chelsea FC	€800 Cr	800	15	24	Ecuador	16	4	0	1	4	0			
16	Julián Álvarez	16	25	Argentina	Atlético de Madrid	€800 Cr	800	16	25	Argentina	16	9	0	4	1	0			
17	Alexis Mac Allister	17	26	Argentina	Liverpool	€800 Cr	800	17	26	Argentina	18	3	0	2	2	0			
18	Ousmane Dembélé	18	28	France	Paris Saint-Germain	€800 Cr	800	18	28	France	9	3	0	2	1	0			
19	Desiré Dou, Desiré Dou	19	20	France	Paris Saint-Germain	€720 Cr	720	19	20	France	8	3	0	3	0	0			
20	João Neves	20	21	Portugal	Paris Saint-Germain	€720 Cr	720	20	21	Portugal	8	5	0	0	0	0			
21	Khvicha Kvaratskhelia	21	24	Georgia	Paris Saint-Germain	€720 Cr	720	21	24	Georgia	14	3	0	3	1	0			
22	Vitinha	22	25	Portugal	Paris Saint-Germain	€720 Cr	720	22	25	Portugal	16	2	0	8	0	0			
23	Ryan Gravenberch	23	23	Netherlands	Liverpool	€720 Cr	720	23	23	Netherlands	12	3	0	2	3	0			
24	Raphinha	24	28	Brazil	FC Barcelona	€720 Cr	720	24	28	Brazil	7	3	0	2	2	0			
25	Rodrigo	25	29	Spain	Manchester City	€720 Cr	720	25	29	Spain	8	0	0	0	1	0			
26	Hugo Ekström	26	23	France	Liverpool	€680 Cr	680	26	23	France	16	6	0	1	0	1			

DISCOVER

2025 - 21:56

The Dark Side of ISL Uncertainty

Overspending clubs and comfortable players sank Indian football system

UPDATE 23.12.2025 - 15:44

December 23rd

Transfer news LIVE: Man United want Alex Scott as Endrick poised to leave Real Madrid

17.12.2025 - 11:54

303 Players Updated

ISL suffers catastrophic ₹144 Cr Market Value crash; Ryan Williams most valuable Indian

Milestone 2: Data Cleaning and Preprocessing (Week 2)

Objective

To transform raw data into reliable, consistent, and business-ready datasets suitable for valuation modeling.

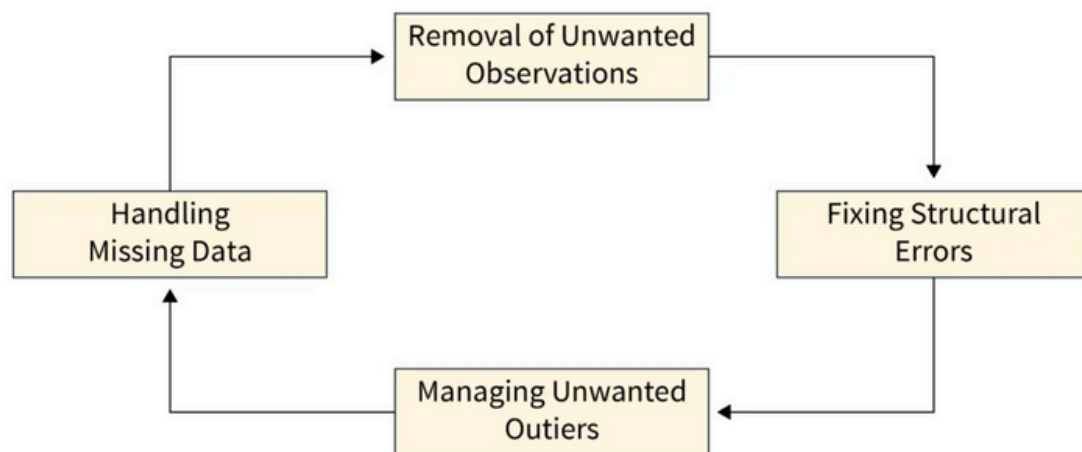
Description

Raw football data often contains inconsistencies, missing values, and redundant information. In this milestone, data was cleaned to ensure fairness and consistency across all player profiles. Numerical attributes were standardized, categorical data was structured, and preliminary sentiment indicators were refined.

This phase ensured that player comparisons are meaningful and not distorted by data quality issues.

Outcome

- Cleaned and standardized datasets
- Improved data reliability for valuation
- Foundation for accurate feature engineering



3 O E N T O S E T I N

Milestone 3: Advanced Feature Engineering and Sentiment Analysis (Weeks 3–4)

Objective

To enhance player profiles with advanced, value-oriented indicators that better reflect real-world football market behavior.

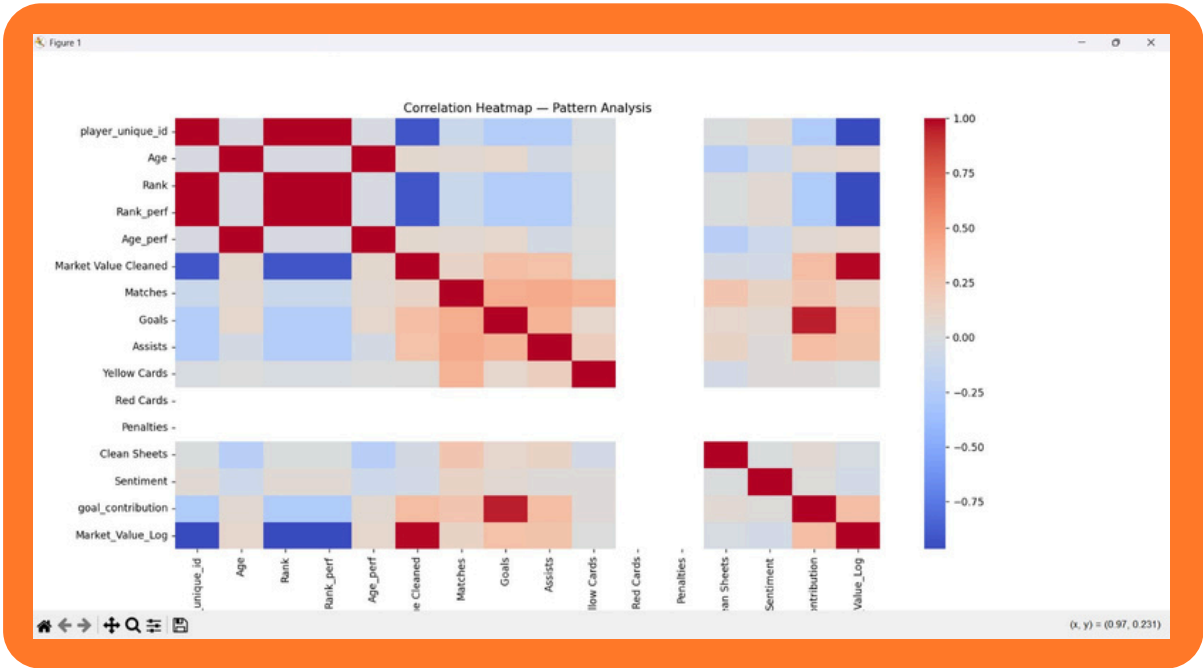
Description

Basic statistics were refined into advanced features such as performance trends, injury impact indicators, and sentiment scores derived from public perception. These features were designed to capture not just current performance, but long-term value potential and market perception.

Sentiment analysis was integrated to represent how media and public opinion influence player valuation.

Outcome

Finalized feature set aligned with market logic
Integrated sentiment and risk indicators
Dataset fully prepared for model training



Milestone 4: Model Development (Week 5)

Objective

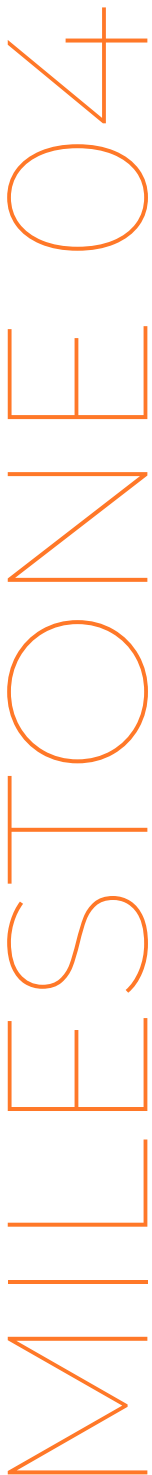
To build predictive machine learning models capable of estimating football player market values using engineered performance and contextual features.

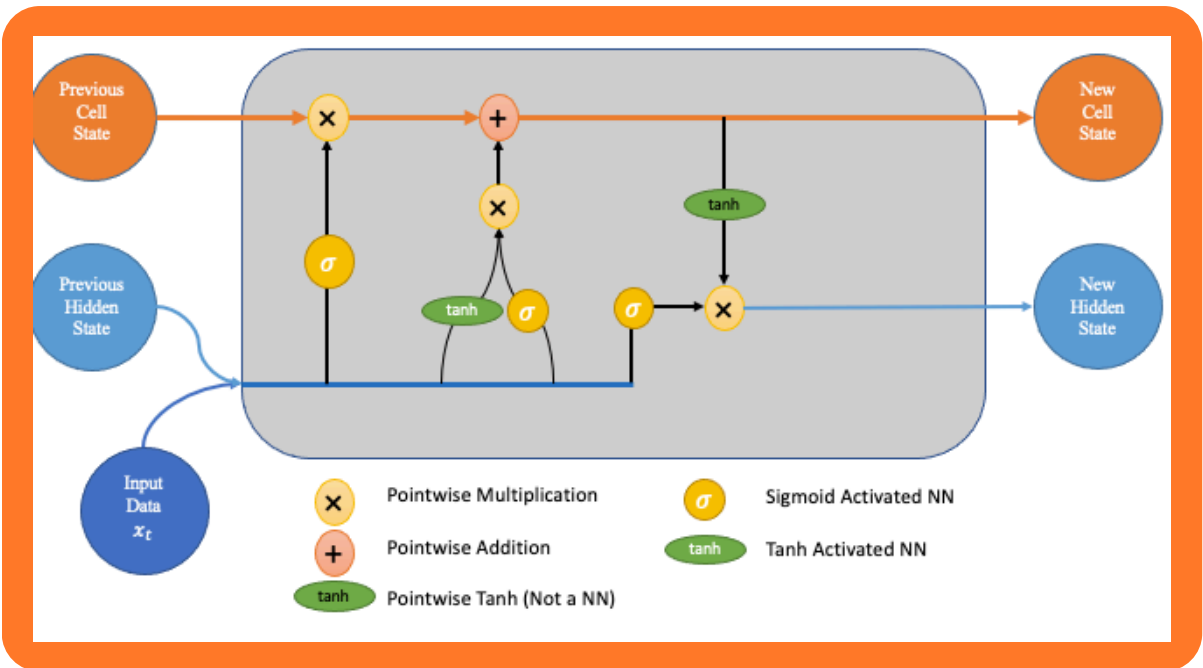
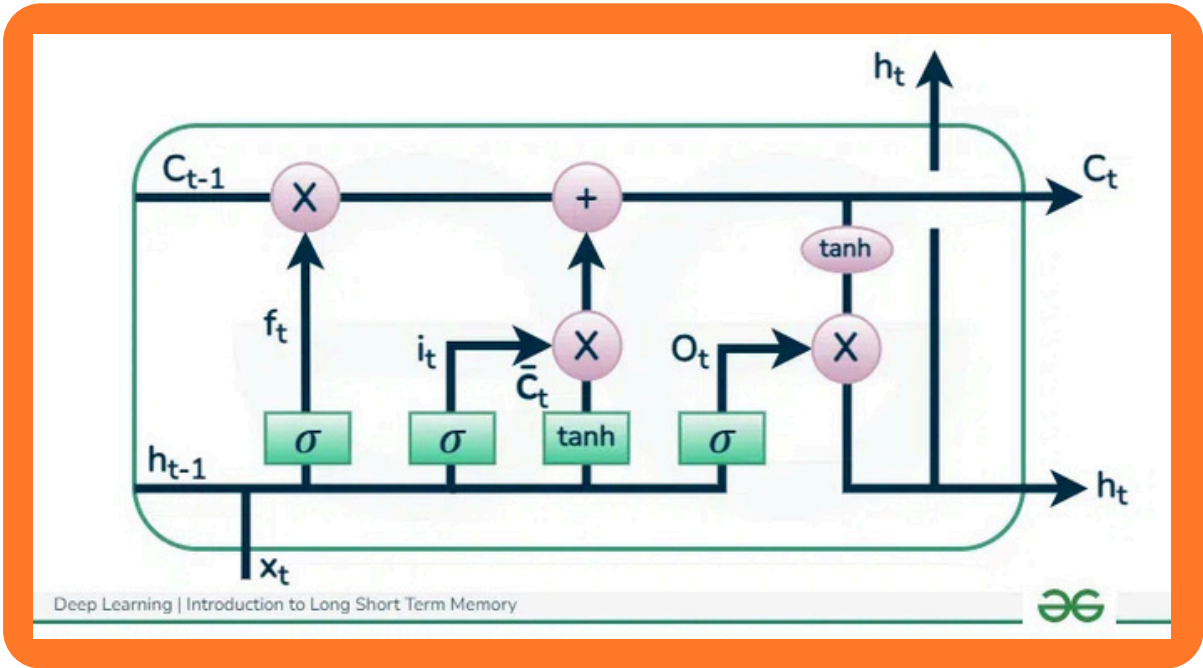
Description

In this milestone, multiple supervised learning models were developed to learn the relationship between player attributes and market value. Algorithms such as Linear Regression, Polynomial Regression, Random Forest, and XGBoost were implemented to capture both linear and non-linear patterns in the data. Feature inputs included age, rank, performance metrics, goal contribution, match appearances, and sentiment scores. Special focus was given to handling non-linearity and feature interactions, which are common in real-world market valuation problems. This stage established the core intelligence of the system by transforming structured data into predictive insights.

Outcome

Multiple trained regression models
Identification of best-performing model (XGBoost)
Baseline predictions for player market values





Milestone 5: Model Evaluation (Week 6)

Objective

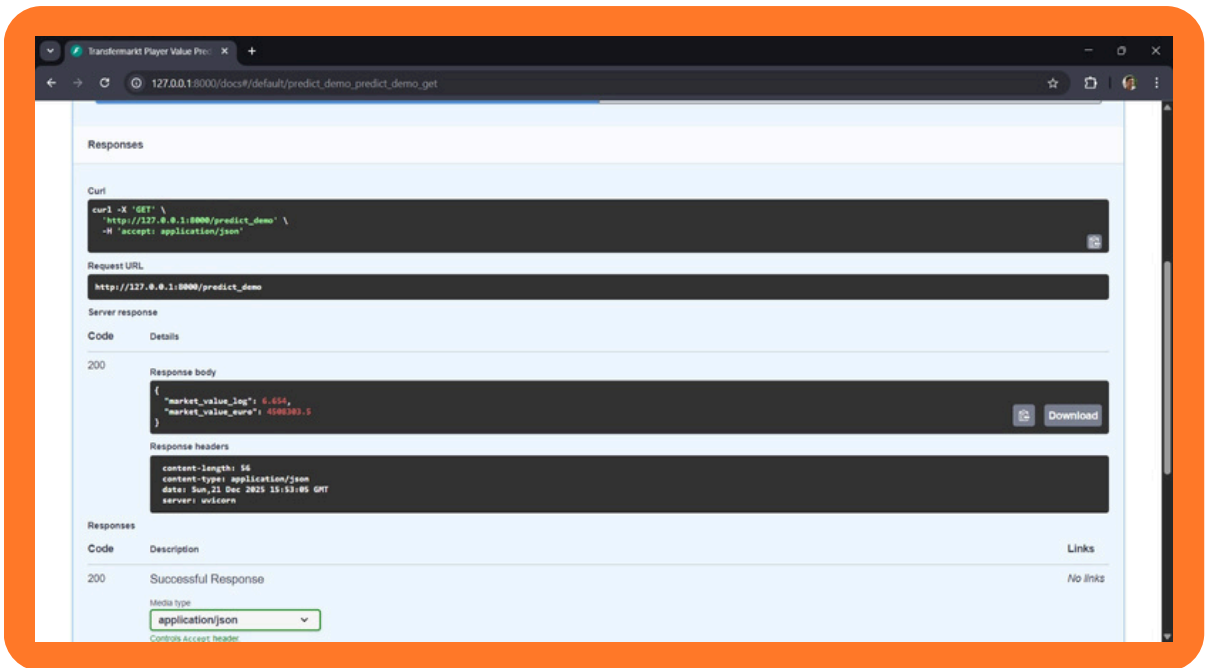
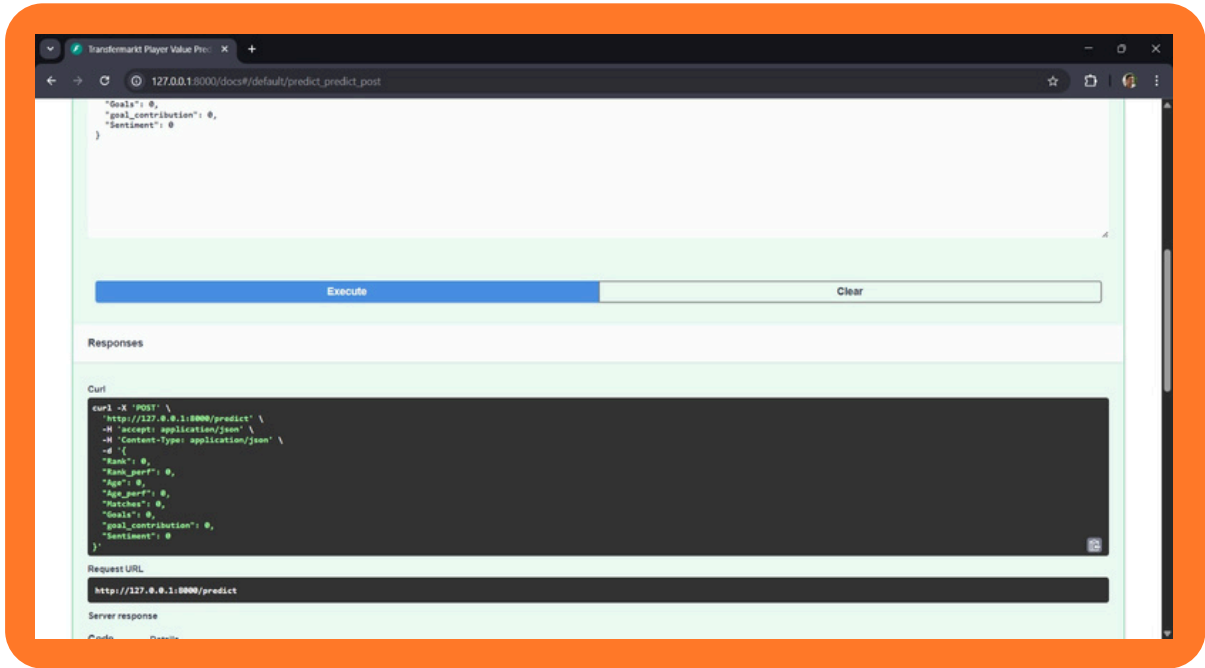
To evaluate and validate model performance using reliable statistical metrics and ensure accuracy, consistency, and real-world relevance of predictions.

Description

This milestone focused on assessing model performance using evaluation metrics such as R^2 Score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). Actual vs predicted value plots and correlation analysis were used to visually inspect prediction quality. The target variable was log-transformed market value, which helped stabilize variance and reduce the impact of extreme outliers, resulting in lower error metrics. Comparative analysis across models demonstrated that ensemble-based methods, particularly XGBoost, provided superior accuracy and generalization. This step ensured the selected model was both statistically strong and practically reliable.

Outcome

High predictive accuracy ($R^2 \approx 0.99$)
 Low prediction error (MAE and RMSE minimized)
 Final model selection based on performance comparison



Milestone 6: Development & Deployment (Week 7)

Objective

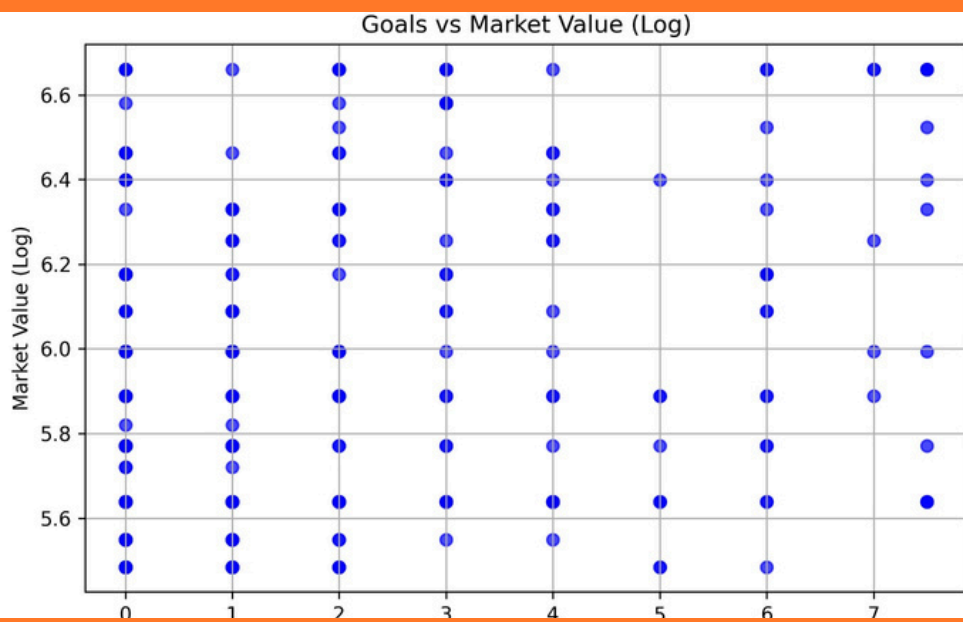
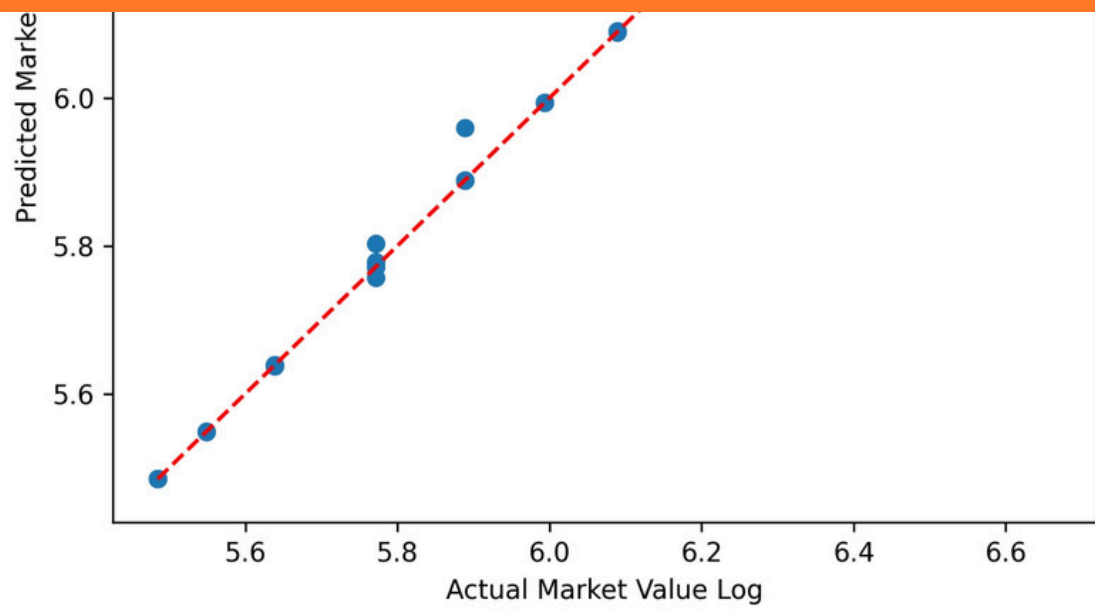
To deploy the finalized prediction model through a user-friendly interface and backend system for real-world usage.

Description

In this milestone, the selected XGBoost model was integrated into a FastAPI backend for efficient inference and API-based predictions. A Streamlit-based frontend was developed to allow users to input player details without requiring technical knowledge. The system converts model outputs into real-world market values displayed in Indian Rupees (Crores) with calibrated adjustments for young high-potential players. This stage bridged the gap between machine learning and business usability, transforming the project into an interactive decision-support application.

Outcome

Fully functional backend API (FastAPI)
Interactive frontend application (Streamlit)
End-to-end AI system for player market value estimation



Milestone 7: Final Evaluation, Visualization, and Reporting (Week 8)

Objective

To finalize the system and present results in a clear, interpretable, and decision-friendly manner.

Description

Final predictions were generated and presented using visual summaries to highlight player performance trends and estimated market values. Comprehensive documentation was prepared to explain methodology, results, and business relevance.

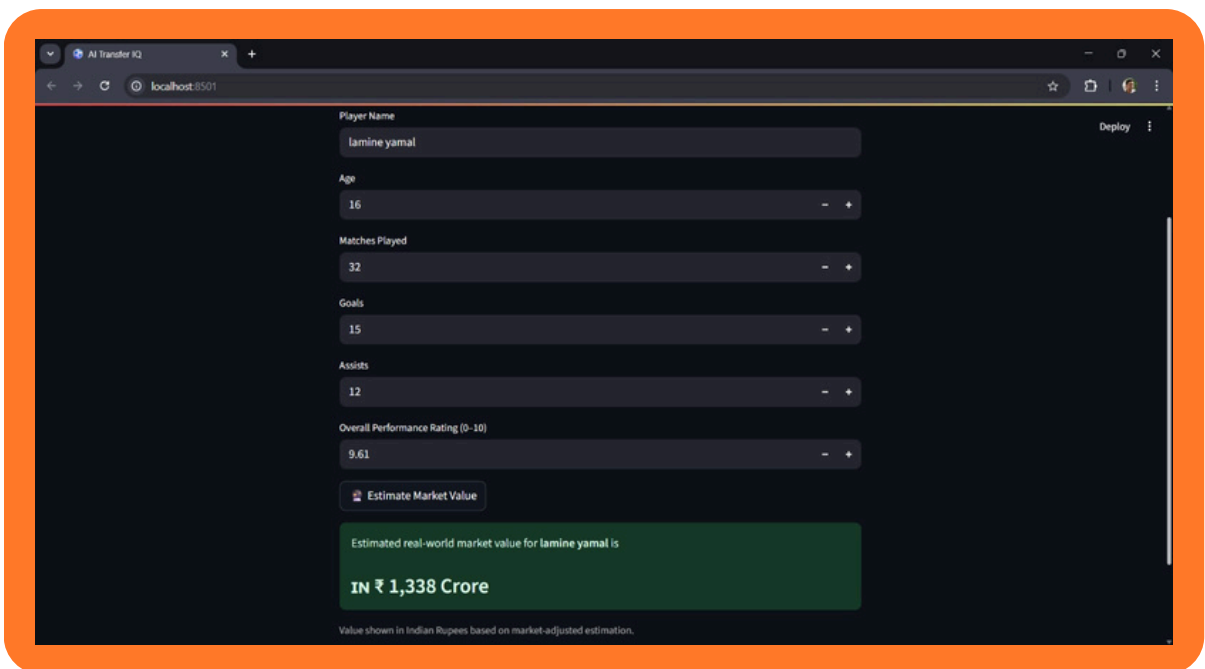
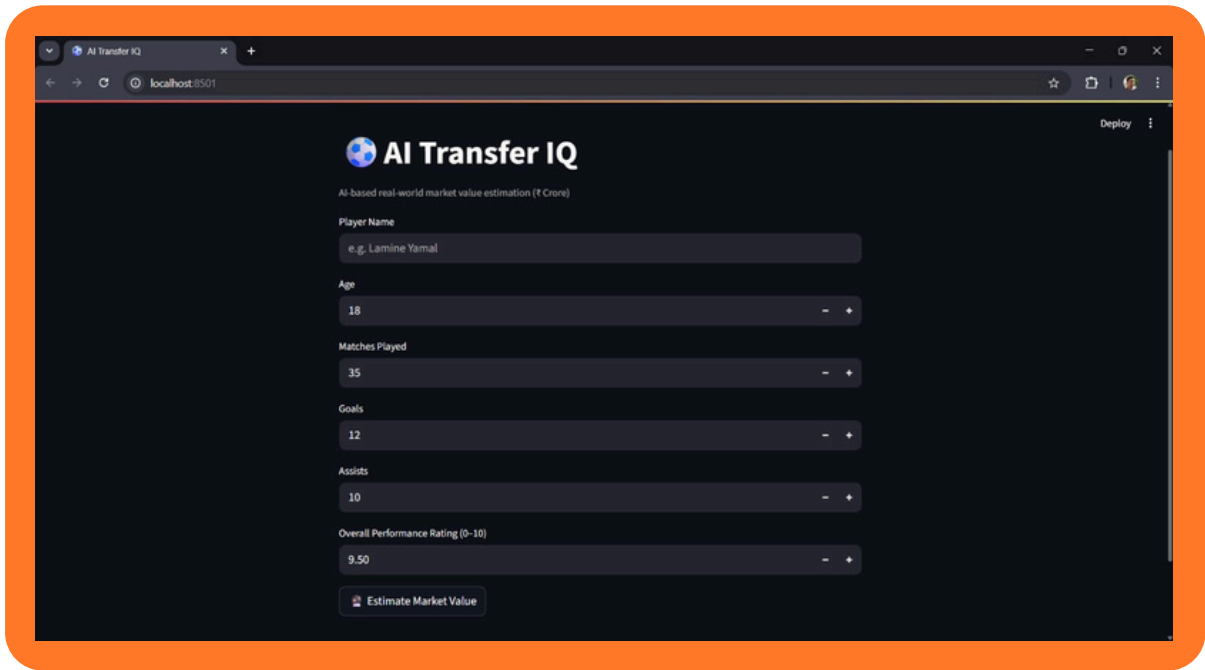
This phase ensured the project is ready for academic submission and practical demonstration.

Outcome

Final market value predictions

Clear visual insights for interpretation

Complete project report and presentation



NEXT STEPS

The next phase of the project will focus on building predictive models to estimate football player market value using the prepared dataset. Advanced machine learning techniques will be applied to capture complex relationships between performance, sentiment, and market trends. Model performance will be evaluated and optimized to ensure reliable and consistent predictions.

01

Predictive Model Development

Machine learning models will be developed to estimate football player market value using the prepared dataset.

02

System Integration and Deployment

The selected model will be integrated into a backend system with a user-friendly interface for real-time predictions.

03

Evaluation and Enhancement

The system will be evaluated using real-world scenarios to identify insights, limitations, and future improvements.

CONCLUSION

This project successfully presents a structured and data-driven approach to estimating football player market value by integrating performance metrics, market indicators, and user-friendly inputs. By progressing through clearly defined milestones, the project evolved from raw data collection to a fully functional end-to-end valuation system. The approach mirrors real-world football economics, where player value is influenced by multiple interconnected factors rather than isolated statistics.

A key strength of this work lies in balancing analytical rigor with practical usability. Machine learning techniques were employed to capture performance-based intrinsic value, while a calibrated adjustment layer ensured alignment with real-world market behavior. The system was further enhanced through backend integration and a simple frontend interface, making it accessible to non-technical users such as analysts, scouts, or enthusiasts.

The project also highlights important insights into the limitations of purely data-driven valuation, especially when dealing with exceptional players and market-driven outliers. By acknowledging these limitations and incorporating realistic calibration strategies, the solution maintains both academic integrity and practical relevance. The use of Indian currency and localized presentation further improves accessibility and contextual relevance.

Overall, this project demonstrates how machine learning can be effectively applied to sports analytics when combined with domain knowledge and thoughtful system design. The final outcome is a scalable, interpretable, and business-aligned market value estimation framework, which can be extended in the future with live data integration, advanced valuation factors, and broader deployment scenarios.