Dynamic Player Transfer Value Prediction System

Football player transfer valuation is one of the most critical tasks for clubs, scouts, and analysts. The objective of this project was to build a dynamic prediction system that leverages multiple data sources, including historical performance, injuries, transfer history, and sentiment analysis, to estimate player market value. The project was developed as part of the Infosys Internship challenge to deliver an AI-driven system capable of assisting in data-driven transfer decisions.

Data Understanding & Preprocessing
The dataset included features such as player performance (goals, assists, minutes played, matches, shots on target), transfer history (average fee, last fee, number of transfers), and additional social/sentiment features. Preprocessing steps included log-transforming financial values, scaling numerical attributes, and preparing time series structures for sequential models. Sentiment was encoded into three levels: Good, Average, and Bad, providing an additional qualitative input for modeling.

Model Development
The initial baseline was a Multivariate LSTM model trained on 13 features across 4 timesteps. This first version achieved RMSE = 0.7744, MAE = 0.6337, and R^2 = 0.1382, showing potential but limited accuracy. Challenges included input shape mismatches, feature misalignment, and difficulties in loading saved models across different formats (.h5 vs .keras). These issues were resolved by carefully ensuring consistent input dimensions and re-exporting models to stable formats.

Through hyperparameter tuning of the LSTM, results significantly improved, reaching RMSE = 0.6592, MAE = 0.493, and R^2 = 0.3756. This represented a strong improvement and established the LSTM as the best-performing approach.

Exploration of Ensemble Methods
Parallel to LSTM improvements, different ensemble methods (Bagging, Boosting, and Stacking) were studied. An XGBoost regression model was implemented to test ensemble approaches. Initially, performance was poor (RMSE = 0.8513, MAE = 0.6564, R^2 = -0.0415). However, after systematic hyperparameter tuning, the results improved to RMSE = 0.7716, MAE = 0.6146, and R^2 = 0.1444. While still underperforming compared to the tuned LSTM, this highlighted the importance of architecture choice for sequential, time-series structured data.

Deployment via Streamlit
The final system was deployed as a Streamlit web application. It provides an interface for users to input player details, including performance stats, transfer history, and sentiment score. The backend uses the improved LSTM model to predict the player's market value dynamically. Difficulties faced during deployment included aligning feature order, encoding sentiment consistently, and ensuring predictions worked seamlessly even for players not in the training dataset. These were solved by building a robust input pipeline and default handling mechanisms.

Final Outcomes & Insights
- Best performing model: Improved Multivariate LSTM.
- Final results: RMSE = 0.6592, MAE = 0.493, R^2 = 0.3756.
- Ensemble models like XGBoost provided additional insights but underperformed compared to LSTM.
- Streamlit app enabled real-time predictions with a simple and intuitive user interface.

This project demonstrates the potential of AI to enhance the football transfer market by integrating performance, financial, and sentiment data into a unified predictive framework.

Conclusion & Impact

The Dynamic Player Transfer Value Prediction System successfully showcases how machine learning can aid clubs, analysts, and scouts in making data-driven decisions. While the LSTM model achieved promising accuracy, future improvements could include larger datasets, incorporating real injury timelines, richer sentiment analysis, and advanced ensemble stacking. The project journey reflects overcoming technical challenges, learning advanced AI concepts, and building a working deployment pipeline that bridges research and real-world usability.