“Expected Assist (xA) Estimation in Football Using Deep Learning and Contextual Modeling"

ABSTRACT

In modern football analytics, Expected Assists (xA) provide a quantitative assessment of a player's ability to create goal-scoring opportunities through passes. Traditional approaches equate xA with the Expected Goals (xG) of the resulting shot, neglecting important contextual elements such as the passer’s positioning, intent, and pressure. This paper introduces a two-stage deep learning framework to estimate xA more accurately. First, we model the probability that a given pass leads to a shot using spatial and contextual features. Second, we estimate the expected quality (xG) of that shot. The final xA is computed as the product of these two components, enhancing interpretability and tactical relevance. We train a fully connected neural network on a clustered pass dataset with pseudo-xA labels and integrate it into a visual rating system using Streamlit, analyzing Bayern Munich players’ contributions over a season.

---

1. INTRODUCTION

Expected Assists (xA) have emerged as a cornerstone metric in football analytics, offering insight into a player’s creative output. However, most implementations oversimplify the metric by assigning the xA value of a pass as equal to the xG of the resulting shot. This undermines the influence of the passer’s positioning, decision-making, and the defensive pressure applied during the pass. Our work proposes a context-aware, data-driven model for xA estimation that decouples the likelihood of a shot from the quality of that shot. The objective is to provide a richer understanding of player contributions beyond goals and assists, especially for midfielders and fullbacks whose creative actions often go unnoticed in traditional stats.

---

2. RELATED WORK

Earlier works have primarily focused on:

- **xG models**: Logistic regression or deep learning models predicting goal probability from shot features (e.g., StatsBomb, Opta).
- **Basic xA**: Assigning xA = xG for the resulting shot, without modeling pass-specific likelihood.
- **DeepPass** and other neural architectures for pass classification using positional embeddings.

Unlike prior approaches, our model introduces an intermediate probability model that estimates **P(shot | pass)**, allowing xA to reflect both pass quality and intent.

---

3. DATASET AND PREPROCESSING

We use three datasets:

- **passes_data.csv**: Includes positional (x, y), pass outcome, angle, length, play pattern, and xG of follow-up shots.
- **final_ratings.csv**: Player-wise ratings per match.
- **`bayern_match_summary.csv`:** Match-level breakdown of Bayern Munich player performances.

We have scraped the following data from StatsBomb.

3.1 Encoding and Scaling

- **Categorical encoding**: One-hot encoded **pass_technique, pass_type, and play_pattern.**

- **Feature scaling**: Min-max normalization applied to positional and length features for model stability.

---

4. METHODOLOGY

Our xA estimation pipeline has three core components:

4.1 Clustering with DBSCAN

To estimate pseudo-xA, we assume similar passes should lead to similar scoring chances. We use DBSCAN to group passes based on:

- Start and end positions

- Pass length, angle, pressure

- Play pattern context

Each cluster is assigned the mean xG of the shots resulting from its passes. This serves as a **pseudo-xA label** for supervised training.

4.2 Neural Network Architecture

We train a **Fully Connected Neural Network (FCNN)** to regress pseudo-xA. Input features include:

- Encoded categorical features

- Scaled numerical features

- Under pressure and spatial dynamics

Loss: Mean Squared Error (MSE)
Optimizer: Adam
Activation: ReLU with Dropout for regularization

4.3 Predictive Model

A function takes new pass inputs and computes:

$$xA = P(\text{pass} \rightarrow \text{shot}) \times E[xG \mid \text{shot}]$$

ensuring the model accounts for both the strategic value and execution quality of the pass.

---

5. APPLICATION:PLAYER RATING UI

Using Streamlit, we built a real-time rating interface for Bayern Munich players. Each player is evaluated by:

- Total xA contribution
- Actual goal involvement (xG + assists)
- Pass completion rate
- Rating standardization using z-scores

This allows users to visually compare player performances across matches and identify under-the-radar contributors.

---

6. RESULTS AND OBSERVATIONS

Our two-stage deep learning framework for estimating Expected Assists (xA) was evaluated using match data from Bayern Munich's season. The key findings include:

- **Enhanced Player Assessment**: The model effectively identifies players who consistently deliver high-quality passes leading to goal-scoring opportunities, even if these do not result in actual assists. This aligns with the approach of assigning credit to creators based on the quality of chances they generate, as discussed in Stats Perform's analysis.Stats Perform

- **Effective Clustering with DBSCAN**: Utilizing DBSCAN for clustering passes based on spatial and contextual features proved effective. This method aligns with the findings of Mead et al. (2023), who emphasized the importance of incorporating contextual features in xG models.PLOS

- **User-Friendly Interface**: The Streamlit-based UI allows for intuitive visualization of player ratings and performance metrics, facilitating better understanding and analysis for coaches and analysts.

---

7. LIMITATIONS

While the model shows promise, certain limitations were identified:

- **Reliance on Pseudo-xA Labels**: The use of pseudo-xA labels, derived from clustering, may not capture the full complexity of in-game scenarios. Future work could involve integrating actual shot outcome data to refine these labels.

- **DBSCAN Parameter Sensitivity**: The effectiveness of DBSCAN clustering is sensitive to parameter selection. Improper tuning can lead to suboptimal clustering results.

- **Limited Scope of Data**: The current model does not account for dynamic factors such as player movements, defensive pressure, and off-the-ball runs, which are crucial in real-game situations.

---

8. FUTURE WORK

To address the identified limitations and enhance the model's applicability:

- **Integration of Sequential Event Data**: Incorporating sequential event data can provide a more accurate representation of play development, as suggested by Yeung et al. (2023), who proposed a Transformer-Based Neural Marked Spatio Temporal Point Process Model for football match events analysis.arXiv

- **Inclusion of Tracking Data**: Utilizing player tracking data can offer insights into player positioning, movement patterns, and spatial control, contributing to a more comprehensive xA estimation.

- **Application of Advanced Machine Learning Techniques**: Exploring advanced machine learning models, such as gradient boosting and deep neural networks, can enhance prediction accuracy, as demonstrated in Hewitt and Karakuş (2023).arXiv

---

9. CONCLUSION

This study presents a novel approach to estimating Expected Assists (xA) in football by combining clustering techniques with deep learning models. By considering both the likelihood of a pass leading to a shot and the quality of the resulting shot, the model provides a more nuanced assessment of a player's creative contributions. The integration of this model into a user-friendly interface demonstrates its potential for practical application in player performance analysis and scouting.

---

10. REFERENCES

1. Mead, J., O'Hare, A., & McMenemy, P. (2023). Expected goals in football: Improving model performance and demonstrating value. *PLOS ONE*, 18(4), e0282295.PLOS+1ResearchGate+1

2. Yeung, C. C. K., Sit, T., & Fujii, K. (2023). Transformer-Based Neural Marked Spatio Temporal Point Process Model for Football Match Events Analysis. *arXiv preprint arXiv:2302.09276*.arXiv

3. Hewitt, J. H., & Karakuş, O. (2023). A Machine Learning Approach for Player and Position Adjusted Expected Goals in Football (Soccer). *arXiv preprint arXiv:2301.13052*.ScienceDirect+2arXiv+2Data Science Association+2

4. Stats Perform. (2017). Expected assists in context. Retrieved from https://www.statsperform.com/resource/expected-assists-in-context-2/Stats Perform