

Phân tích các feature cho vị trí Tiền vệ (MF) trong dự đoán giá trị chuyển nhượng (ETV)

Tác giả: Grok 3, được tạo bởi xAI

Ngày: 3 tháng 5 năm 2025

Tài liệu này phân tích các feature được chọn để dự đoán giá trị chuyển nhượng ước tính (ETV) cho vị trí Tiền vệ trong bóng đá, sử dụng dữ liệu từ tệp result.csv.

Mục lục

- 1 Tổng quan về feature của tiền vệ
- 2 Phân tích chi tiết từng feature
 - 2.1 Chuyển tạo ra cơ hội ghi bàn (Key Passes)
 - 2.1.1 Định nghĩa
 - 2.1.2 Lý do chọn
 - 2.1.3 Cách xử lý trong mã
 - 2.1.4 Ví dụ minh họa
 - 2.2 Đường chuyền tiến công (PrgP - Progressive Passes)
 - 2.2.1 Định nghĩa
 - 2.2.2 Lý do chọn
 - 2.2.3 Cách xử lý trong mã
 - 2.2.4 Ví dụ minh họa
 - 2.3 Hành động tạo cơ hội (SCA - Shot-Creating Actions)
 - 2.3.1 Định nghĩa
 - 2.3.2 Lý do chọn
 - 2.3.3 Cách xử lý trong mã
 - 2.3.4 Ví dụ minh họa
 - 2.4 Kiến tạo kỳ vọng (xAG - Expected Assisted Goals)
 - 2.4.1 Định nghĩa
 - 2.4.2 Lý do chọn
 - 2.4.3 Cách xử lý trong mã
 - 2.4.4 Ví dụ minh họa
 - 2.5 Tỷ lệ chuyền chính xác (Cmp% - Pass Completion Percentage)
 - 2.5.1 Định nghĩa
 - 2.5.2 Lý do chọn
 - 2.5.3 Cách xử lý trong mã
 - 2.5.4 Ví dụ minh họa
 - 2.6 Số đường chuyền vào vòng cấm đối phương (PPA - Passes into Penalty Area)
 - 2.6.1 Định nghĩa
 - 2.6.2 Lý do chọn
 - 2.6.3 Cách xử lý trong mã
 - 2.6.4 Ví dụ minh họa
 - 2.7 Số đường chuyền vào 1/3 sân cuối (Pass_into_1/3)
 - 2.7.1 Định nghĩa
 - 2.7.2 Lý do chọn

- 2.7.3 Cách xử lý trong mã
- 2.7.4 Ví dụ minh họa
- 2.8 Số pha kiến tạo (Ast - Assists)
 - 2.8.1 Định nghĩa
 - 2.8.2 Lý do chọn
 - 2.8.3 Cách xử lý trong mã
 - 2.8.4 Ví dụ minh họa
- 2.9 Số lần chạm bóng (Touches)
 - 2.9.1 Định nghĩa
 - 2.9.2 Lý do chọn
 - 2.9.3 Cách xử lý trong mã
 - 2.9.4 Ví dụ minh họa
- 2.10 Số lần mang bóng vào 1/3 sân cuối (Carries 1_3)
 - 2.10.1 Định nghĩa
 - 2.10.2 Lý do chọn
 - 2.10.3 Cách xử lý trong mã
 - 2.10.4 Ví dụ minh họa
- 2.11 Tổng khoảng cách bóng được đưa lên phía trước (ProDist - Progressive Distance)
 - 2.11.1 Định nghĩa
 - 2.11.2 Lý do chọn
 - 2.11.3 Cách xử lý trong mã
 - 2.11.4 Ví dụ minh họa
- 2.12 Số lần nhận bóng thành công (Rec - Receptions)
 - 2.12.1 Định nghĩa
 - 2.12.2 Lý do chọn
 - 2.12.3 Cách xử lý trong mã
 - 2.12.4 Ví dụ minh họa
- 2.13 Số lần mất bóng do đối thủ đoạt được (Dis - Dispossessed)
 - 2.13.1 Định nghĩa
 - 2.13.2 Lý do chọn
 - 2.13.3 Cách xử lý trong mã
 - 2.13.4 Ví dụ minh họa
- 2.14 Số lần xử lý bóng lỗi (Mis - Miscontrols)
 - 2.14.1 Định nghĩa
 - 2.14.2 Lý do chọn
 - 2.14.3 Cách xử lý trong mã
 - 2.14.4 Ví dụ minh họa

- 3 Tại sao các feature này được chọn cùng nhau?
- 4 Cách các feature tương tác với mô hình
- 5 Hạn chế và đề xuất cải thiện
- 6 Kết luận

1. Tổng quan về feature của tiền vệ

Trong bài toán dự đoán giá trị chuyển nhượng ước tính (ETV) cho vị trí Tiền vệ (MF), các feature được chọn cần phản ánh các khía cạnh chính của vai trò tiền vệ, bao gồm:

- Khả năng sáng tạo: Tạo cơ hội ghi bàn thông qua chuyền bóng chính xác, chuyền tạo cơ hội, và kiến tạo.
- Khả năng tiến công: Đưa bóng lên phía trước qua chuyền hoặc mang bóng, đặc biệt vào các khu vực nguy hiểm như 1/3 sân cuối hoặc vòng cấm.
- Khả năng kiểm soát bóng: Tham gia vào lối chơi thông qua số lần chạm bóng và nhận bóng thành công.
- Độ tin cậy kỹ thuật: Giảm thiểu sai lầm như mất bóng hoặc xử lý bóng lỗi.

Các feature được chia thành:

- Feature liên quan tới sáng tạo: Chuyền tạo ra cơ hội ghi bàn (Key Passes), Hành động tạo cơ hội (SCA), Kiến tạo kỳ vọng (xAG), Số pha kiến tạo (Ast).
- Feature liên quan tới tiến công: Đường chuyền tiến công (PrgP), Số đường chuyền vào vòng cấm đối phương (PPA), Số đường chuyền vào 1/3 sân cuối (Pass_into_1/3), Số lần mang bóng vào 1/3 sân cuối (Carries 1_3), Tổng khoảng cách bóng được đưa lên phía trước (ProDist).
- Feature liên quan tới kiểm soát bóng: Tỷ lệ chuyền chính xác (Cmp%), Số lần chạm bóng (Touches), Số lần nhận bóng thành công (Rec).
- Feature liên quan tới kỹ thuật tiêu cực: Số lần mất bóng do đối thủ đoạt được (Dis), Số lần xử lý bóng lỗi (Mis).

Lý do chọn: Các feature này bao quát vai trò của tiền vệ, phản ánh đúng giá trị thị trường, và có sẵn trong tệp result.csv. Chúng được chọn dựa trên:

- Liên quan đến vai trò tiền vệ: Tiền vệ cần sáng tạo, tiến công, kiểm soát bóng tốt, và giảm thiểu sai lầm.
- Tầm quan trọng trong thị trường chuyển nhượng: Các chỉ số này là yếu tố chính mà các CLB xem xét khi định giá tiền vệ.
- Dữ liệu sẵn có: Các feature được thu thập từ thống kê bóng đá tiêu chuẩn.

2. Phân tích chi tiết từng feature

2.1. Chuyển tạo ra cơ hội ghi bàn (Key Passes)

2.1.1. Định nghĩa

Chuyển tạo ra cơ hội ghi bàn (Key Passes) là số đường chuyền dẫn trực tiếp đến một cú sút của đồng đội. Ví dụ: Một tiền vệ thực hiện 60 đường chuyền tạo cơ hội trong mùa giải.

2.1.2. Lý do chọn

- Cốt lõi của sự sáng tạo: Key Passes thể hiện khả năng tạo cơ hội ghi bàn, là kỹ năng quan trọng của tiền vệ tấn công.
- Tương quan với ETV: Tiền vệ có Key Passes cao (như Kevin De Bruyne) thường có ETV cao hơn, vì họ trực tiếp đóng góp vào bàn thắng.
- Phổ biến: Key Passes là chỉ số tiêu chuẩn, có sẵn trong result.csv.

2.1.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Chuyển thành số, điền giá trị thiếu bằng trung vị (hoặc 0 nếu trung vị là NaN).
- Biến đổi log: Áp dụng `np.log1p` để giảm độ lệch, vì Key Passes có phân phối lệch.
- Tăng trọng số: Là `important_features`, được nhân với 2.0 để nhấn mạnh vai trò sáng tạo.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler` để đưa về thang đo tương đương.

2.1.4. Ví dụ minh họa

Tiền vệ A có Key Passes = 70 và tiền vệ B có Key Passes = 30. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta tạo nhiều cơ hội hơn.

2.2. Đường chuyền tiến công (PrgP - Progressive Passes)

2.2.1. Định nghĩa

Đường chuyền tiến công (PrgP) là số đường chuyền đưa bóng tiến gần hơn đến khung thành đối phương. Ví dụ: Tiền vệ thực hiện 200 đường chuyền tiến công trong mùa giải.

2.2.2. Lý do chọn

- Khởi tạo tấn công: PrgP thể hiện khả năng đưa bóng lên phía trước, đặc biệt quan trọng với tiền vệ trung tâm.
- Tương quan với ETV: Tiền vệ có PrgP cao thường được định giá cao hơn, vì họ hỗ trợ lối chơi tấn công.
- Phổ biến: PrgP là chỉ số nâng cao, có sẵn trong result.csv.

2.2.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Tăng trọng số: Là `important_features`, được nhân với 2.0.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.2.4. Ví dụ minh họa

Tiền vệ A có $\text{PrgP} = 250$ và tiền vệ B có $\text{PrgP} = 100$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta đóng góp nhiều vào tấn công.

2.3. Hành động tạo cơ hội (SCA - Shot-Creating Actions)

2.3.1. Định nghĩa

Hành động tạo cơ hội (SCA) là số hành động dẫn đến một cú sút, bao gồm chuyền bóng, rê bóng, hoặc phạm lỗi dẫn đến sút. Ví dụ: Tiền vệ thực hiện 150 hành động tạo cơ hội trong mùa giải.

2.3.2. Lý do chọn

- Đo lường sự sáng tạo tổng thể: SCA bao quát các hành động sáng tạo, không chỉ chuyền bóng.
- Tương quan với ETV: Tiền vệ có SCA cao thường có ETV cao hơn, vì họ đóng góp vào các pha tấn công.
- Phổ biến: SCA là chỉ số nâng cao, có sẵn trong result.csv.

2.3.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.

- Tăng trọng số: Là `important_features`, được nhân với 2.0.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.3.4. Ví dụ minh họa

Tiền vệ A có $SCA = 180$ và tiền vệ B có $SCA = 90$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta sáng tạo hơn.

2.4. Kiến tạo kỳ vọng (xAG - Expected Assisted Goals)

2.4.1. Định nghĩa

Kiến tạo kỳ vọng (xAG) là giá trị kỳ vọng của số bàn thắng từ các đường chuyền của tiền vệ, dựa trên chất lượng cơ hội. Ví dụ: Tiền vệ có $xAG = 8.5$ trong mùa giải.

2.4.2. Lý do chọn

- Đo lường chất lượng sáng tạo: xAG tập trung vào chất lượng cơ hội, không chỉ số lượng.
- Tương quan với ETV: Tiền vệ có xAG cao thường có ETV cao hơn, vì họ tạo cơ hội chất lượng.
- Phổ biến: xAG là chỉ số nâng cao, có sẵn trong `result.csv`.

2.4.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Tăng trọng số: Là `important_features`, được nhân với 2.0.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.4.4. Ví dụ minh họa

Tiền vệ A có $xAG = 10.0$ và tiền vệ B có $xAG = 4.0$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta tạo cơ hội chất lượng hơn.

2.5. Tỷ lệ chuyền chính xác (Cmp% - Pass Completion Percentage)

2.5.1. Định nghĩa

Tỷ lệ chuyền chính xác (Cmp%) là tỷ lệ phần trăm đường chuyền thành công:

$$\text{Cmp\%} = \frac{\text{Số đường chuyền hoàn thành}}{\text{Tổng số đường chuyền}} \times 100$$

Ví dụ: Tiền vệ hoàn thành 1,200/1,300 đường chuyền, thì $\text{Cmp\%} = 92.3\%$.

2.5.2. Lý do chọn

- Đo lường độ tin cậy: Cmp% thể hiện khả năng kiểm soát bóng và chuyền bóng chính xác.
- Tương quan với ETV: Tiền vệ có Cmp% cao thường được định giá cao hơn, đặc biệt ở các đội chơi kiểm soát bóng.
- Phổ biến: Cmp% là chỉ số tiêu chuẩn, có sẵn trong result.csv.

2.5.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Không tăng trọng số: Không nằm trong `important_features`.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.5.4. Ví dụ minh họa

Tiền vệ A có $\text{Cmp\%} = 94\%$ và tiền vệ B có $\text{Cmp\%} = 85\%$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta đáng tin cậy hơn khi chuyền bóng.

2.6. Số đường chuyền vào vòng cấm đối phương (PPA - Passes into Penalty Area)

2.6.1. Định nghĩa

Số đường chuyền vào vòng cấm đối phương (PPA) là số đường chuyền đưa bóng vào khu vực 16m50 của đối phương. Ví dụ: Tiền vệ thực hiện 50 đường chuyền vào vòng cấm trong mùa giải.

2.6.2. Lý do chọn

- Tạo nguy hiểm trực tiếp: PPA thể hiện khả năng đưa bóng vào khu vực nguy hiểm.

- Tương quan với ETV: Tiền vệ có PPA cao thường được định giá cao hơn, vì họ tạo cơ hội gần khung thành.
- Phổ biến: PPA là chỉ số nâng cao, có sẵn trong result.csv.

2.6.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Tăng trọng số: Là `important_features`, được nhân với 2.0.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.6.4. Ví dụ minh họa

Tiền vệ A có $PPA = 60$ và tiền vệ B có $PPA = 20$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta tạo nguy hiểm nhiều hơn.

2.7. Số đường chuyền vào 1/3 sân cuối (`Pass_into_1/3`)

2.7.1. Định nghĩa

Số đường chuyền vào 1/3 sân cuối (`Pass_into_1/3`) là số đường chuyền đưa bóng vào 1/3 sân gần khung thành đối phương. Ví dụ: Tiền vệ thực hiện 100 đường chuyền vào 1/3 sân cuối trong mùa giải.

2.7.2. Lý do chọn

- Thúc đẩy tấn công: `Pass_into_1/3` thể hiện khả năng đưa bóng vào khu vực tấn công.
- Tương quan với ETV: Tiền vệ có `Pass_into_1/3` cao thường được định giá cao hơn.
- Phổ biến: `Pass_into_1/3` là chỉ số nâng cao, có sẵn trong result.csv.

2.7.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Tăng trọng số: Là `important_features`, được nhân với 2.0.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.7.4. Ví dụ minh họa

Tiền vệ A có $\text{Pass_into_1/3} = 120$ và tiền vệ B có $\text{Pass_into_1/3} = 50$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta đưa bóng vào khu vực tấn công nhiều hơn.

2.8. Số pha kiến tạo (Ast - Assists)

2.8.1. Định nghĩa

Số pha kiến tạo (Ast) là số lần tiền vệ chuyền bóng dẫn đến bàn thắng trực tiếp. Ví dụ: Tiền vệ có 12 pha kiến tạo trong mùa giải.

2.8.2. Lý do chọn

- Kết quả sáng tạo: Ast đo lường hiệu quả cuối cùng của khả năng sáng tạo.
- Tương quan với ETV: Tiền vệ có Ast cao thường có ETV cao hơn, vì họ trực tiếp góp phần vào bàn thắng.
- Phổ biến: Ast là chỉ số tiêu chuẩn, có sẵn trong result.csv.

2.8.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng np.log1p .
- Tăng trọng số: Là `important_features`, được nhân với 2.0.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.8.4. Ví dụ minh họa

Tiền vệ A có $\text{Ast} = 15$ và tiền vệ B có $\text{Ast} = 5$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta góp phần vào nhiều bàn thắng hơn.

2.9. Số lần chạm bóng (Touches)

2.9.1. Định nghĩa

Số lần chạm bóng (Touches) là số lần tiền vệ chạm vào bóng trong trận đấu. Ví dụ: Tiền vệ có 2,000 lần chạm bóng trong mùa giải.

2.9.2. Lý do chọn

- Tham gia lối chơi: Touches thể hiện mức độ tham gia vào trận đấu.

- Tương quan với ETV: Tiền vệ có Touches cao thường được định giá cao hơn, vì họ kiểm soát trận đấu.
- Phổ biến: Touches là chỉ số tiêu chuẩn, có sẵn trong result.csv.

2.9.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Không tăng trọng số: Không nằm trong `important_features`.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.9.4. Ví dụ minh họa

Tiền vệ A có Touches = 2,500 và tiền vệ B có Touches = 1,500. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta tham gia nhiều hơn.

2.10. Số lần mang bóng vào 1/3 sân cuối (Carries 1_3)

2.10.1. Định nghĩa

Số lần mang bóng vào 1/3 sân cuối (Carries 1_3) là số lần tiền vệ rê bóng hoặc mang bóng vào 1/3 sân gần khung thành đối phương. Ví dụ: Tiền vệ thực hiện 80 lần mang bóng vào 1/3 sân cuối trong mùa giải.

2.10.2. Lý do chọn

- Tấn công trực tiếp: Carries 1_3 thể hiện khả năng tự mình đưa bóng vào khu vực nguy hiểm.
- Tương quan với ETV: Tiền vệ có Carries 1_3 cao thường được định giá cao hơn.
- Phổ biến: Carries 1_3 là chỉ số nâng cao, có sẵn trong result.csv.

2.10.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Tăng trọng số: Là `important_features`, được nhân với 2.0.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.10.4. Ví dụ minh họa

Tiền vệ A có Carries $1_3 = 100$ và tiền vệ B có Carries $1_3 = 40$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta tấn công trực tiếp hơn.

2.11. Tổng khoảng cách bóng được đưa lên phía trước (ProDist - Progressive Distance)

2.11.1. Định nghĩa

Tổng khoảng cách bóng được đưa lên phía trước (ProDist) là tổng khoảng cách (tính bằng mét) mà bóng được đưa lên phía trước qua chuyền hoặc mang bóng. Ví dụ: Tiền vệ có ProDist = 5,000 mét trong mùa giải.

2.11.2. Lý do chọn

- Hiệu quả tiến công: ProDist đo lường mức độ đóng góp vào việc đưa bóng lên phía trước.
- Tương quan với ETV: Tiền vệ có ProDist cao thường được định giá cao hơn.
- Phổ biến: ProDist là chỉ số nâng cao, có sẵn trong result.csv.

2.11.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Tăng trọng số: Là `important_features`, được nhân với 2.0.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.11.4. Ví dụ minh họa

Tiền vệ A có ProDist = 6,000 mét và tiền vệ B có ProDist = 3,000 mét. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta đưa bóng lên phía trước nhiều hơn.

2.12. Số lần nhận bóng thành công (Rec - Receptions)

2.12.1. Định nghĩa

Số lần nhận bóng thành công (Rec) là số lần tiền vệ nhận bóng từ đồng đội mà không mất kiểm soát. Ví dụ: Tiền vệ nhận bóng thành công 1,500 lần trong mùa giải.

2.12.2. Lý do chọn

- Khả năng kiểm soát bóng: Rec thể hiện khả năng tham gia vào lối chơi và duy trì quyền kiểm soát.

- Tương quan với ETV: Tiền vệ có Rec cao thường được định giá cao hơn.
- Phổ biến: Rec là chỉ số tiêu chuẩn, có sẵn trong result.csv.

2.12.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Không tăng trọng số: Không nằm trong `important_features`.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.12.4. Ví dụ minh họa

Tiền vệ A có $\text{Rec} = 1,800$ và tiền vệ B có $\text{Rec} = 1,200$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta tham gia lối chơi nhiều hơn.

2.13. Số lần mất bóng do đối thủ đoạt được (Dis - Dispossessed)

2.13.1. Định nghĩa

Số lần mất bóng do đối thủ đoạt được (Dis) là số lần tiền vệ mất bóng do bị đối thủ cướp bóng. Ví dụ: Tiền vệ bị cướp bóng 25 lần trong mùa giải.

2.13.2. Lý do chọn

- Đo lường sai lầm kỹ thuật: Dis phản ánh độ tin cậy khi kiểm soát bóng.
- Tương quan với ETV: Tiền vệ có Dis thấp thường được định giá cao hơn, vì họ ít gây sai lầm.
- Phổ biến: Dis là chỉ số tiêu chuẩn, có sẵn trong result.csv.

2.13.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Không tăng trọng số: Không nằm trong `important_features`.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.13.4. Ví dụ minh họa

Tiền vệ A có $\text{Dis} = 15$ và tiền vệ B có $\text{Dis} = 35$. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta ít bị cướp bóng.

2.14. Số lần xử lý bóng lỗi (Mis - Miscontrols)

2.14.1. Định nghĩa

Số lần xử lý bóng lỗi (Mis) là số lần tiền vệ kiểm soát bóng không tốt, dẫn đến mất bóng mà không phải do đối thủ cướp. Ví dụ: Tiền vệ xử lý bóng lỗi 20 lần trong mùa giải.

2.14.2. Lý do chọn

- Do lường độ tin cậy kỹ thuật: Mis phản ánh kỹ năng kiểm soát bóng.
- Tương quan với ETV: Tiền vệ có Mis thấp thường có ETV cao hơn, vì họ ít gây sai lầm.
- Phổ biến: Mis là chỉ số tiêu chuẩn, có sẵn trong result.csv.

2.14.3. Cách xử lý trong mã

- Chuyển đổi và điền giá trị thiếu: Tương tự Key Passes.
- Biến đổi log: Áp dụng `np.log1p`.
- Không tăng trọng số: Không nằm trong `important_features`.
- Chuẩn hóa: Chuẩn hóa bằng `StandardScaler`.

2.14.4. Ví dụ minh họa

Tiền vệ A có Mis = 10 và tiền vệ B có Mis = 30. Tiền vệ A có khả năng được định giá cao hơn, vì anh ta kiểm soát bóng tốt hơn.

3. Tại sao các feature này được chọn cùng nhau?

- Phản ánh toàn diện vai trò tiền vệ:
 - Sáng tạo: Key Passes, SCA, xAG, Ast đo lường khả năng tạo cơ hội và kiến tạo.
 - Tiến công: PrgP, PPA, Pass_into_1/3, Carries 1_3, ProDist thể hiện khả năng đưa bóng vào khu vực nguy hiểm.
 - Kiểm soát bóng: Cmp%, Touches, Rec đo lường khả năng tham gia và duy trì quyền kiểm soát.
 - Độ tin cậy kỹ thuật: Dis, Mis đánh giá khả năng tránh sai lầm.
- Bổ sung lẫn nhau:

- Key Passes và SCA đo lường số lượng, xAG và Ast đo lường chất lượng sáng tạo.
- PrgP, PPA, Pass_into_1/3, Carries 1_3, ProDist cung cấp các khía cạnh khác nhau của tiến công.
- Cmp%, Touches, Rec đo lường mức độ tham gia và độ tin cậy.
- Dis và Mis cung cấp góc nhìn tiêu cực, bổ sung cho các chỉ số tích cực.
- Tương quan với thị trường chuyển nhượng: Tiền vệ toàn diện (sáng tạo, tấn công, kiểm soát bóng tốt, ít sai lầm) như Kevin De Bruyne có ETV cao nhờ các chỉ số này.
- Phù hợp với dữ liệu: Các feature đều có sẵn trong result.csv.

4. Cách các feature tương tác với mô hình

Mô hình hồi quy tuyến tính giả định ETV là tổ hợp tuyến tính:

$$ETV = \beta_0 + \beta_1 \cdot \text{Key Passes} + \beta_2 \cdot \text{PrgP} + \beta_3 \cdot \text{SCA} + \dots + \beta_{13} \cdot \text{Dis} + \beta_{14} \cdot \text{Mis}$$

Trong đó:

- $\beta_1, \beta_2, \beta_3, \beta_4, \beta_6, \beta_7, \beta_8, \beta_{10}, \beta_{11}$ (cho Key Passes, PrgP, SCA, xAG, PPA, Pass_into_1/3, Ast, Carries 1_3, ProDist) thường là số dương và lớn, do được nhân trọng số 2.0.
- $\beta_5, \beta_9, \beta_{12}$ (cho Cmp%, Touches, Rec) là số dương, nhưng nhỏ hơn.
- β_{13}, β_{14} (cho Dis, Mis) thường là số âm.

Tiền xử lý:

- Biến đổi log và chuẩn hóa đảm bảo các feature có thang đo tương đương.
- Trọng số 2.0 cho các feature sáng tạo và tiến công nhấn mạnh vai trò cốt lõi.

5. Hạn chế và đề xuất cải thiện

Hạn chế:

- Key Passes, SCA, xAG, Ast: Không tính đến độ khó của tình huống.
- PrgP, PPA, Pass_into_1/3, Carries 1_3, ProDist: Phụ thuộc vào phong cách chơi của đội.
- Cmp%, Touches, Rec: Có thể bị nhiễu bởi vai trò tiền vệ.

- Dis, Mis: Không phân biệt mức độ nghiêm trọng của sai lầm.

Đề xuất cải thiện:

- Thêm xG từ cơ hội tạo ra để đánh giá chất lượng sáng tạo.
- Sử dụng vị trí chuyền/mang bóng để tăng độ chính xác.
- Bổ sung tỷ lệ chuyền bóng dưới áp lực.
- Dùng mô hình phi tuyến (như Random Forest) để nắm bắt tương tác phức tạp.

6. Kết luận

Các feature Key Passes, PrgP, SCA, xAG, Cmp%, PPA, Pass_into_1/3, Ast, Touches, Carries 1_3, ProDist, Rec, Dis, và Mis được chọn vì chúng bao quát vai trò sáng tạo, tiến công, kiểm soát bóng, và độ tin cậy kỹ thuật của tiền vệ. Chúng phản ánh đúng giá trị thị trường, được xử lý kỹ lưỡng để phù hợp với mô hình hồi quy tuyến tính, và cung cấp cơ sở mạnh mẽ để dự đoán ETV.