

Legenda:

Identificadores

Variavel Dependente/Resultado

Variaveis a serem mantidas

Variaveis a serem transformadas/adaptadas

Variaveis que talvez possam ser usadas

Variáveis a serem removidas

---

[1] "Equipe"

"ID" \* → usar apenas para conseguir as medidas agregadas das últimas n rodadas

"Data"

"Horário" \* talvez analisar se um time joga melhor durante um horario

"Rodada"

[6] "Dia"

"Local" → sempre olhar pela ótica do dono de casa

→ Métrica derivada = aproveitamento em casa (para o time da casa)

→ Aproveitamento como visitante (para o time visitante)

"Resultado" → Y; variável dependente

→ Somar os pontos e fazer o aproveitamento

"GP"

"GC"

→ métrica derivada: saldo de gols

→ manter as duas, pois elas indicam capacidade ofensiva e defensiva, remover apenas em caso de multicolinearidade

[11] "Oponente"

"Posse" \* analisar a métrica e seu significado

"Publico" \* talvez, mas improvável; possível de ser removida

"Capitao"

"Formacao" \* muitas; avaliar se faz diferenca

[16] "Arbitro"

"TotalChutes"

"ChutesAGol" \* remover devido a multicolinearidade

"ChGol%"

"G/Ch"

[21] "G/ChGol"

"PenFeitos"

"PenBatidos" \* remover devido a multicolinearidade

➔ Métrica derivada: %Penaltis = penfeitos/penbatidos

"CaGC"

"Defesas" \* remover devido a multicolinearidade

[26] "%Defesas"

"CleanSheet"

➔ Métrica derivada: % de clean sheets; qtd de clean sheets

"PenAdv"

"GPenAdv"

"DefPenAdv"

[31] "ErroPenAdv"

➔ métrica derivada: %PenAdversarios = gpdenadv/PenAdv;

%DefPen = defPenadv/PenAdv

"CrtsA" \*testar em outro modelo talvez com as variáveis originais ao invés do score

"CrTV"

"2CrtsA"

➔ Métrica derivada: score de faltas = (crtsA\*valor + crtsV\*valor)

➔ Atribuir valores encontrando a razão; media de cartões por jogo

"FaltasCometidas" \* talvez usar as originais

[36] "FaltasSofridas"

➔ saldo de faltas= cometidas - sofridas

"Impedimentos"

"Cruzamentos"

"Cortes"

"RoubadasDeBola"

[41] "GolsContra"

➔ métrica derivada: %GolsContra = golscontra/gc

[1] "Equipe"

"Rodada"

"Local" → sempre olhar pela ótica do dono de casa

→ Métrica derivada = aproveitamento em casa (para o time da casa)

→ Aproveitamento como visitante (para o time visitante)

"Resultado" → Y; variável dependente

"GP"

"GC"

→ métrica derivada: saldo de gols

→ manter as duas, pois elas indicam capacidade ofensiva e defensiva, remover apenas em caso de multicolinearidade

[11] "Oponente"

"Posse" \* analisar a métrica e seu significado

"TotalChutes"

"ChGol%"

"G/Ch"

[21] "G/ChGol"

"PenFeitos"

→ Métrica derivada: %Penaltis = penfeitos/penbatidos

"CaGC"

[26] "%Defesas"

→ Métrica derivada: % de clean sheets; qtd de clean sheets

"GPenAdv"

→ métrica derivada: %PenAdversarios = gpenadv/PenAdv;

%DefPen = defPenadv/PenAdv

→ Métrica derivada: score de faltas = (crtsA\*valor + crtsV\*valor)

→ saldo de faltas= cometidas - sofridas

"Impedimentos"

"Cruzamentos"

"Cortes"

"RoubadasDeBola"

→ métrica derivada: %GolsContra = golscontra/gc

Possíveis modelos:

- Dados do time com o do adversário (feito)
- Substrair dados pareados (ex: sgHome – sgAway, etc.)
- Resumir e criar índices que agregam diversas variáveis: off, deff e outros, além de métricas auxiliares

*Índices:*

→ Ofensivo:

Em relação ao rótulo:

- Vitória, empate e derrota
- Vitória ou não vitória

---

Inserir dados em produção: groupby

Dados para treinar modelo: média móvel (última rodada, 3 rodadas, 5 rodadas, 10 rodadas, campeonato inteiro)

Obter dados para estimar modelo:

Home -

Away –

Join das duas – e depois fazer as coisas descritas acima

> #métricas auxiliares: formacao

> #métricas agregadas: posse-avg, totalchutes - avg, chgol% - avg, g/ch - avg, \*\*%gols de pen - sum, cagc - avg, %defesas - avg, cleansheets - sum, %defpen - avg, saldfaltas -avg, stats -avg, \*\*% gols contra, pts - sum, saldogols - sum, cardscore – avg

\*\* apresentam problemas quando se usa a média mas por enquanto é o jeito mais fácil de se obter essas métricas de maneira temporal/roll