

در مجموعه دادگان، 240 بازی مختلف وجود دارد.

برای تعیین شوت زن های برتر، ویژگی playType بسیار مهم است. چون سوباسا در تیم خود نیاز به پناستی زن های خوب، شوت زن های خوب در جریان بازی و ... دارد. اما وزن اهمیت انواع شوت با هم یکی نیست! چون در خیلی بازی ها ممکن است اصلا پناستی رخ ندهد یا کم رخ دهد اما تعداد بالایی شوت در جریان بازی همیشه زده خواهد شد. برای تعیین اهمیت حالت های مختلف playType، میانگین وقوع هر حالت را در همه match ها محاسبه میکنم. این میانگین ها در شکل 1 نشان داده شده اند.

(وزن هر حالت به صورت میانگین آن حالت تقسیم بر مجموع میانگین ها محاسبه میشود)

با توجه به اینکه مدل آموزش دیده در سوال "احتمال گل" برآوردی از میانگین عملکرد همه بازیکن ها است، مطابق جدول 1، پاداش برای عملکرد بازیکنان در هر سطر از داده آموزش در نظر میگیرم. در ابتدا پاداش همه بازیکن ها 0 در نظر گرفته میشود. پاداشی که در جدول نوشته شده در وزنی که برای هر حالت playType به دست آوردم ضرب میشود.

در نهایت، مجموع پاداش همه بازیکن ها مطابق جدول 1، سطر های داده آموزش و وزن های حالات مختلف شوت محاسبه خواهد شد و تقسیم بر تعداد بازی هایی میشود که بازیکن در آن ها شرکت کرده است. 5 بازیکن اول بعد از مرتب سازی نزولی این اعداد، معرفی میشوند.

(دلیل اینکه مجموع پاداش را تقسیم بر تعداد شوت نکردم این است که تعداد شوت بیشتر هم، خود میتواند یک شاخص برای تعیین شوت زن خوبی بودن باشد. اما تعداد بازی در داده ها ممکن است به نفع تعدادی از بازیکن ها bias باشد)

توضیح جدول: همانطور که گفته شد، مدل سوال احتمال گل برآورد عملکرد همه بازیکنان است، پس اگر در موقعیتی معمولاً بازیکنان گل میزنند و واقعا هم بازیکنی گل زده، پاداش +1 میگیرد. اما اگر در موقعیتی معمولاً بازیکنان گل نمیزنند، با این حال بازیکنی در آن موقعیت گل زده است، پاداش +2 میگیرد.

(برای منفی ها هم به همین صورت)

برای classify کردن خروجی های مدل هم، $\text{threshold} = 0.5$ در نظر گرفته میشود.

خروجی 5 playerId بازیکن شوت زن برتر به ترتیب:

p_18

p_44

p_2

p_115

p_66

در شکل 2 تعداد وقوع مغایرت نتیجه مدل با نتیجه واقعی (0) و تطابق نتیجه مدل با نتیجه واقعی به نمایش در آمده است.

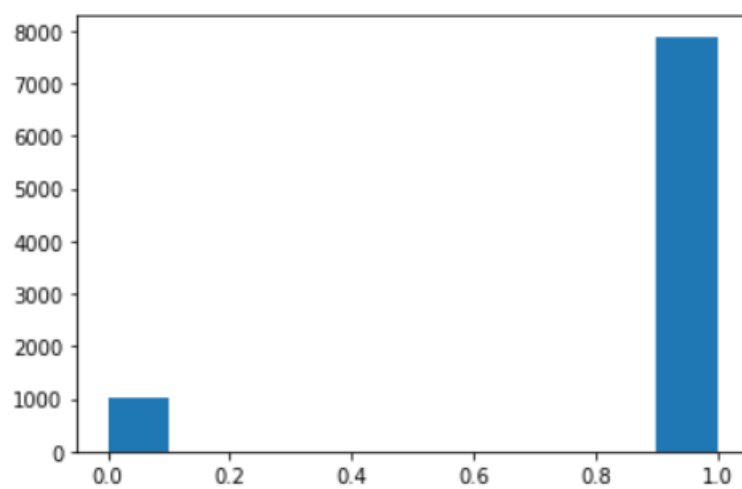
یعنی تعداد پاداش ها با ضرایب +2 و تعداد پاداش ها با ضرایب -1.

	جریان بازی	پنالتی	ضربه آزاد مستقیم	مستقیم از کرنر
count	240.000000	240.000000	240.000000	240.000000
mean	34.908333	0.525000	1.737500	0.016667

شکل 1

پاداش	خروجی واقعی	خروجی مدل
1+	گل شده	گل میشود
2-	گل نشده	گل میشود
2+	گل شده	گل نمیشود
1-	گل نشده	گل نمیشود

جدول 1



شکل 2