

Finał UCL 2023

Rafał Mielniczuk

12 czerwca 2023

Spis treści

1	Problem badawczy	3
2	Dane	3
2.1	Pozyskane dane	3
2.1.1	Zakres	3
2.1.2	Szczegóły	3
2.1.3	Sposób	5
2.2	Odrzucone dane	5
2.2.1	Zakres	5
2.2.2	Szczegóły	5
3	Analiza danych	5
3.1	Rozgrywki ligowe	5
3.1.1	AC Milan	6
3.1.2	Internazionale Mediolan	9
3.1.3	Real Madrid	12
3.1.4	Manchester City	15
3.2	Rozgrywki UCL	18
3.2.1	Real Madrid	19
3.2.2	Manchester City	22
4	Modelowanie danych	25
4.1	Problem	25
4.1.1	Objaśnienie	25
4.2	Real Madrid - Manchester City	26

4.3	AC Milan - Internazionale Mediolan	28
4.4	Finał	29
5	Wnioski	30
5.1	Podsumowanie	30
5.2	Prawdziwy wynik fianłu	30
5.3	Komentarze i ulepszenia	30

1 Problem badawczy

W piłce nożnej wszystko jest możliwe. Nawet słabsza drużyna zawsze ma realne szanse na wygraną meczu, chociaż jego wynik sugerowałby zupełnie co innego. Zgodnie z tą myślą finał UEFA Champions League (UCL) może przynieść zaskakujące rezultaty. Co więcej, wyniki półfinałów mogłyby być już dla niektórych zaskoczeniem.

Przedstawia to oczywisty problem: czy można na podstawie dostępnych danych z obecnych sezonów ligowych i zeszłych rozgrywek UCL wyłonić zasłużonych finalistów oraz zwycięzcę tegorocznego UCL. Ze względu na rozważanie finału problem jest ograniczony tylko do tegorocznych półfinalistów - AC Milan, Internazionale Mediolan, Real Madrid, Manchester City.

2 Dane

2.1 Pozyskane dane

2.1.1 Zakres

Najważniejszym czynnikiem pozwalającym na ocenę szans danej drużyny jest jej obecna forma, dlatego należy uwzględnić w pozyskanych danych obecny sezon. Ponadto drużyny nie zawsze przywiązują taką samą wagę do rozgrywek domowych, co do rozgrywek europejskich. Z tego powodu należy również zawrzeć w danych informacje o poprzednich rozgrywkach UCL. Ze względu na dynamikę piłki nożnej i częste rotacje w składach drużyn tylko dane z ostatnich 5 lat zostaną uwzględnione. Ponadto będą to dane tylko rozgrywek 1/8 bądź wyżej, ponieważ odpadnięcie w grupie dla drużyny na tym poziomie jest nieudanym sezonem.

2.1.2 Szczegóły

Do zakresu analizowanych danych należą statystyki strzałów oraz piłkarzy drużyn w meczu. Przykładowo:

	player	shirtnumber	nationality	position	age	minutes	goals	assists	pens_made	pens_att	...	sca	gca	passes_completed	passes	passes_pct
0	Ante Rebic	12	CRO	FW	28	57	0	0	0	0	...	4	0	5	11	45.5
1	Olivier Giroud	9	FRA	FW	35	33	0	0	0	0	...	0	0	3	4	75.0
2	Rafaël Leão	17	POR	LW	23	65	0	0	0	0	...	4	0	12	17	70.6
3	Divock Origi	27	BEL	LW	27	25	0	0	0	0	...	1	0	4	4	100.0
4	Junior Messias	30	BRA	RW	31	65	0	0	0	0	...	0	0	18	22	81.8
5	Alexis Saelemaekers	56	BEL	RW	23	25	0	1	0	0	...	2	1	8	14	57.1
6	Brahim Díaz	10	ESP	AM	23	57	0	0	0	0	...	2	0	16	19	84.2
7	Charles De Ketelaere	90	BEL	AM	21	33	0	0	0	0	...	1	0	6	9	66.7
8	Sandro Tonali	8	ITA	DM	22	90	0	0	0	0	...	3	0	46	58	79.2
9	Ismaël Bennacer	4	ALG	DM	24	90	1	0	0	0	...	11	1	77	90	85.6
10	Theo Hernández	19	FRA	LB	24	90	0	0	0	0	...	3	0	55	66	83.2
11	Fikayo Tomori	23	ENG	CB	24	90	0	0	0	0	...	1	0	72	76	94.7
12	Pierre Kalulu	20	FRA	CB	22	90	0	0	0	0	...	0	0	56	58	96.6
13	Davide Calabria	2	ITA	RB	25	83	0	0	0	0	...	3	0	62	75	82.7
14	Alessandro Florenzi	25	ITA	RB	31	7	0	0	0	0	...	0	0	6	9	66.7
15	Mike Maignan	16	FRA	GK	27	90	0	0	0	0	...	0	0	35	38	92.1
16 rows × 17 columns																

16 rows • 31 columns

minute	player	team	xg_shot	psxg_shot	outcome	distance	body_part	notes	sca_1_player	sca_1_type	sca_2_player	sca_2_type	
0	1	Rafael Leão	Milan	0.09	NaN	Off Target	14	Right Foot	NaN	Brahim Diaz	Pass (Live)	Ante Rebić	Pass (Live)
1	3	Rafael Leão	Milan	0.05	0.12	Saved	16	Right Foot	NaN	Ante Rebić	Pass (Live)	Rafael Leão	Pass (Live)
2	19	Sandro Tonali	Milan	0.05	NaN	Off Target	26	Right Foot	NaN	Ismaïl Bennacer	Pass (Live)	Davide Calabria	Pass (Live)
3	22	Theo Hernández	Milan	0.02	NaN	Blocked	31	Left Foot	NaN	Ismaïl Bennacer	Pass (Dead)	Ante Rebić	Fouled
4	23	Junior Messias	Milan	0.12	NaN	Off Target	10	Left Foot	Volley	Theo Hernández	Pass (Live)	Ismaïl Bennacer	Pass (Live)
5	30	Theo Hernández	Milan	0.03	NaN	Off Target	29	Left Foot	NaN	Ismaïl Bennacer	Pass (Live)	Davide Calabria	Pass (Live)
6	32	Pierre Kalulu	Milan	0.12	NaN	Off Target	9	Head	NaN	Ismaïl Bennacer	Pass (Dead)	Rafael Leão	Pass (Live)
7	34	Rafael Leão	Milan	0.02	0.21	Saved	32	Right Foot	NaN	Brahim Diaz	Pass (Live)	Sandro Tonali	Pass (Live)
8	43	Theo Hernández	Milan	0.04	NaN	Blocked	27	Left Foot	NaN	Ismaïl Bennacer	Pass (Live)	Ante Rebić	Pass (Live)
9	45	Davide Calabria	Milan	0.02	NaN	Off Target	30	Right Foot	Volley	Ismaïl Bennacer	Pass (Live)	Theo Hernández	Shot
10	50	Pierre Kalulu	Milan	0.36	NaN	Off Target	7	Head	NaN	Ismaïl Bennacer	Pass (Dead)	Rafael Leão	Fouled
11	53	Rafael Leão	Milan	0.04	NaN	Off Target	29	Right Foot	NaN	Ismaïl Bennacer	Pass (Live)	NaN	NaN
12	60	Sandro Tonali	Milan	0.05	0.10	Saved	14	Right Foot	NaN	Charles De Ketelaere	Pass (Live)	Theo Hernández	Pass (Live)
13	63	Olivier Giroud	Milan	0.02	0.11	Saved	12	Head	NaN	Ismaïl Bennacer	Pass (Live)	Rafael Leão	Pass (Live)
14	65	Junior Messias	Milan	0.11	NaN	Off Target	10	Head	NaN	Sandro Tonali	Pass (Live)	Davide Calabria	Pass (Live)
15	68	Ismaïl Bennacer	Milan	0.03	0.15	Saved	29	Left Foot	NaN	Alexis Saelemaekers	Pass (Live)	Divock Origi	Pass (Live)
16	68	Ismaïl Bennacer	Milan	0.09	0.53	Goal	13	Left Foot	NaN	Alexis Saelemaekers	Pass (Live)	Ismaïl Bennacer	Shot
17	86	Theo Hernández	Milan	0.02	0.07	Saved	27	Left Foot	Deflected	Sandro Tonali	Pass (Live)	Fikayo Tomori	Pass (Live)

W danych piłkarzy można wyróżnić kolejno: imię i nazwisko, numer koszulki, pozycja, wiek, rozegrane minuty, gole, asysty, rzuty karne wykonane, próby rzutów karnych, ilość strzałów, ilość strzałów na bramkę, żółte kartki, czerwone kartki, kontakty, przejęcia, przechwycenia, bloki, oczekiwane gole, oczekiwane gole bez rzutów karnych, oczekiwane gole po asyście, liczba akcji kreujących strzał, liczba akcji kreujących gol, udane podania, podania, skuteczność podań, progresywne podania, kontrola piłki, progresywna kontrola piłki, dryblingi, udane dryblingi.

W danych strzałów można wyróżnić kolejno: minuta, imię i nazwisko, klub, oczekiwane gole, oczekiwane gole uwzględniając pozycję bramkarza, rezultat, dystans do bramki, część ciała, adnotacje, gracze i wydarzenia kreujące akcję.

2.1.3 Sposób

Dane zostały pozyskane poprzez web scraping[1].

2.2 Odrzucone dane

2.2.1 Zakres

Analizując występy drużyn należy skupiać się również na poszczególnych jednostkach, które się na nią składają. Nie zawsze jednak można porównać dwóch zawodników. Niektórzy z nich pojawiają się na murawie tylko w celach strategicznych, aby skrócić rozgrywany czas trwania meczu. Wchodzą w ostatnich minutach meczu i ich występ statystycznie wypada bardzo słabo na tle innych zawodników. Ponadto bramkarze mają specyficzną rolę, która jest ciężka do porównania do innych zawodników w polu.

Z wyżej wymienionych powodów statystyki zawodników będących na murawie poniżej 10 minut i bramkarzy nie będą brane pod uwagę.

2.2.2 Szczegóły

Ponieważ szczegółowość dostarczonych danych jest zbyt duża na potrzeby rozważanego raportu, dane zostaną ograniczone.

Dane graczy kolejno: imię i nazwisko, pozycja, rozegrane minuty, gole, asysty, oczekiwane gole, oczekiwane gole po asyście, liczba akcji kreujących strzał, liczba akcji kreujących gol, udane podania, podania, skuteczność podań.

Dane strzałów kolejno: minuta, imię i nazwisko, klub, oczekiwane gole, rezultat.

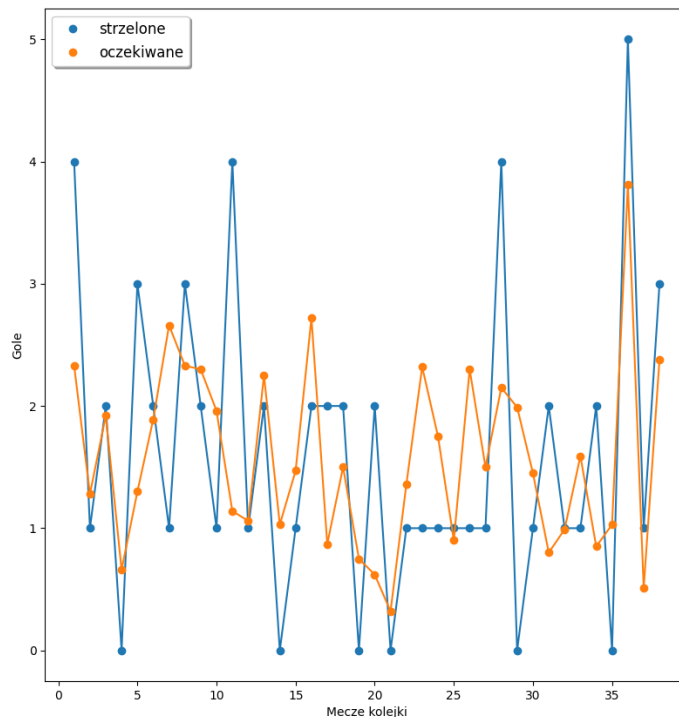
3 Analiza danych

3.1 Rozgrywki ligowe

Każda z analizowanych drużyn rozegrała w tym sezonie 38 spotkań w rozgrywkach domowych. Do analizy zostaną poddane najważniejsze statystyki

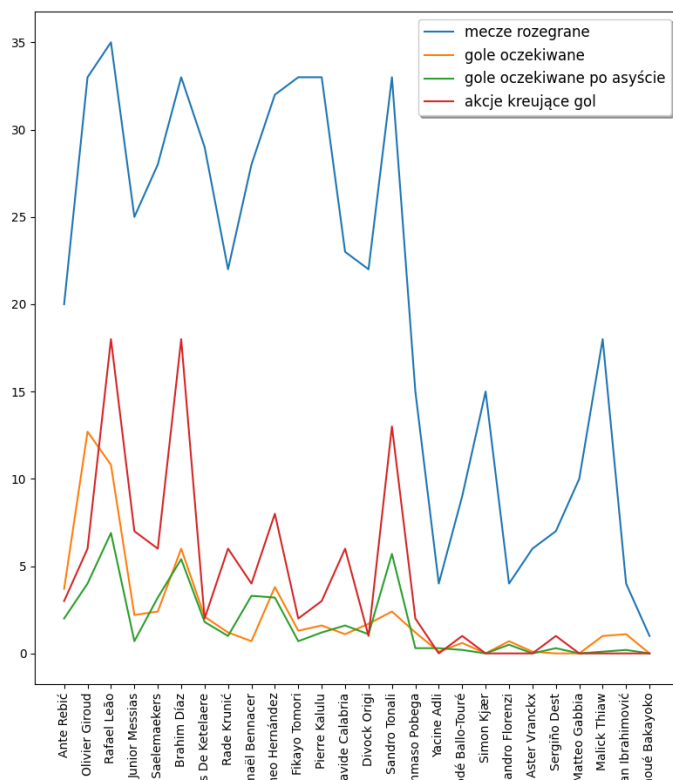
pozwalające zwyciężać, czyli okazje strzeleckie, szanse na zdobywanie bramek oraz dokładność podań drużyny. Zwycięstwa nie mają aż tak dużego znaczenia, ponieważ piłka nożna jest nieprzewidywalna. Natomiast stałe dobre rezultaty pozwalają zobaczyć, czy drużyna rzeczywiście reprezentowała wysoką formę. Szczególną uwagę można zwrócić na okres fazy pucharowej ligi mistrzów zaczynający się mniej więcej w trakcie kolejki 22.

3.1.1 AC Milan



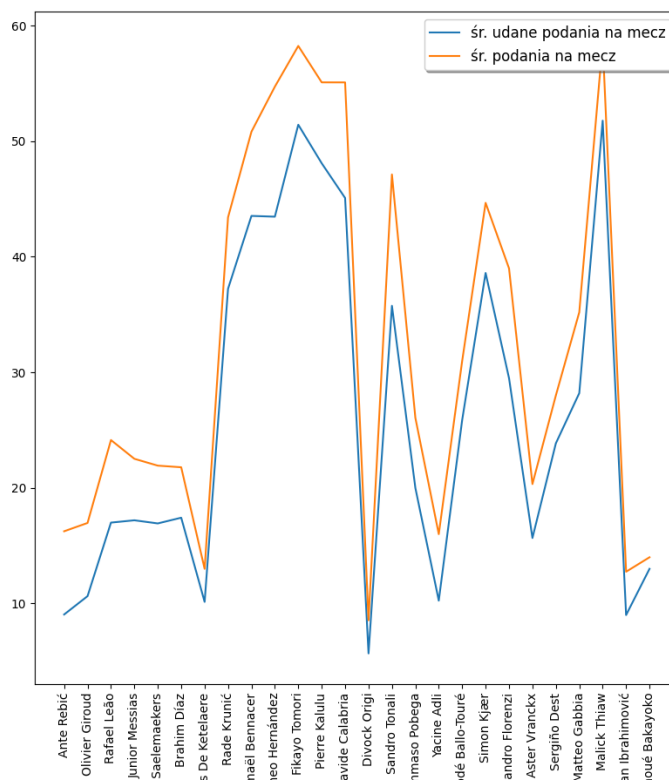
Rysunek 1: Gole AC Milanu

W przypadku goli można zauważyć stałą tendencję do 1 lub 2 bramek na mecz. Można uznać AC Milan pod tym względem za dość stabilną drużynę.



Rysunek 2: Gole graczy AC Milanu

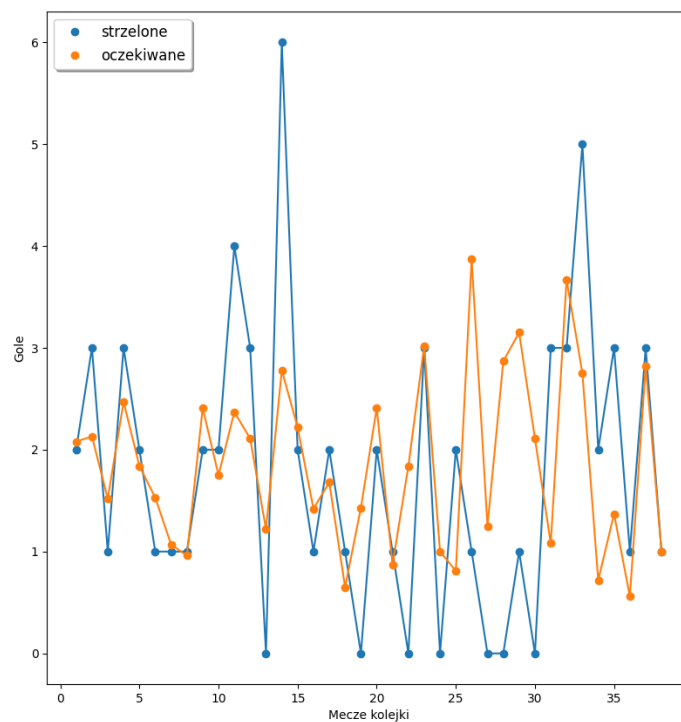
Można zauważyć, że analizowany zespół opiera swoją grę głównie na kilku zawodnikach, gdzie przeważnie Olivier Giroud i Rafael Leao mają okazje strzeleckie, natomiast akcje wyprowadzane są głównie przez S.Tonalego, B.Diaza oraz ponownie R.Leao.



Rysunek 3: Podania graczy AC Milanu

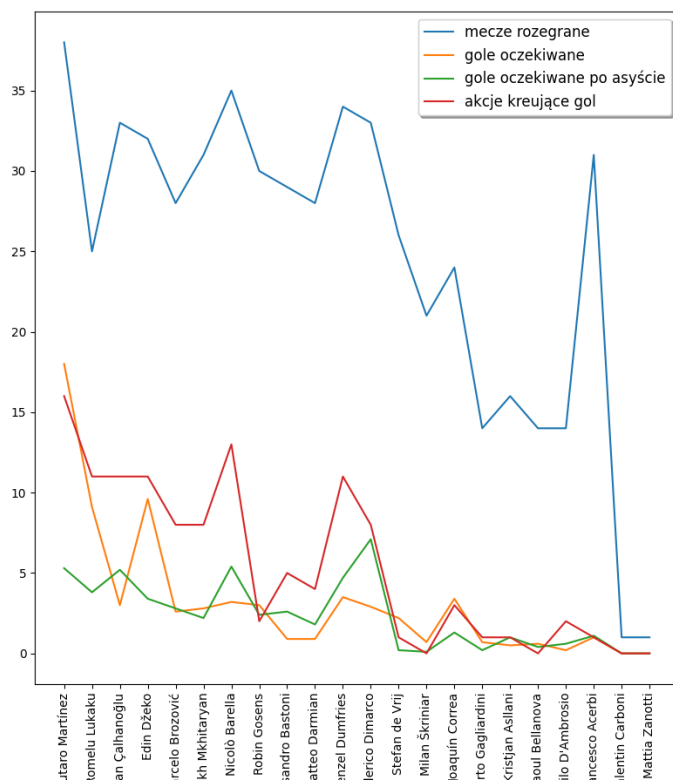
Drużyna bardzo słabo wypada na europejskim poziomie w kwestii podań i ich dokładności. U wielu zawodników można zauważyć dużą, jak na reprezentowany przez nich poziom, różnicę między udanymi, a wykonanymi podaniami. Ponadto liczba podań jest przeważnie mała. Sugeruje to długie piłki i grę z kontry.

3.1.2 Internazionale Mediolan



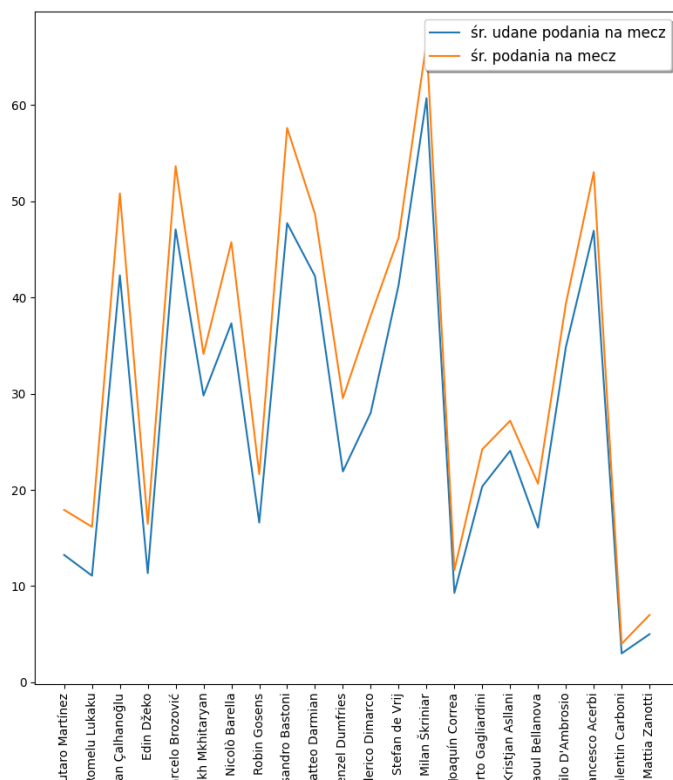
Rysunek 4: Gole Internazionale Mediolan

Gole na mecz Internazionale Mediolan, zwanego również potocznie Nerazzurri bądź Inter, również oscylują w przedziale od 1 do 2. Nawet można zauważyć wzrost goli oczekiwanych pod koniec sezonu, to jest w okresie trwających rozgrywek UCL.



Rysunek 5: Gole graczy Internazionale Medional

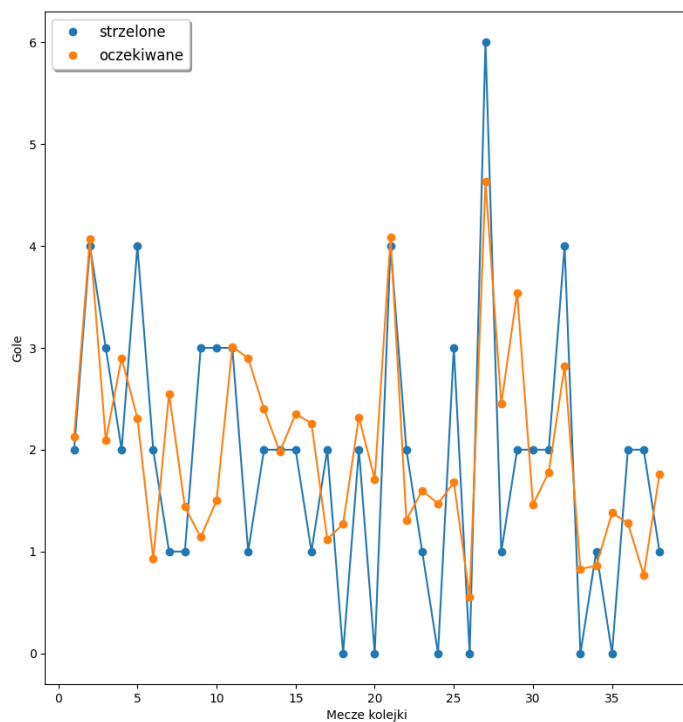
Nerazzurri ma dużo solidnych graczy wyprowadzających akcje. Ich gra może się wydawać bardziej rozłożona i niezależna od kilku jednostek. Natomiast gole są strzelane przeważnie przez ich napastników - R.Lukaku, L.Martineza i E.Dzeko. Godny wyróżnienia jest również F.Dimarco, którego podania stwarzają duże zagrożenie bramkowe.



Rysunek 6: Podania graczy Internazionale Mediolan

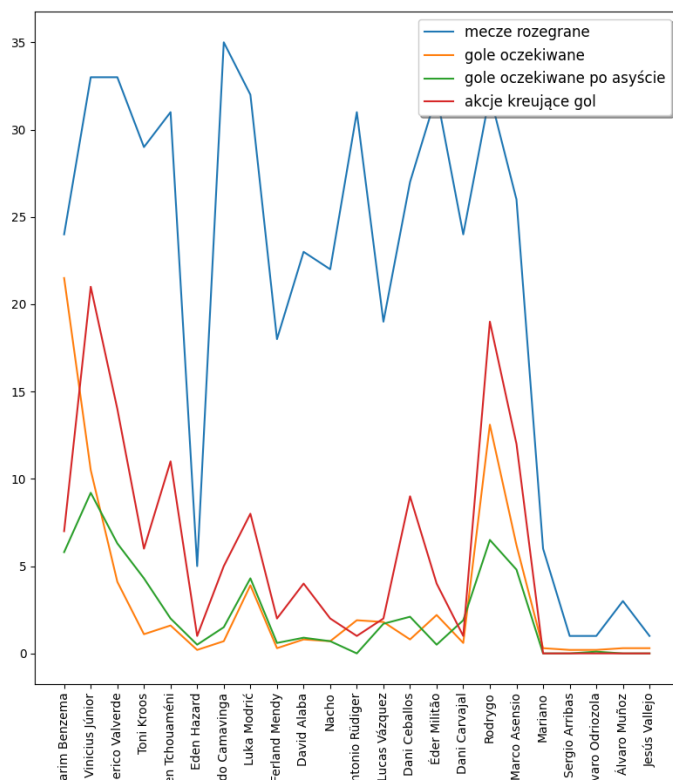
Podania zawodników Interu cechują się przeważnie w dużą dokładnością. Ponadto ich średnia jest stosunkowo duża, co może świadczyć o defensywnej grze lub wolno budowanych akcjach.

3.1.3 Real Madrid



Rysunek 7: Gole Realu Madrid

Real ma dość wysoką średnią goli w okolicach 2. Ponadto ich oczekiwane gole często sięgają aż 3.



Rysunek 8: Gole graczy Realu Madrid

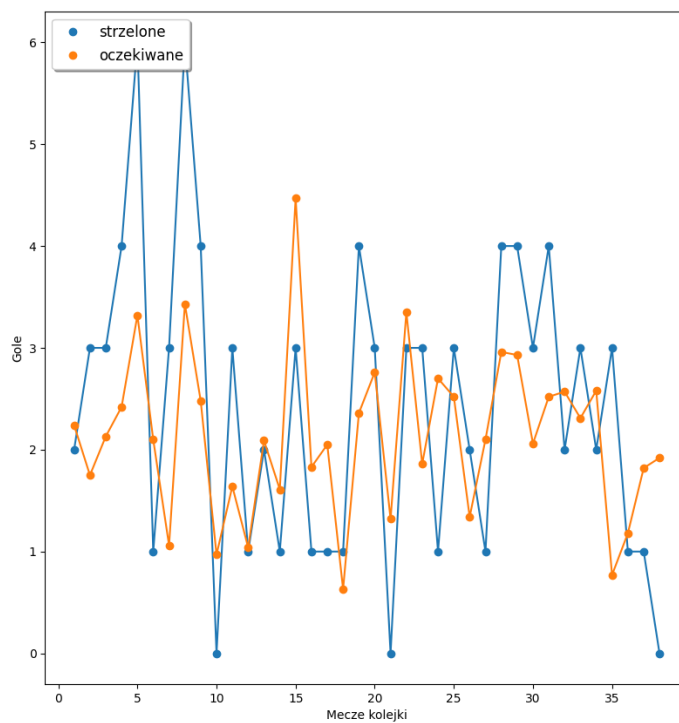
Z wykresu można zauważyć, że drużyna ma nadzwyczaj uniwersalną drużynę, gdzie dużo osób jest w stanie kreować akcje bramkowe i wykonywać podania z dużą szansą na bramkę. Jednak da się zauważyć trzech głównych strzelców, czyli K.Benzemę, Viniciusa oraz Rodrygo.



Rysunek 9: Podania graczy Realu

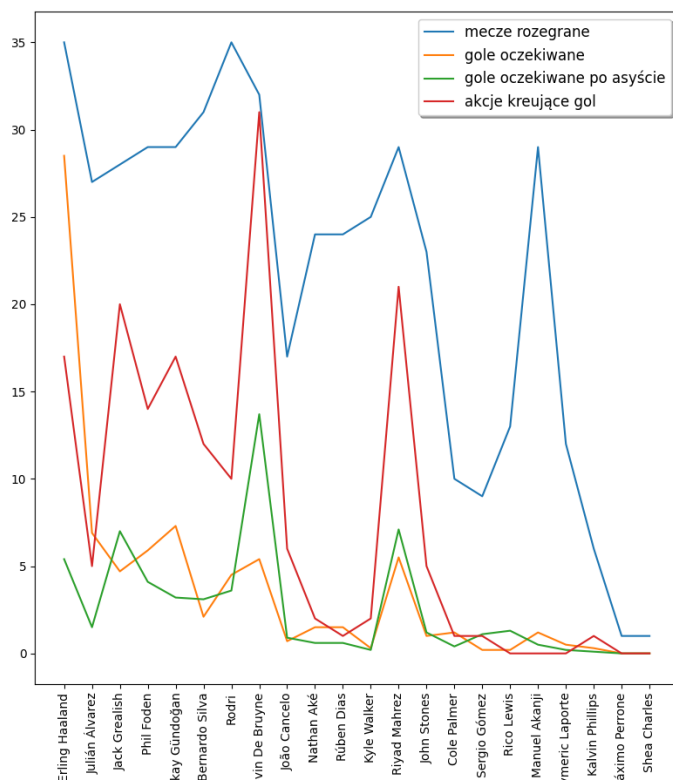
Podania graczy analizowanego zespołu są równomiernie rozłożone, co sugeruje, że grają równo na szerokości całego boiska. Zwłaszcza duża ilość podań można zauważyć u T.Krossa, F.Valverde i A.Tchouameni, czyli dużo akcji przechodzi przez środek pola.

3.1.4 Manchester City



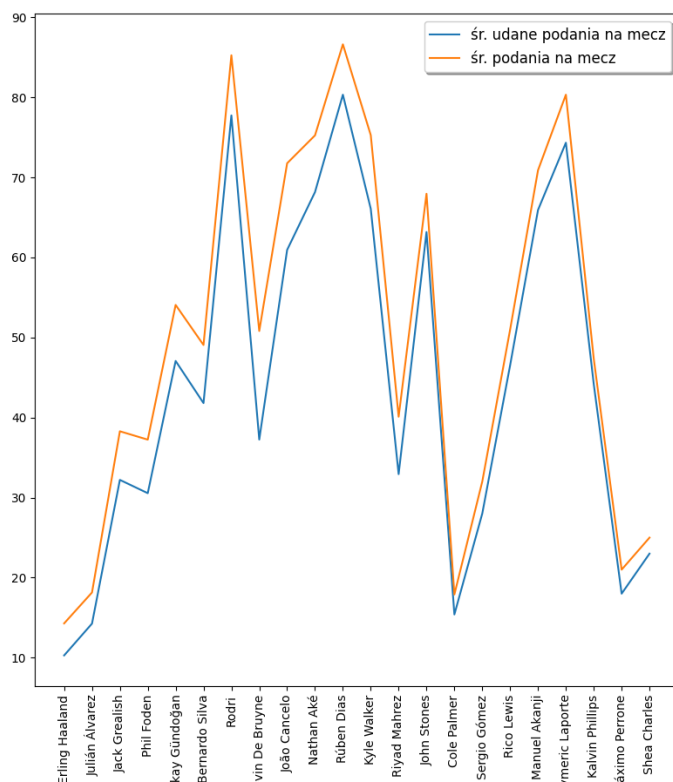
Rysunek 10: Gole Manchesteru City

Manchester ma ponadprzeciętną średnią goli, a ponadto nie zalicza spadków formy, jak można to było zauważyć u innych drużyn. Ich średnia oscyluje w okolicach 2 i powyżej. Szczególnie wysoką formę można zauważyć w okolicach 30. kolejki.



Rysunek 11: Gole graczy Manchesteru City

Wykres przedstawia zaskakująco dobre statystyki. Można tutaj również wyróżnić kilku graczy, ale z pewnością cała drużyna prezentuje się nadzwyczaj solidnie. Z pewnością dwóch najbardziej odznaczających się graczy to Kevin de Bruyne, kreujący najwięcej akcji, i Erling Haaland, strzelający najwięcej goli.



Rysunek 12: Podania graczy Manchesteru City

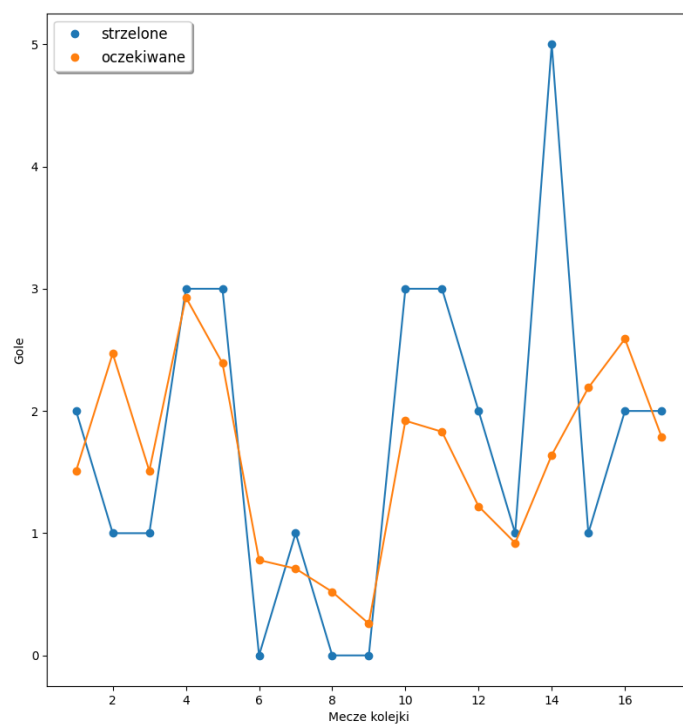
Podania zawodników Manchesteru City również prezentują się nadzwyczaj dobrze. Ich ilość jest nieporównywalnie większa od innych analizowanych klubów oraz dokładność jest również na wysokim poziomie.

3.2 Rozgrywki UCL

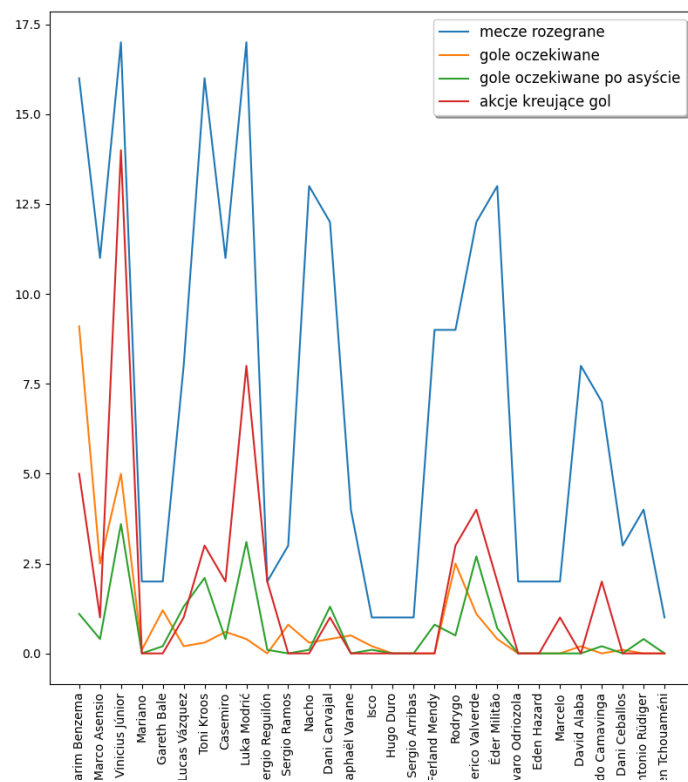
Niestety Milan i Inter w ostatnich latach bardzo słabo wypadały w lidze mistrzów. W ciągu ostatnich 5 lat tylko Interowi raz udało się wejść do fazy pucharowej, odpadając w 1/8 finału. Tym samym oba kluby dostarczają za mało informacji do analizy. Tylko Manchester City i Real Madrid prezentowały wyniki, które można przeanalizować.

Dla obywateli drużyn można zauważyć trend. Naturalnie zwiększa się liczba podań, spada liczba goli oraz akcji kreujących gole ze względu na trudność rozgrywek. Pomimo tego dalej są gracze wybitni na tle reszty zespołu, przykładowo Vinicius i Modric z Realu oraz Kevin de Bruyne i Phil Foden z Manchesteru. Niemniej jednak dalej można zauważyć o wiele lepsze statystyki Manchesteru jako całego zespołu, czy pojedynczych jednostek tworzących zespół.

3.2.1 Real Madrid



Rysunek 13: Gole Realu Madrid

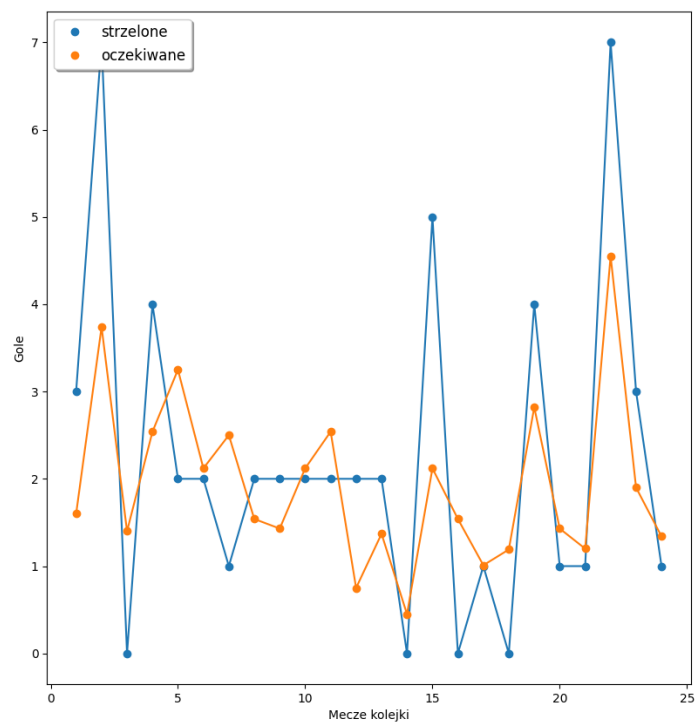


Rysunek 14: Gole graczy Realu Madrid

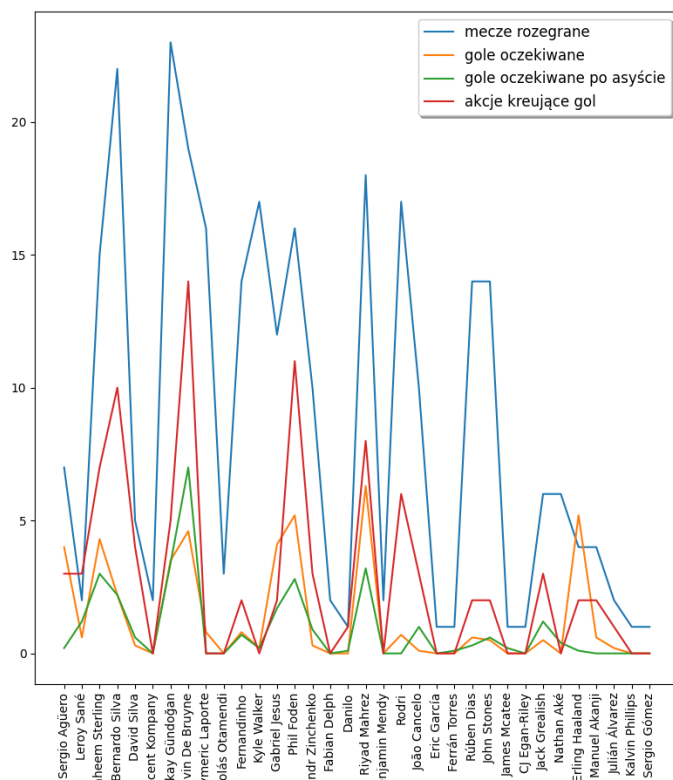


Rysunek 15: Podania graczy Realu Madrid

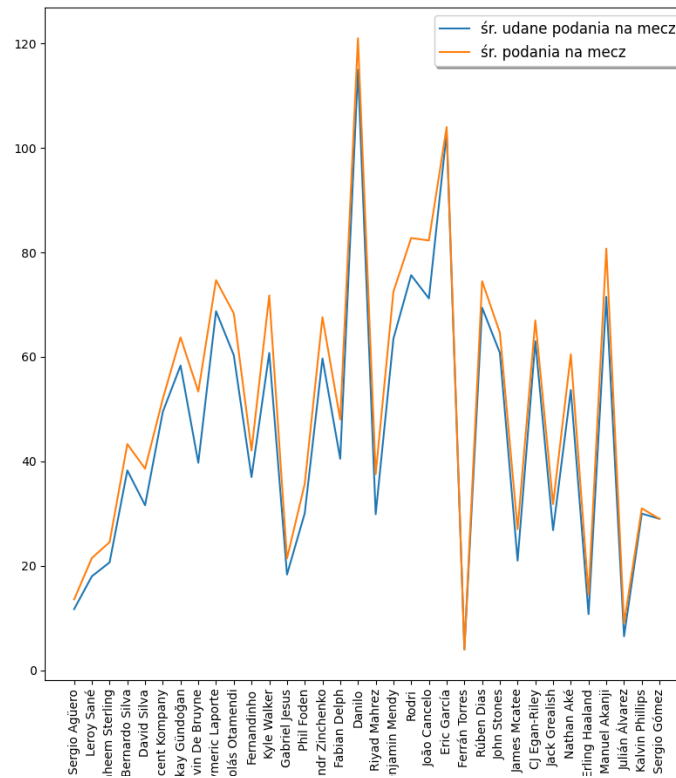
3.2.2 Manchester City



Rysunek 16: Gole Manchesteru City



Rysunek 17: Gole graczy Manchesteru City



4 Modelowanie danych

4.1 Problem

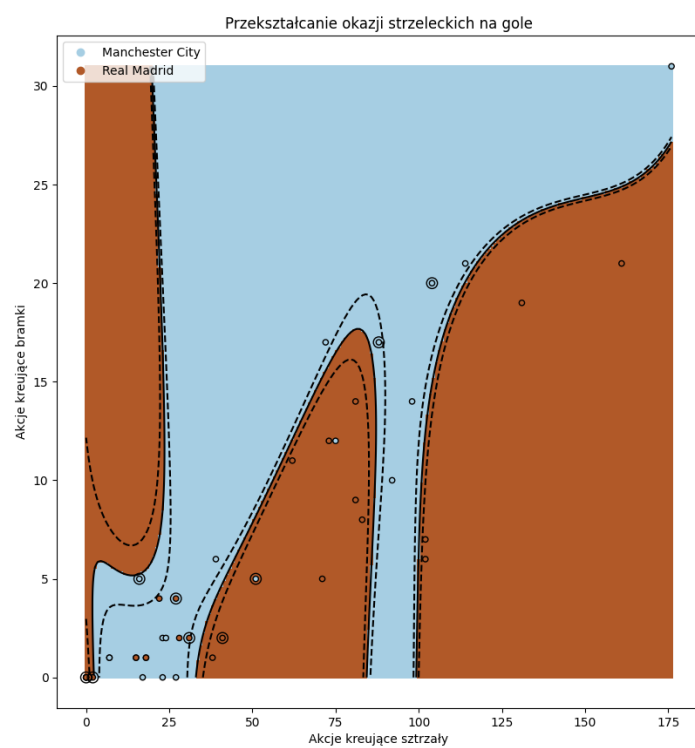
Podane powyżej dane pozwalają zobrazować formę danych drużyn. Jednak najistotniejsze według wielu analityków jest przekształcanie okazji na bramki. Tym samym dla rozwiązywanego problemu kluczowa jest klasyfikacja zawodników obu drużyn w stosunku akcji kreujących bramki do akcji kreujących strzały.

4.1.1 Objaśnienie

W celu zobrazowania przewagi drużyny posłużę się SVM (Support Vector Machines) służącymi między innymi do klasyfikacji. Metoda pozwala w przystępny sposób na odczytanie z wykresu żądanych informacji. W tym przypadku ma pokazywać, że klaster danej drużyny znajduje się w większości w strefie większej skuteczności względem drużyny przeciwnej. Tym samym dowodząc, że ten zespół ma większe szanse na wygraną.

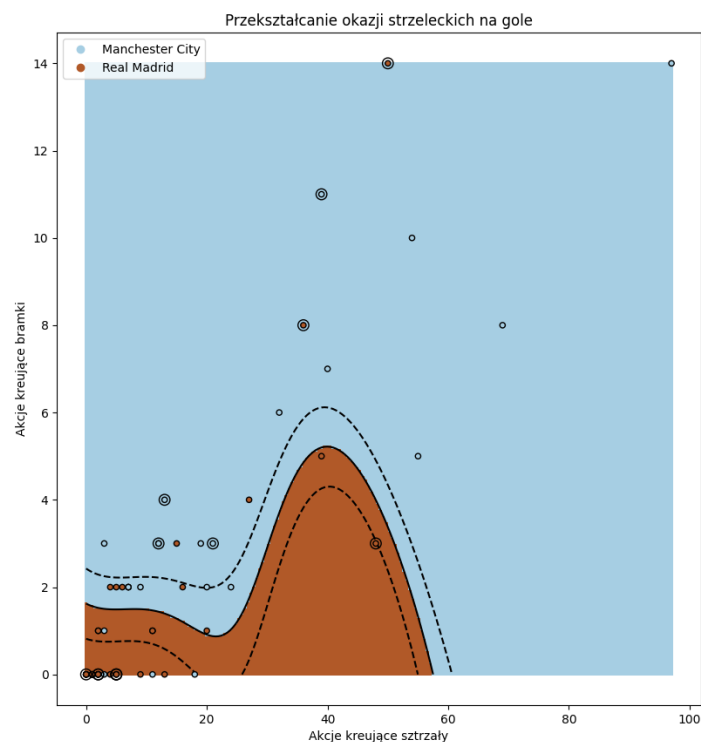
SVM polega na wyznaczeniu płaszczyzny znajdującej się w jak największej odległości od punktów treningowych jakiegokolwiek klasy. Odległość, zwaną marginesem, od płaszczyzny do punktów SVM maksymalizuje przy pewnych założeniach. Przykładowo parametr C umożliwia regulację pewnego dopuszczalnego dystansu, w którym się mogą znaleźć punkty treningowe. Z kolei γ określa jądro, które udostępnia funkcję mierzącą podobieństwo między danym punktem treningowym i innymi.

4.2 Real Madrid - Manchester City



Rysunek 19: Klasyfikacja - Real Madrid vs Manchester City

Dla powyższego wykresu: accuracy=67%, precision=75%, recall=60%.

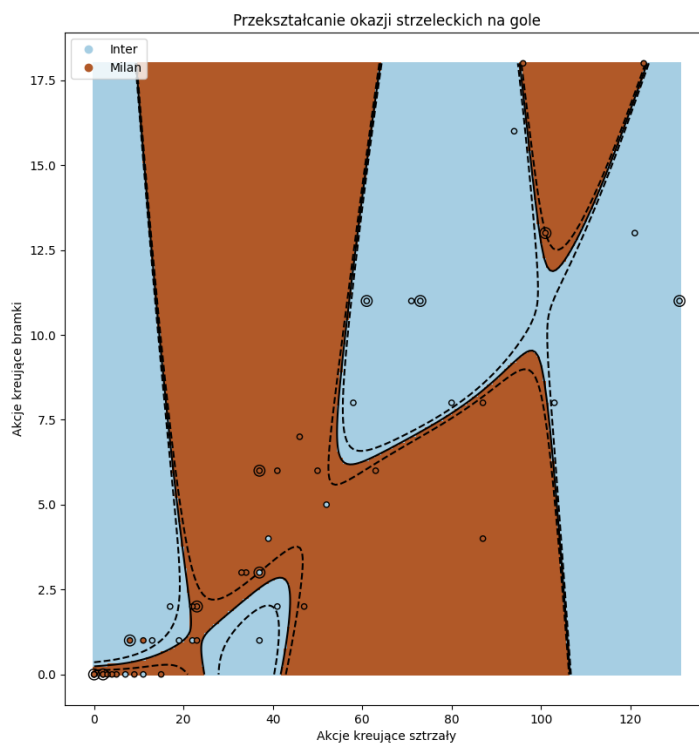


Rysunek 20: Klasyfikacja UCL - Real Madrid vs Manchester City

Dla powyższego wykresu: $\text{accuracy}=62\%$, $\text{precision}=57\%$, $\text{recall}=67\%$.

Jak można zauważyć dla dwóch powyższych wykresów Manchester zdecydowanie dominuje skutecznością nad Realem. Gracze Manchesteru zwłaszcza w rozgrywkach UCL mają więcej akcji kreujących strzały oraz większe ich przekształcanie na strzały kreujące bramki. Co jednak najważniejsze w obu przypadkach da się zauważyć gracza z niesamowitą formą (skrajny prawy górny róg). Jest to Kevin de Bruyne, który jest niewątpliwie ogromnym atutem Manchesteru.

4.3 AC Milan - Internazionale Mediolan

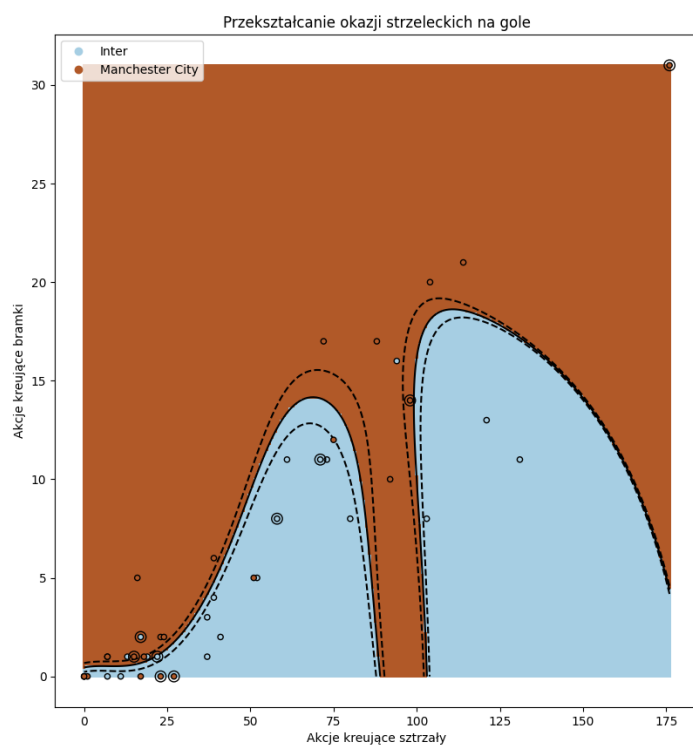


Rysunek 21: Klasyfikacja - AC Milan vs Internazionale Mediolan

Dla powyższego wykresu: $\text{accuracy}=70\%$, $\text{precision}=67\%$, $\text{recall}=80\%$.

Sytuacja odnośnie dwóch mediolańskich drużyn nie jest już taka oczywista. Zawodnicy Interu z całą pewnością prezentują bardziej solidną formę. Nie da jednak nie zauważyć, że Milan posiada 3 utalentowanych graczy (Diaz, Tonali, Leao), którzy mogliby odwrócić bieg spotkania.

4.4 Finał



Rysunek 22: Klasyfikacja - Manchester City vs Internazionale Mediolan

Dla powyższego wykresu: $\text{accuracy}=67\%$, $\text{precision}=75\%$, $\text{recall}=60\%$.

Wykres prezentuje niezmienną przewagę Manchesteru nad rywalami. Choć można zauważyć, że gracze Interu kreują dużo akcji strzeleckich, nie potrafią ich wykorzystać. Jakość gry świadczy zdecydowanie na korzyść Manchesteru City.

5 Wnioski

Na podstawie wstępnej analizy danych jak również wykonanego modelu obrazującego stosunek kreowania akcji bramkowych do akcji strzeleckich można z całą pewnością założyć, że zwycięzcą tegorocznej edycji UCL będzie Manchester City. Ponadto znani już finaliści, czyli Internazionale Mediolan i Manchester City, również zdają się być zasłużeni. Największe wątpliwości można wobec Interu, którego model niejednoznacznie pokazał silniejszą drużynę.

5.1 Podsumowanie

Rozwiązując problem wyłonienia najlepszej drużyny ciężko jest uzasadnić swoją decyzję, ponieważ na wygraną składa się bardzo wiele czynników. Niemniej w sporcie często tylko jedno ma znaczenie - wynik. Uznałem tym samym, że w piłce nożnej nic innego, jak tworzenie akcji strzeleckich oraz ich skuteczność może zobrazować potęgę drużyny.

Ponadto rzeczywisty wynik meczu może nie odwzorowywać statystyk i prognoz. Nie da się jednoznacznie go stwierdzić. Można udostępnić najwyżej wykresy i zestawienia do własnej analizy odbiorcy dodając również coś od siebie.

5.2 Prawdziwy wynik finału

Na chwilę pisania tego raportu wynik finału UCL 2023 jest już znany. Wygrał Manchester City, co odpowiada przewidywaniu na podstawie wykorzystanego modelu.

5.3 Komentarze i ulepszenia

Obrany model SVM nie cechował się zbyt dużą dokładnością nawet przy zmianie parametrów, np. wykorzystując inne jądro. Z pewnością modele lepiej by sobie poradziły, jeżeli nie uwzględnić w analizowanych danych graczy tworzących mniej niż X akcji, którzy są zazwyczaj obrońcami i nie mają zbyt dużego wkładu ofensywnego. Przykładową wartością X na podstawie modeli można przyjąć 30.

Ponadto zmienna C odpowiadająca za regularyzację dopasowania krzywej musiała przyjmować duże wartości ze względu na duże zagęszczenie wartości graczy innych drużyn. Natomiast zmienny γ , czyli wpływ jednego

treningowego obiektu, powodowała bardzo nieprzewidywalne zmiany, tym samym była ciężka do własnoręcznego ustawienia.

Literatura

[1] FBRef. <https://fbref.com/en/>, 2023. Last accessed 10 June 2023.