

Analiza procentu zwycięstw drużyny NBA na podstawie transferów wykonanych w offseason

Igor Podgórniak, 272640

9 czerwca 2024

1 Wstęp

1.1 Wprowadzenie do tematu

Koszykówka to dynamiczna gra zespołowa, w której dwie drużyny, każda składająca się z pięciu zawodników, rywalizują o zdobycie jak największej liczby punktów, wrzucając piłkę do kosza przeciwnika. Wartość punktów zależy od miejsca, z którego oddano rzut. Poza zdobytymi punktami w koszykówce wyróżnia się inne ważne statystyki:

- Zbiórki - sytuacje, w których zawodnik przejmując piłkę po nieudanym rzucie, czyli po tym, jak piłka odbije się od obręczy lub tablicy.
- Asysty - przyznawane zawodnikowi, który poda piłkę do kolegi z drużyny w taki sposób, że ten ostatni zdobywa punkty bez konieczności dodatkowych podań czy dryblingu.
- Straty - sytuacje, w których drużyna traci posiadanie piłki na skutek błędów, takich jak niecelne podania, faul ofensywny, przekroczenie linii bocznej lub inne naruszenia przepisów.
- Przechwyty - sytuacje, w których zawodnik odbiera piłkę przeciwnikowi, najczęściej poprzez celne i szybkie działanie w obronie.
- Bloki - zaliczane, gdy zawodnik udaremni próbę rzutu przeciwnika, dotykając piłki w momencie jej wznoszenia się w kierunku kosza.
- Faule - naruszenia przepisów gry, które są popełniane przez zawodników w stosunku do przeciwników.

Poza wyżej wymienionymi, występuje jeszcze wiele innych, bardziej dogłębnych, którymi jednak nie będziemy się zajmować w projekcie. W koszykówce drużyny dzielą się na ligi. NBA to profesjonalna liga koszykówki w Stanach Zjednoczonych, uważana za najlepszą i najbogatszą ligę koszykówki na świecie. Składa się z 30 drużyn podzielonych na dwie konferencje: Wschodnią i Zachodnią, a każda konferencja jest dodatkowo podzielona na trzy dywizje. Podział na konferencje ma znaczenie przede wszystkim podczas sezonu regularnego, ponieważ drużyny klasyfikuje się w konferencjach oraz w fazie playoff, gdzie drużyny rywalizują najpierw w ramach swoich konferencji.

W NBA, podobnie jak w innych sportach występuje tak zwany offseason - okres między zakończeniem jednego sezonu a rozpoczęciem następnego. W tym czasie drużyny mogą dokonywać transferów zawodników, podpisywać kontrakty z wolnymi agentami oraz przygotowywać się do nowego sezonu poprzez treningi i mecze towarzyskie.

NBA wyróżnia się na tle innych lig draftem. Draft NBA to coroczne wydarzenie, w którym drużyny NBA wybierają nowych zawodników spośród graczy z college'ów, lig zagranicznych oraz innych młodych talentów. Jest to główny na pozyskanie przyszłych gwiazd ligi.

1.2 Opis problemu

Problemem wybranym przeze mnie do badania jest zależność między procentem zwycięstw drużyny w NBA, który jest bezpośrednim czynnikiem wpływającym na pozycję drużyny w konferencji a transferami, które drużyna wykonała przed sezonem, bądź w trakcie trwania poprzedniego sezonu. Analiza tego problemu może pomóc w pewnym stopniu ocenić zasadność transferów. Celem projektu jest zbadanie zmiany procentu wygranych spotkań przez drużynę w sezonie w stosunku do poprzedniego sezonu na podstawie danych dotyczących:

- drużyn:
 1. liczby punktów zdobytych na mecz
 2. liczby asyst zdobytych na mecz
 3. liczby bloków zdobytych na mecz
 4. liczby zbiórek zdobytych na mecz
 5. liczby przechwytów zdobytych na mecz
 6. liczby strat popełnionych na mecz
 7. liczby fauli personalnych popełnionych na mecz
- zawodników:
 1. liczby punktów zdobytych na mecz
 2. liczby asyst zdobytych na mecz
 3. liczby bloków zdobytych na mecz
 4. liczby zbiórek zdobytych na mecz
 5. liczby przechwytów zdobytych na mecz
 6. liczby strat popełnionych na mecz
 7. liczby fauli personalnych popełnionych na mecz
- zmian zawodników w drużynach

a także stworzenie modelu, który będzie estymował zmianę procentu wygranych drużyny na podstawie zawodników przychodzących i odchodzących z drużyny.

2 Dane

2.1 Źródło i zakres danych

Dane zostały uzyskane ze strony basketball-reference.com i zawierają informacje na temat statystyk drużyn oraz statystyk graczy. Statystyki drużyn i zawodników są podane na mecz. Dane dotyczące zawodników obejmują sezony od 1989-90 do 2021-22. Dane drużyn obejmują wszystkie lata funkcjonowania zespołów, zapewniając szeroki kontekst historyczny do analizy problemu. Oprócz informacji wymienionych w punkcie 2.1, dane zawierają również niewykorzystane w projekcie informacje takie jak wiek, waga, wzrost zawodników czy informacje dotyczące punktów za 2, za 3 i rzutów wolnych. Każdy wiersz dotyczący statystyk zawodnika zawiera również informację o drużynie, w której dany zawodnik grał w danym sezonie, a wiersze zawierające dane drużyn informację o ilości wygranych i przegranych spotkań.

2.2 Przetwarzanie wstępne (pre-processing)

Dane wykorzystywane do konstrukcji modelu posiadały puste komórki, które zostały zastąpione wartościami średnimi obliczonymi na podstawie pozostałych komórek w kolumnie. Na podstawie informacji dotyczących ilości wygranych i przegranych spotkań z pobranych danych, wyznaczono procent wygranych spotkań drużyny w danym sezonie i dodano jako nową kolumnę. Jeśli zawodnik występował w sezonie w więcej niż jednej drużynie, to wybrano drużynę, w której rozegrał większą ilość spotkań, a dane z innych drużyn nie zostały wzięte pod uwagę.

2.3 Przykładowe dane

Oczyszczone dane dotyczące zawodników

	Rk	Player	Tm	G	TRB	AST	STL	BLK	TOV	PF	PTS
0	1	Mark Acres	ORL	80	5.4	0.8	0.5	0.3	0.9	3.1	4.5
1	2	Michael Adams	DEN	79	2.8	6.3	1.5	0.0	1.8	1.7	15.5
2	3	Mark Aguirre	DET	78	3.9	1.9	0.4	0.2	1.6	2.6	14.1
3	4	Danny Ainge	SAC	75	4.3	6.0	1.5	0.2	2.5	3.2	17.9
4	5	Mark Alarie	WSB	82	4.6	1.7	0.7	0.5	1.2	2.7	10.5
5	6	Steve Alford	DAL	41	0.6	1.0	0.4	0.1	0.4	0.5	4.1
6	7	Randy Allen	SAC	63	2.2	0.4	0.3	0.3	0.4	1.6	3.7
7	8	Greg Anderson	MIL	60	6.2	0.4	0.5	0.9	1.3	2.9	8.8
8	9	Nick Anderson	ORL	81	3.9	1.5	0.9	0.4	1.7	1.7	11.5
9	10	Richard Anderson	CHH	54	2.4	1.0	0.4	0.2	0.5	1.2	4.3

Oczyszczone dane dotyczące drużyn

	Season	Lg	W	G	TRB	AST	STL	BLK	TOV	PF	PTS	win_percentage
0	2023-24	NBA	22	82	44.2	29.9	7.1	6.3	15.1	17.2	112.1	0.268
1	2022-23	NBA	22	82	43.7	27.2	7.0	3.9	15.3	19.9	113.0	0.268
2	2021-22	NBA	34	82	45.3	27.9	7.6	4.9	12.7	18.1	113.2	0.415
3	2020-21	NBA	33	72	43.9	24.4	7.0	5.1	11.4	18.0	111.1	0.458
4	2019-20	NBA	32	71	44.6	24.7	7.3	5.5	12.6	19.4	114.1	0.451
5	2018-19	NBA	48	82	44.7	24.5	6.1	4.7	12.1	18.1	111.7	0.585
6	2017-18	NBA	47	82	44.2	22.8	7.7	5.6	13.1	17.2	102.7	0.573
7	2016-17	NBA	61	82	43.9	23.8	8.0	5.9	13.4	18.3	105.3	0.744
8	2015-16	NBA	67	82	43.9	24.5	8.3	5.9	13.1	17.5	103.5	0.817
9	2014-15	NBA	55	82	43.6	24.4	8.0	5.4	14.0	19.1	103.2	0.671

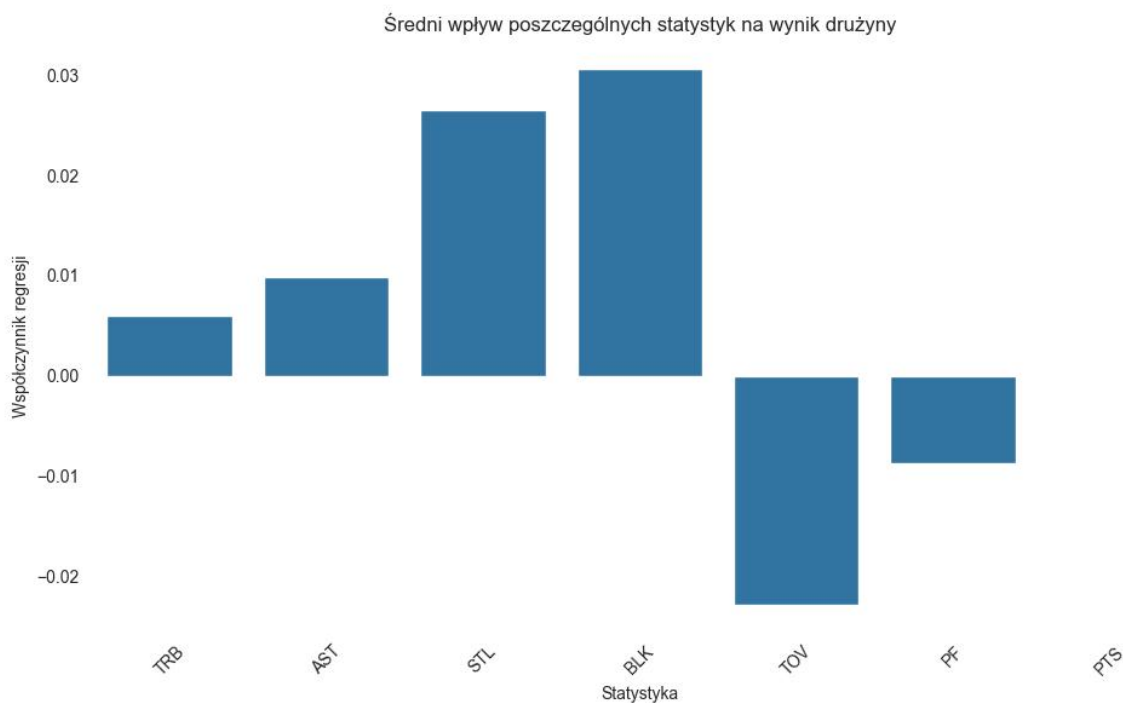
Dane dotyczące drużyn, wykorzystane w sposobie drugim

	Rk	G	TRB	AST	STL	BLK	TOV	PF	PTS	year	prev_year	team
0	-280.0	-66.0	-4.300000e+00	-5.900000e+00	-1.400000e+00	-8.000000e-01	-8.000000e-01	-5.100000e+00	-17.5	0.524	0.500	ATL
1	291.0	-1.0	2.100000e+00	2.700000e+00	1.600000e+00	0.000000e+00	7.000000e-01	4.400000e+00	1.5	0.463	0.524	ATL
2	-1000.0	-38.0	2.300000e+00	-3.800000e+00	-1.100000e+00	2.000000e-01	-1.100000e+00	-5.000000e-01	-6.1	0.524	0.463	ATL
3	-153.0	90.0	-1.700000e+00	1.800000e+00	6.000000e-01	4.000000e-01	-6.000000e-01	-1.900000e+00	1.7	0.695	0.524	ATL
4	260.0	113.0	2.000000e-01	3.300000e+00	1.300000e+00	6.000000e-01	1.700000e+00	5.700000e+00	-4.1	0.512	0.695	ATL
5	31.0	-60.0	-1.900000e+00	-1.500000e+00	-4.000000e-01	-4.000000e-01	1.100000e+00	-2.100000e+00	4.6	0.561	0.512	ATL
6	110.0	120.0	1.540000e+01	-8.000000e-01	-6.000000e-01	1.800000e+00	8.000000e-01	3.400000e+00	6.7	0.683	0.561	ATL
7	8.0	-53.0	-3.600000e+00	-1.000000e-01	-5.000000e-01	2.220446e-16	-2.000000e+00	-1.600000e+00	-6.3	0.610	0.683	ATL
8	180.0	-122.0	-1.400000e+00	-2.300000e+00	-1.400000e+00	-5.000000e-01	-1.600000e+00	-1.500000e+00	-3.9	0.620	0.610	ATL
9	-379.0	179.0	-1.400000e+00	4.800000e+00	-1.200000e+00	-1.000000e-01	2.200000e+00	1.800000e+00	10.7	0.341	0.620	ATL

3 Eksperymenty

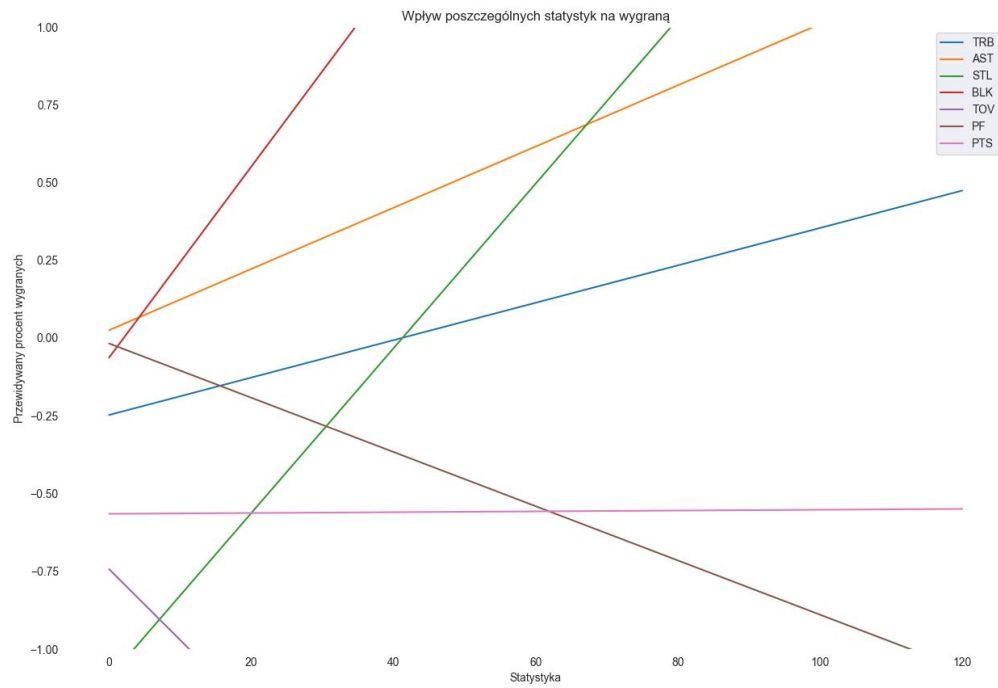
3.1 Analiza danych

W celu analizy danych, sprawdzono wpływ statystyk, które będą wzięte pod uwagę podczas późniejszego tworzenia modelu, na wynik drużyny. Do analizy wykorzystano współczynniki regresji liniowej modelu liniowego z biblioteki scikit-learn w Pythonie, na których trenowany był później model.



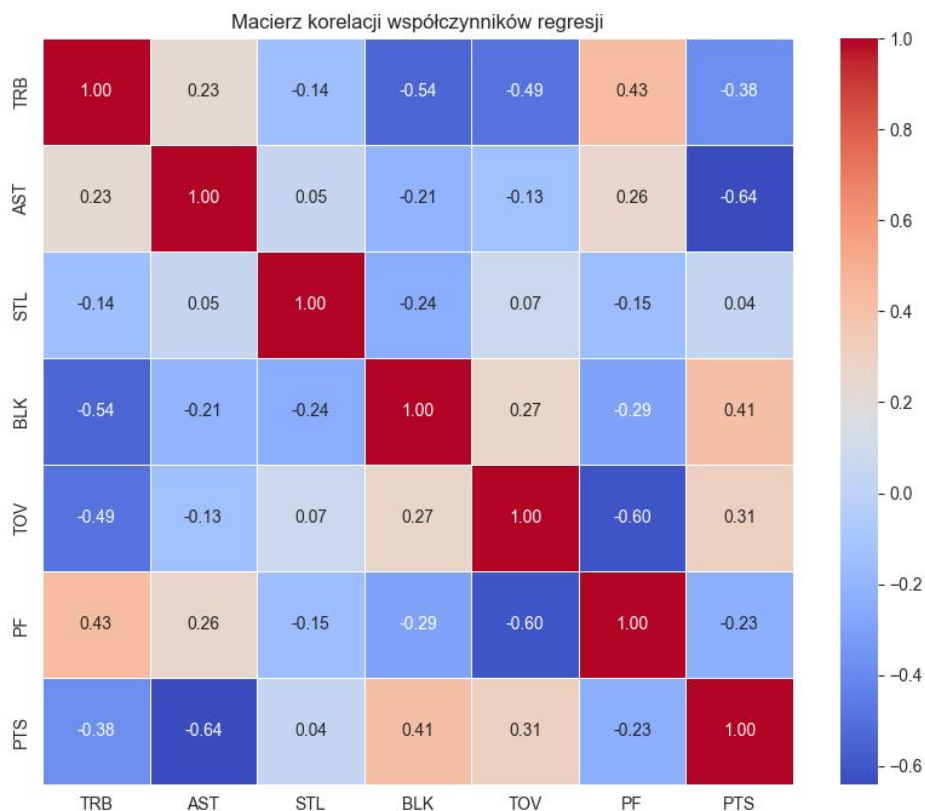
Na wykresie słupkowym przedstawiono wpływ, jaki poszczególne statystyki mają na procent wygranych zespołu w sezonie. Jak widać, największy wpływ na wynik drużyny w sezonie, mają zdobyte bloki i przechwyty, niewielki, ale pozytywny wpływ mają także zbiórki. Straty i faule osobiste wpływają negatywnie na procent zwycięstw. Punkty zdobywane na mecz mają bardzo niski wpływ na wynik zespołu.

Te statystyki zilustrowane zostały również za pomocą regresji liniowej:



Ten wykresy pokazuje, że mimo dużego pozytywnego wpływu niektórych statystyk na mecz na wynik drużyny w sezonie, ich brak może powodować odmienne skutki, tak jak np. w przypadku przechwytów i zbiórek. Ta zależność działa też w drugą stronę. Przy braku fauli osobistych, ta statystyka nie wpływa na procent wygranych zespołu.

Kolejny wykres przedstawia macierz korelacji współczynników regresji poszczególnych statystyk.

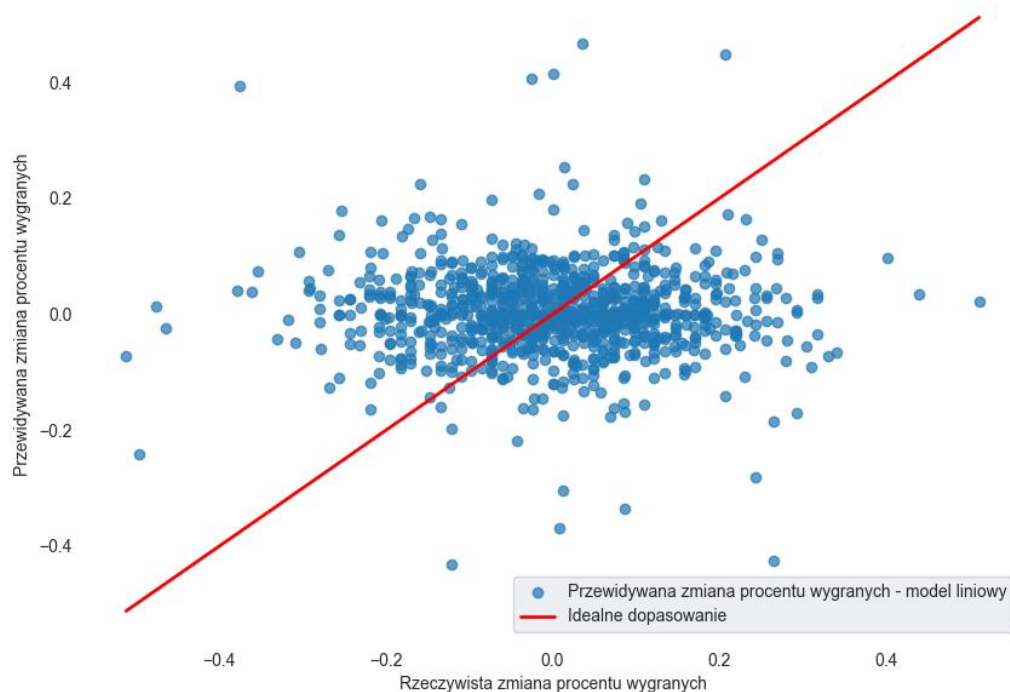


Na początku należy zaznaczyć, że dane te dotyczą jednej drużyny, nie są oparte na danych meczów i statystykach przeciwników. Na podstawie macierzy, można zauważyć kilka zależności między statystykami. Na pewno interesujący może być związek asyst i punktów na mecz. Zgodnie z danymi, kiedy rośnie liczba zdobytych punktów przez drużynę, zazwyczaj maleje liczba asyst i odwrotnie. Zaskakujący jest też brak większego wpływu liczby przechwytów na zdobywane punkty na mecz. Można także zobaczyć, że wraz z liczbą zbiórek, rośnie liczba fauli wykonywanych przez drużynę, co sugeruje, że walka zbiórek generuje faule.

Podsumowując, największą uwagę przykuwa bardzo niski, wręcz niezauważalny wpływ ilości zdobywanych punktów na mecz oraz znaczna przewaga statystyk defensywnych nad ofensywnymi. Te dane potwierdzają przewijającą się w świecie koszykówki tezę, że aktualnie zawodnicy zaniedbują defensywną grę ataku, na korzyść większej efektywności w ofensywie. Z tego powodu, samo zdobywanie punktów przez zawodnika, nie ma większego wpływu na przebieg meczu, ważniejsza jest jednak jego skuteczność oraz gra w defensywie. Ponadto negatywne zależności punktów z asystami i zbiórkami w macierzy korelacji, pokazują, że ta statystyka nie ma tak znaczącego wpływu na grę, ponieważ standardem stało się, że drużyny zdobywają podobne, wysokie ilości punktów na mecz.

3.2 Wyniki

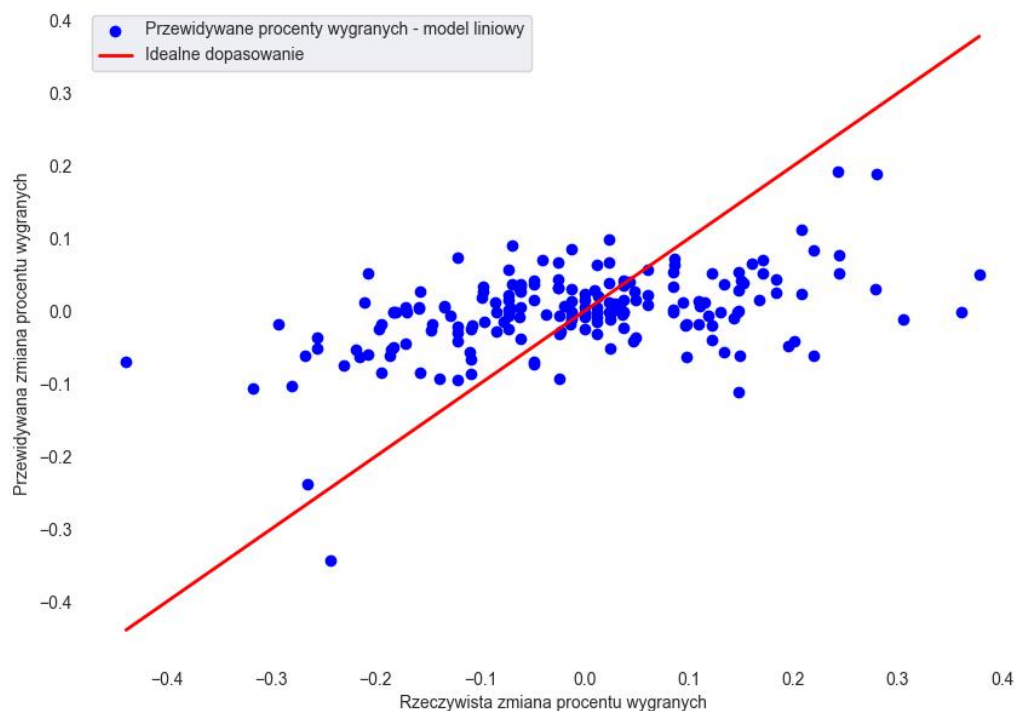
3.2.1 Sposób 1



W sposobie pierwszym wykorzystano modele użyte wcześniej do analizy danych. Na podstawie ich średnich współczynników regresji wyznaczono wpływ pojedynczego zawodnika na mecz. Następnie znając statystyki graczy, którzy opuścili drużynę i tych, którzy do niej dołączyli, zsumowano wpływ zawodnika z procentem wygranych spotkań drużyny z poprzedniego sezonu i porównano z wartością w poprzednim sezonie. Model regresji liniowej z biblioteki scikit-learn w pythonie przewidywał współczynniki regresji dla każdej zmiennej, które potem były uśredniane dla całej ligi. Najdokładniejsze wyniki znajdują się przy niskich zmianach procentu wygranych.

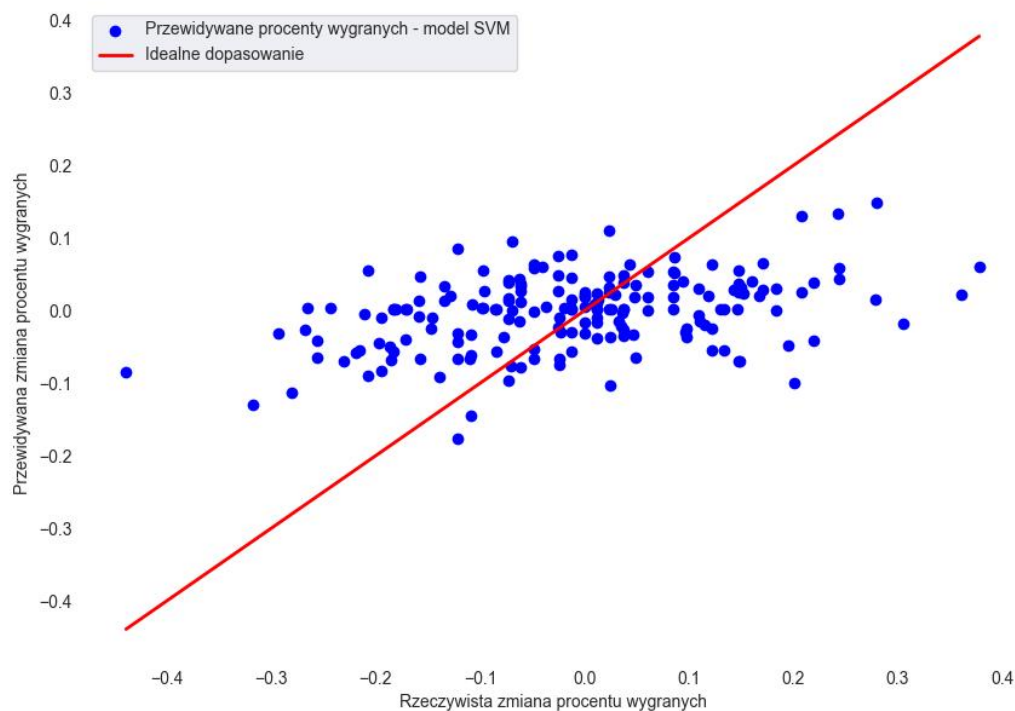
Obliczono dokładność modelu, wyznaczając jego błąd średniokwadratowy, którego największe wartości osiągały około 0.03.

3.2.2 Sposób 2



W drugim sposobie wykorzystano połączone dane dla wszystkich drużyn w celu wytrenowania modelu bezpośrednio danymi dotyczącymi transferów w offseason i ich wpływu na procent zwycięstw drużyny. Model którego użyto to model regresji liniowej z biblioteki scikit-learn w pythonie. Po kilkunastu uruchomieniach, można zauważyć znacznie mniejszy rozrzut punktów w zależności przewidywanych zmian procentu wygranych od rzeczywistych oraz nieznacznie zmieniony rozkład niż w poprzednim podejściu. Model zazwyczaj przewiduje, że wzrost lub spadek procentu wygranych będzie większy niż rzeczywisty.

Błędy średniokwadratowe modelu przyjmowały wartości w granicach od 0.010 do 0.018.



Do sposobu drugiego wykorzystano dodatkowo algorytm uczenia maszynowego SVM (Support Vector Machine). Zależność rzeczywistej zmiany procentu wygranych od tej przewidzianej przez model jest tutaj bardzo podobna jak w poprzednim modelu.

Błędy średniokwadratowe w tym modelu przyjmowały bardzo podobne wartości jak w modelu liniowym w sposobie 2.

3.2.3 Porównanie

W porównaniu obu metod można zauważyć, że drugi sposób, który bezpośrednio uwzględniał dane dotyczące transferów, osiągnął lepsze wyniki. Modele oparte na danych transferowych wykazały mniejszy rozrzut w przewidywanych zmianach procentu wygranych drużyny oraz mniejszy błąd średniokwadratowy w porównaniu do modeli opartych na analizie statystyk graczy z poprzedniego sezonu.

Wynika to z faktu, że drugi sposób bardziej precyzyjnie uwzględniał specyfikę transferów i bezpośrednio analizował wpływ nowych zawodników na drużynę. Brak pośrednich obliczeń, jak w przypadku pierwszej metody, zmniejszył ryzyko błędów i poprawił dokładność modeli.

3.3 Wnioski

Na podstawie przeprowadzonych eksperymentów można wywnioskować, że skuteczność transferów w kontekście procentu zwycięstw drużyny w sezonie NBA jest ściśle powiązana z umiejętnościami oraz formą zawodników transferowanych między sezonami lub w trakcie poprzedniego sezonu. Wyniki analizy statystyk graczy z poprzedniego sezonu pokazują, że ich osiągnięcia mają istotny wpływ na przewidywanie efektów ich późniejszych transferów.

Analiza danych potwierdza, że statystyki defensywne, takie jak bloki, przechwyty i zbiórki, mają większy wpływ na wynik drużyny w NBA niż statystyki ofensywne, takie jak zdobywane punkty. Negatywny wpływ na procent wygranych mają przede wszystkim straty i faule osobiste, co sugeruje, że skuteczna obrona oraz dyscyplina w grze są kluczowe dla sukcesu drużyny.

W obu badanych metodach wykorzystano modele regresji liniowej oraz uczenia maszynowego. Druga metoda, która uwzględniała bezpośrednio dane dotyczące transferów, okazała się skuteczniejsza i dokładniejsza niż pierwsza metoda. Wyniki modeli były bardziej zbliżone do rzeczywistych danych, co sugeruje, że analiza bezpośrednio oparta na danych transferowych może być bardziej efektywna w prognozowaniu zmian procentu wygranych drużyny.

Wytrenowane modele mogą być użyteczne dla menedżerów drużyn NBA w podejmowaniu decyzji dotyczących transferów graczy. Mogą one pomóc w przewidywaniu, jakie zmiany w składzie drużyny mogą wpłynąć na jej wyniki w kolejnym sezonie. Pomimo tego, że modele osiągnęły stosunkowo niskie błędy średniokwadratowe, należy pamiętać, że istnieją inne czynniki, które mogą wpływać na wyniki drużyny, takie jak kontuzje, taktyka gry przyjęta przez trenera, atmosfera w szatni czy forma indywidualnych zawodników. Dlatego też modele te powinny być traktowane jedynie jako narzędzia wspomagające proces podejmowania decyzji.