

Przewidywanie pensji zawodników NBA na podstawie ich statystyk

Mikołaj Olesiński

Informatyka Stosowana, Politechnika Wrocławska

272702@student.pwr.edu.pl

Maj 2024

Spis treści

1	Wstęp	2
1.1	Istota problemu	2
1.2	Dane	2
2	Analiza danych	4
2.1	Pensja zawodników	4
2.2	Pozycje zawodników	6
2.3	Macierz korelacji danych	7
3	Model Regresji	9
3.1	Wybranie odpowiedniego modelu	9
3.2	Analiza wyników przewidywań	11
3.3	Odziaływanie zmiennych na pensje w RF	12
3.4	Model dla ograniczonych danych	13
3.5	Model dla poszczególnych pozycji	15
3.5.1	Rozgrywający	16
3.5.2	Skrzydłowy	17
3.5.3	Środkowy	18
4	Wnioski	19

1 Wstęp

1.1 Istota problemu

NBA (National Basketball Association) to prestiżowa, profesjonalna liga koszykówki, która cieszy się ogromną popularnością na całym świecie, szczególnie w Stanach Zjednoczonych. Koszykówka to dynamiczna gra zespołowa, której mecze przyciągają miliony widzów zarówno na żywo, jak i przed telewizorami oraz ekranami komputerów. Popularność NBA stale rośnie, przyciągając coraz więcej fanów z różnych zakątków świata.

Wzrost zainteresowania ligą wiąże się nie tylko z samymi rozgrywkami, ale także z osobami zawodników. Kibice są zafascynowani nie tylko ich umiejętnościami sportowymi, ale również życiem prywatnym oraz zarobkami. Pensje graczy NBA stały się jednym z najczęściej dyskutowanych tematów w mediach sportowych i społeczno-sciowych. Wysokie kontrakty i ich wpływ na zespoły, jak również na indywidualnych graczy, budzą powszechną ciekawość i spekulacje.

Czy można przewidzieć wysokość zarobków zawodnika NBA wyłącznie na podstawie jego statystyk? Które z nich są najważniejsze w określaniu wynagrodzenia? Czy wzrost i waga mają istotny wpływ na pozycję i płace zawodnika?

W celu odpowiedzi na te pytania, przeanalizujemy, jakie czynniki wpływają na wynagrodzenia graczy NBA. Skoncentrujemy się głównie na statystykach zawodników podczas gry, takich jak średnia punktów na mecz, czy ilość zagranych meczy. Przy użyciu modelu regresji przeanalizujemy dane dotyczące pensji zawodników na przestrzeni lat, aby zidentyfikować kluczowe czynniki determinujące zarobki w NBA. Nasza analiza pomoże zrozumieć, które statystyki mają największy wpływ na wysokość wynagrodzeń oraz w jakim stopniu można przewidzieć pensje graczy na podstawie tych informacji.

1.2 Dane

Do analizy przyjrzymy się danym z sezonów 2000 - 2023. Dane na temat wynagrodzeń graczy zostały pobrane ze strony ESPN [1], a dane na temat statystyk graczy i informacji np. na temat ich wagi zostały pobrane z NBA API 2.

Działamy na 6536 rekordach graczy

Zmienne na których będziemy się skupać i ich jednostki(każda zmienna dotyczy pojedynczego sezonu gracza):

Nazwa	Rozwinięcie	Opis
Weight		Ile waży dany zawodnik w funtach
Height		Ile ma wzrostu dany zawodnik w cm
Position		Określa pozycję, na której gra dany zawodnik
Season_Exp	Season experience	Ile lat gra zawodnik w danym sezonie
Draft number		Określa miejsce w drafcie danego zawodnika (im niżej, tym lepiej)
Team abbreviation		Określa drużynę, w jakiej zawodnik grał
GP	Games played	Ile zawodnik zagrał meczy
GS/G	Games Started/Games	Współczynnik gier, w których zawodnik był pierwszym składzie
MIN/G	Minutes/Game	Średnia liczba minut na mecz, jaką grał zawodnik
REB/G	Rebounds/Game	Średnia ilość zbiórek na mecz
AST/G	Assist/Game	Średnia ilość asyst na mecz
STL/G	Steals/Game	Średnia ilość przechwytyń na mecz
BLK/G	Blocks/Game	Średnia ilość bloków na grę
TOV/G	Turnovers/Game	Średnia ilość strat na mecz
PTS/G	Points/Game	Średnia ilość punktów na mecz
Salary		Pensja danego zawodnika na rok, podana jest w milionach i w dolarach

Dane zostały odpowiednio zredagowane:

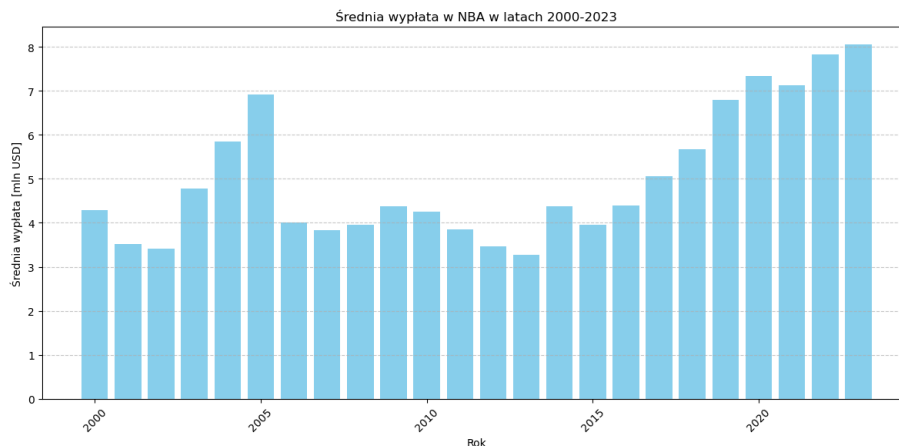
- Wszystkie rekody, które nie miały którejs z tych zmiennych zostały usunięte
- Wszystkie zmienne Draft Number, które miały wartość "undrafted" zostały zmienione na 61

Rekordy, które nie miały którejs ze zmiennej zostały usunięte, aby zapewnić spójność danych i zapobiec jakimkolwiek zakłóceniom z tego wynikających. Natomiast zmiana wartości "undrafted" na 61 w kolumnie Draft Number umożliwia reprezentację braku wyboru w drafcie w formie liczbowej, co ułatwia analizę danych. Jest ona zmieniona akurat na liczbę 61, ponieważ 60, to najwyższa możliwa pozycja w drafcie.

2 Analiza danych

2.1 Pensja zawodników

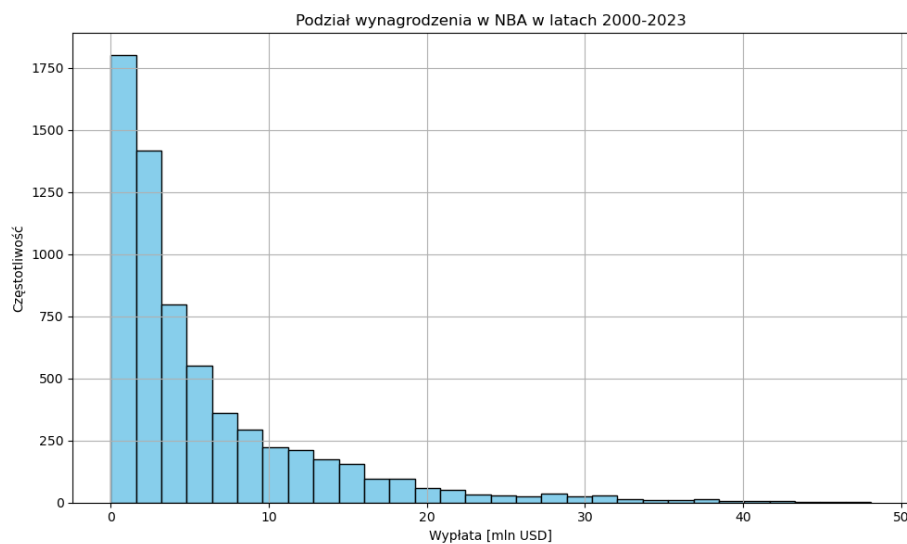
Skupimy się na analizie pensji zawodników NBA, korzystając zarówno z graficznej reprezentacji danych, jak i statystyk opisowych. Naszym celem jest zrozumienie rozkładu pensji wśród graczy ligi.



Rysunek 1: Wykres słupkowy przedstawiający średnią wypłatę w NBA

Analiza wykresu pokazuje, że od roku 2000 do 2016 średnie wynagrodzenie zawodników NBA utrzymywało się na stosunkowo stabilnym poziomie. Niemniej jednak, można zauważyć wyraźny wzrost w okresie od 2003 do 2005 roku, a następnie spadek do wcześniejszego poziomu, po czym przez kolejne 13 lat średnia pensja nie zmieniała się znacząco. Dopiero od roku 2016 zauważalny jest kolejny wzrost, który wydaje się być bardziej wyraźny i schodkowy.

Ten wzrost może być skutkiem różnych czynników, w tym zwiększonych przychodów związanych z ligą NBA. Wzrost popularności ligi, nowe umowy telewizyjne oraz rosnące zainteresowanie międzynarodowe mogły przyczynić się do zwiększenia dostępnych środków na wynagrodzenia zawodników. Ponadto, zmiany w umowach zbiorowych, które ustalają zasady finansowania i podziału dochodów, również mogły wpłynąć na wzrost pensji zawodników od roku 2016.



Rysunek 2: Podział wynagrodzenia w NBA

Statystyka	Pensja
Minimum	0.01
Maksimum	48.07
Średnia	5.99
Odchylenie standardowe	6.92

Tabela 2: Statystyki dotyczące pensji zawodników NBA

Na podstawie wykresu i danych statystycznych możemy stwierdzić, że przeważająca większość zawodników NBA otrzymuje roczne pensje w przedziale od 0 do 5 milionów dolarów. Dalej, obserwuje się wyraźny spadek liczby graczy wraz z wzrostem pensji, szczególnie widocznym w przedziale między 20 a 48 milionów dolarów. Średnia pensja wynosi 6 milionów dolarów, jednakże duże odchylenie standardowe, niemal 7 milionów dolarów, wskazuje na znaczące zróżnicowanie w poziomie zarobków między zawodnikami.

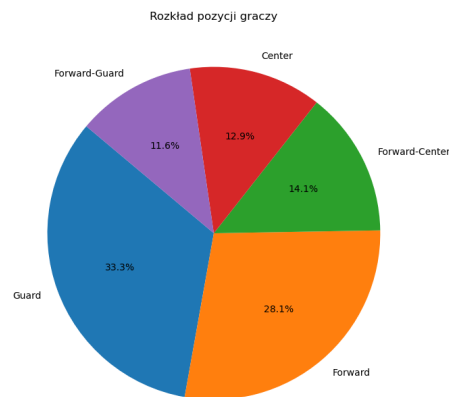
Ta analiza sugeruje, że większość graczy w NBA nie należy do tej elitarnie opiewanej grupy supergwiazd, którą często podkreśla się w mediach. Zamiast tego, liga ta głównie składa się z zawodników, których pensje mieszczą się w zakresie od 0 do 10 milionów dolarów rocznie, ale którzy nie zawsze są szeroko rozpoznawani lub omawiani w mediach. To podkreśla ogromną rolę całej społeczności zawodników w NBA,

którzy, mimo że mogą nie być gwiazdami na pierwszych stronach gazet, odgrywają kluczową rolę w codziennej grze i sukcesie ligi.

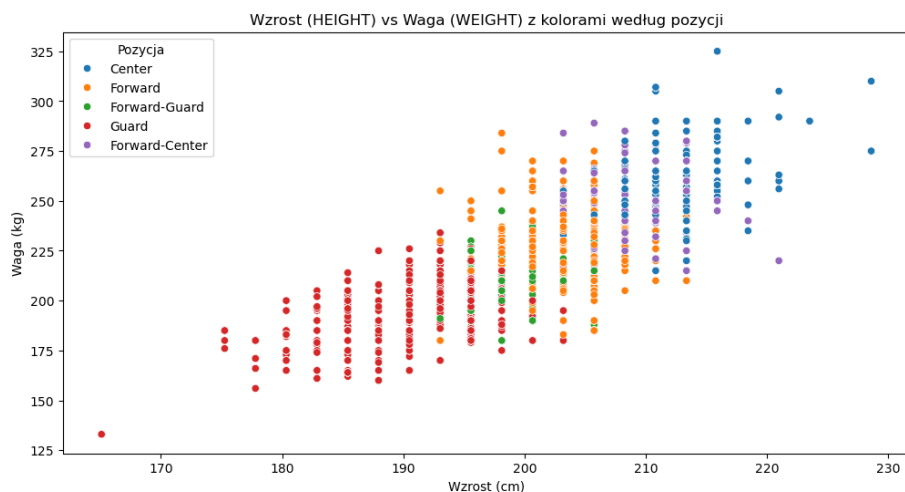
2.2 Pozycje zawodników

W NBA możemy wyróżnić 5 pozycji:

- **Forward (F)** – skrzydłowy, wszechstronni gracze, zdolni do gry zarówno blisko kosza, jak i na zewnątrz. Mają duży udział w zdobywaniu punktów i obronie.
- **Guard (G)** – rozgrywający, kluczowi gracze odpowiedzialnymi za organizację ataku, rzuty na kosz i obronę. Są zazwyczaj szybcy i zwinni.
- **Center (C)** – środkowy, Centrowie są wysocy i potężni, dominujący w grze blisko kosza. Ich zadaniem jest kontrola deski, blokowanie rzutów i obrona niskiego bloku.
- **Forward-Guard (F-G)** łączy cechy skrzydłowego(F) i rozgrywającego(G). Pełnią kluczową rolę w ataku, organizując grę, penetrując obronę i wykonując rzuty z różnych pozycji. Dodatkowo są istotni w obronie dzięki szybkości i zwinności.
- **Forward-Center (F-C)** łączy cechy skrzydłowego(F) i środkowego(C). Dominują blisko kosza dzięki wysokości i sile, ale posiadają także umiejętności gry na zewnątrz. Pełnią kluczową rolę w kontrolowaniu deski, blokowaniu rzutów oraz zdobywaniu punktów z różnych pozycji na boisku



Rysunek 3: Wykres kołowy rozkładu pozycji graczy w NBA



Rysunek 4: Wykres punktowy wzrostu od wagi zawodników w NBA podzielony na pozycje

Z wykresu kołowego 2.2 możemy zauważyć, że dominującą grupę wśród zawodników NBA stanowią skrzydłowi i rozgrywający. Natomiast środkowi stanowią najmniejszy odsetek. Jest to zgodne z typowym układem pięciu graczy na boisku, gdzie drużyna składa się z dwóch rozgrywających, dwóch skrzydłowych i jednego środkowego.

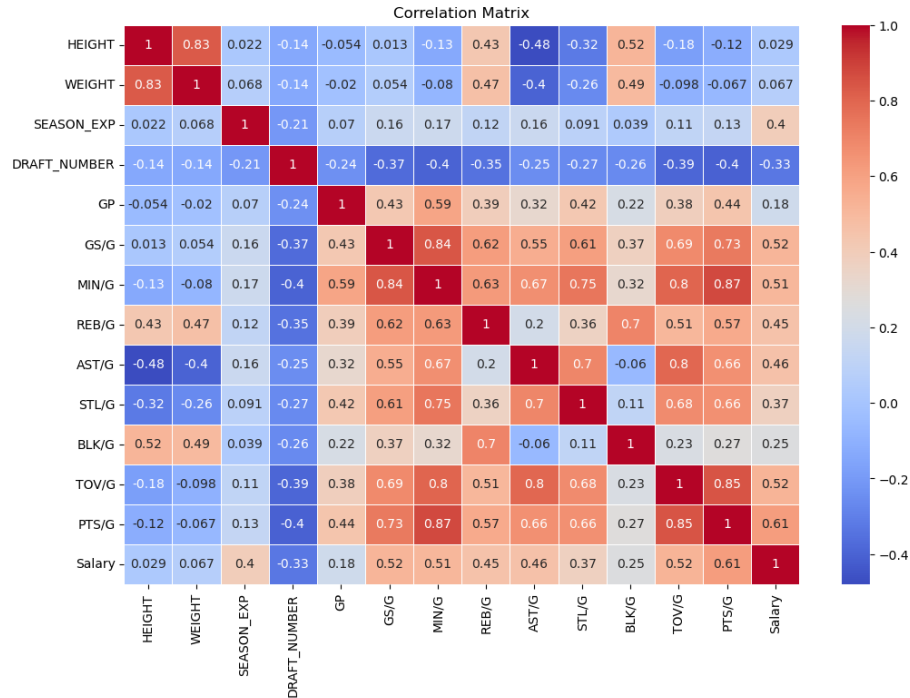
Z drugiego wykresu 2.2 natomiast możemy zaobserwować, że wzrost graczy przeważnie koreluje z ich pozycją na boisku. Środkowi są z reguły najwyżsi, rozgrywający natomiast najniżsi, a skrzydłowi zajmują miejsce między nimi. Jest to związane z rolą każdej pozycji na boisku, gdzie środkowi dominują pod koszem, gdzie przydaje się głównie wzrost, rozgrywający przeważnie pełnią rolę dyrygenta zespołu i nadają dynamikę, do czego potrzebna jest szybkość i zwinność, a skrzydłowi mają wszechstronność, umożliwiającą im zarówno grę na zewnątrz, jak i w pobliżu obręczy.

Widzimy również, że zdecydowana większość zawodników ma wzrost między 190 a 220 cm, co pokazuje charakterystykę fizyczną większości graczy NBA. Wysoki wzrost jest jednym z kluczowych czynników determinujących, kto będzie miał większą kontrolę nad piłką i przewagę na boisku.

2.3 Macierz korelacji danych

Macierz korelacji danych to narzędzie statystyczne używane do analizy zależności między zmiennymi w zbiorze danych. Jest to tabela, w której każda komórka przed-

stawia współczynnik korelacji między dwoma zmiennymi. Użyjemy go właśnie po to by zauważyć te współczynniki między naszymi zmiennymi



Rysunek 5: Macierz korelacji

Po analizie powyższej macierzy, możemy dojść do kilku ciekawych wniosków:

- Draft number ma wszędzie ujemną korelację, czyli zawodnicy z niższym numerem draftu, czyli ci co zostali wybrani jako pierwsi to m.in są wyżsi, rzucają więcej punktów, grają więcej minut, czy więcej zarabiają
- Widzimy bardzo wysoki dodatni współczynnik korelacji przy rzuconych punktach i minutach, bo aż 0.87 i przy rzuconych punktach i stratach, 0.85. Czyli zawodnicy co rzucają więcej punktów, grają więcej, ale też popełniają więcej strat
- Można zauważyć ciekawą zależność przy wzroście i asystach, wyżsi zawodnicy mają mniej asyst
- Ciekawą zależnością jest to, że doświadczenie zawodnika nie ma większego wpływu na statystyki danego zawodnika

- Z macierzy możemy wywnioskować, że największy wpływ na pensje zawodnika mają rzucone punkty

3 Model Regresji

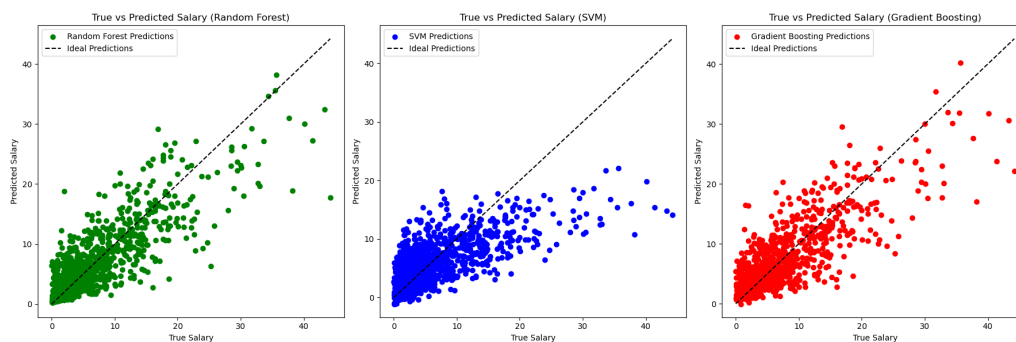
3.1 Wybranie odpowiedniego modelu

W tej sekcji dokonujemy wyboru odpowiedniego modelu regresji do analizy naszych danych. Porównamy trzy popularne modele: Random Forest, Support Vector Machine (SVM) oraz Gradient Boosting

Aby określić, który model jest najlepszy, podzielimy nasze dane na zestaw treningowy i testowy, przetrenujemy każdy z trzech modeli na zestawie treningowym, a następnie ocenimy ich wydajność na zestawie testowym przy użyciu odpowiednich miar, takich jak RMSE (Root Mean Squared Error) i R^2 . Analizujemy również znaczenie zmiennych, aby zidentyfikować, które z nich mają największy wpływ na wynik modelu.

Model	RMSE	R^2
Random Forest	3.44	0.72
SVM	4.57	0.50
Gradient Boosting	3.49	0.71

Tabela 3: Miary danych modeli



Rysunek 6: Wykres przedstawiający przewidywania danych modeli

Z tabeli można wywnioskować, że najlepszym modelem jest Random Forest, ponieważ:

- Posiada najniższą wartość RMSE (3.44), co oznacza, że jego przewidywania są najbliższe rzeczywistym wartościom.
- Wartość R^2 wynosi 0.72, co wskazuje, że model dobrze wyjaśnia zmienność danych.

Gradient Boosting również osiąga dobre wyniki, z RMSE równym 3.49 i R^2 równym 0.71, jednak nieco gorsze niż Random Forest. Model SVM wypada najgorzej, z najwyższym RMSE (4.57) i najniższym R^2 (0.50), co oznacza, że jego przewidywania są najmniej dokładne.

Podsumowując, na podstawie tych wyników, model Random Forest wydaje się być najlepszym wyborem do analizy naszych danych.

3.2 Analiza wyników przewidywań

Poniżej przedstawiamy tabelę z przewidywaniami modelu Random Forest dla wybranych zawodników, gdzie model przewidział pensje zawodników, które były znacznie różne od rzeczywistych wartości.

Na końcu przedstawiamy tabelę z przewidywaniami modelu Random Forest dla zawodników, gdzie model przewidział pensje, które były bardzo bliskie rzeczywistym wartościom.

Imię	Sezon	Prawdziwa Pensja	Pensja przez RF	Różnica
Russell Westbrook	2021-22	44.21	17.76	26.45
Rudy Gobert	2022-23	38.17	18.87	19.30
Joel Embiid	2017-18	25.25	6.29	18.96
Nicolas Batum	2018-19	24.00	9.30	14.70
Evan Turner	2019-20	18.61	4.18	14.43

Tabela 4: Top 5 przewidywań, gdzie model przewidział zbyt niska pensje

Imię	Sezon	Prawdziwa Pensja	Pensja przez RF	Różnica
Chauncey Billups	2011-12	2.0	18.75	-16.75
Tracy McGrady	2006-07	16.9	29.14	-12.24
Peja Stojakovic	2006-07	10.8	21.87	-11.07
Pau Gasol	2015-16	7.45	18.32	-10.87
Derrick Rose	2018-19	1.51	12.24	-10.73

Tabela 5: Top 5 przewidywań, gdzie model przewidział zbyt wysoką pensję

Imię	Sezon	Prawdziwa Pensja	Pensja przez RF	Różnica
Dean Wade	2020-21	1.52	1.52	0.00
Solomon Alabi	2010-11	0.67	0.67	0.00
Daniel Ochefu	2016-17	0.54	0.54	0.00
Isaac Bonga	2021-22	1.67	1.67	0.00
Ben Simmons	2019-20	8.11	8.11	0.00

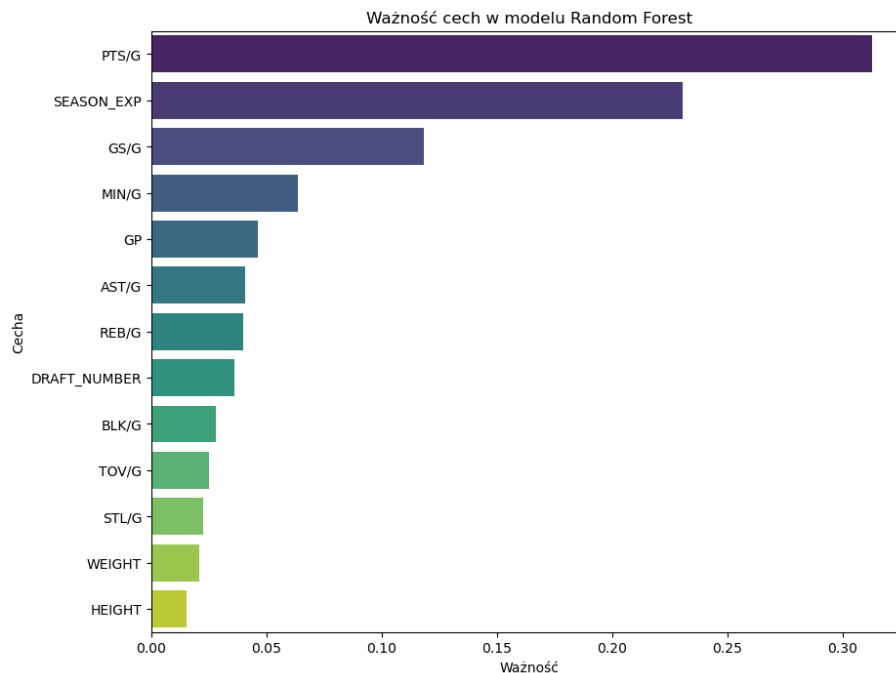
Tabela 6: Top 5 przewidywań, gdzie model przewidział pensję bardzo bliską rzeczywistej wartości

Z tych tabel możemy wywnioskować, że model Random Forest jest w stanie przewidzieć dokładnie pensje zawodników. Dla danych testowych, model przewidział pen-

sje z błędem mniejszym niż milion dolarów dla 593 rekordów. Jednakże, możemy zauważyć, że model przewiduje dokładniej pensje dla zawodników, których zarobki wynoszą między 0,5 a 5 miliony dolarów. W przypadku zawodników zarabiających więcej, model popełnia większe błędy. Może to wynikać z faktu, że liga składa się głównie z zawodników o niższych zarobkach, podczas gdy gwiazdy, które zarabiają duże pieniądze, stanowią mniejszość i są wyjątkiem.

3.3 Odziaływowanie zmiennych na pensje w RF

Warto również przyjrzeć się które zmienne dla modelu RF najbardziej wpływają na to ile zawodnik zarabia. Czy wzrost w lidze ma aż takie znaczenie pod względem zarobków? A może to tylko punkty mają znaczenie? W odpowiedzi na te pytania pomoże nam poniższy wykres:



Rysunek 7: Wykres przedstawiający wagę zmiennych w przewidywaniu pensji NBA przez model RF dla wszystkich danych

Możemy zauważyć, że najważniejszą cechą w przewidywaniu pensji zawodników NBA jest liczba punktów zdobywanych na mecz. Wskazuje to, że zawodnicy, którzy zdobywają więcej punktów, są bardziej cenieni i zarabiają więcej. Kolejną istotną

zmienną jest doświadczenie zawodnika, co sugeruje, że bardziej doświadczeni zawodnicy mają tendencję do wyższych zarobków. Liczba gier, w których zawodnik zaczyna w pierwszej piątce, oraz średnia liczba minut spędzonych na boisku również mają istotne znaczenie, co pokazuje, że zawodnicy regularnie grający i spędzający więcej czasu na boisku mają wyższe wynagrodzenie.

Inne zmienne, takie jak liczba rozegranych meczy, liczba asyst na mecz, liczba zbiórek na mecz, czy numer draftu mają umiarkowane znaczenie. Z kolei zmienne fizyczne, takie jak wzrost i waga, mają stosunkowo niewielki wpływ na zarobki w porównaniu do statystyk związanych z wydajnością na boisku.

Podsumowując, wyniki modelu RF pokazują, że najważniejszymi czynnikami wpływającymi na pensje zawodników NBA są ich osiągnięcia na boisku i doświadczenie, a niekoniecznie ich cechy fizyczne.

3.4 Model dla ograniczonych danych

Jak można zauważyć w 3.2 nasz model głównie działa dla typowych graczy z których składa się liga NBA, czyli graczy którzy nie rzucają najwięcej punktów, nie są gwiazdami, ale wciąż grają i mają jakiś wpływ na grę. Niestety łączy się to z tym, że nie zarabiają najwięcej. Dlatego nasze nowe dane skupią się na graczach którzy zarabiają między 0.5mln a 10mln dolarów. Zmniejszy to ilość rekordów o około 20% i zostanie nam ich 4907.

Więc dla tych ograniczonych danych tak się prezentują statystyki pensji:

Statystyka	Pensja
Minimum	0.51
Maksimum	9.94
Średnia	3.36
Odchylenie standardowe	2.41

Tabela 7: Statystyki dotyczące pensji zawodników NBA z ograniczonych danych

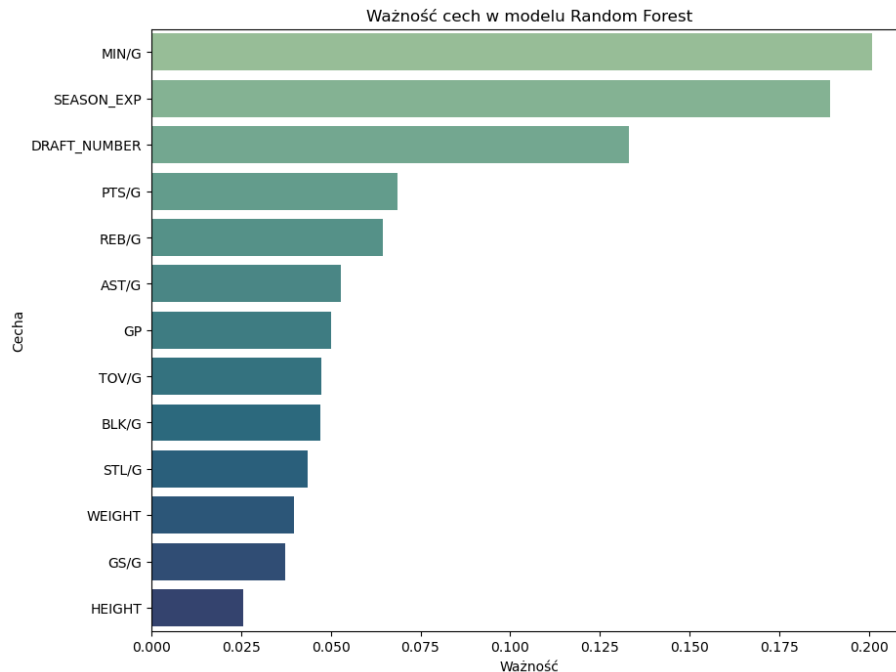
a tak prezentuje się wynik modelu RF:

- RMSE: 1.66
- R^2 : 0.50

RMSE dla ograniczonych danych wynosi 1,66, co jest znacząco lepszym wynikiem w porównaniu do RMSE dla pełnych danych, które wynosiło 3,70. Mniejszy RMSE

wskazuje na mniejsze błędy przewidywań, co oznacza, że model lepiej przewiduje zarobki graczy w węższym przedziale.

Wartość R^2 dla ograniczonych danych wynosi 0,59, co jest niższą wartością w porównaniu do R^2 dla pełnych danych, które wynosiło 0,71. Oznacza to, że model wyjaśnia mniej wariacji w danych, gdy jest ograniczony do węższego przedziału zarobków. Jednakże, biorąc pod uwagę niższy przedział zarobków, niższe R^2 jest spodziewane, ponieważ jest mniejsze zróżnicowanie danych.



Rysunek 8: Wykres przedstawiający wagę zmiennych w przewidywaniu pensji NBA przez model RF dla ograniczonych danych

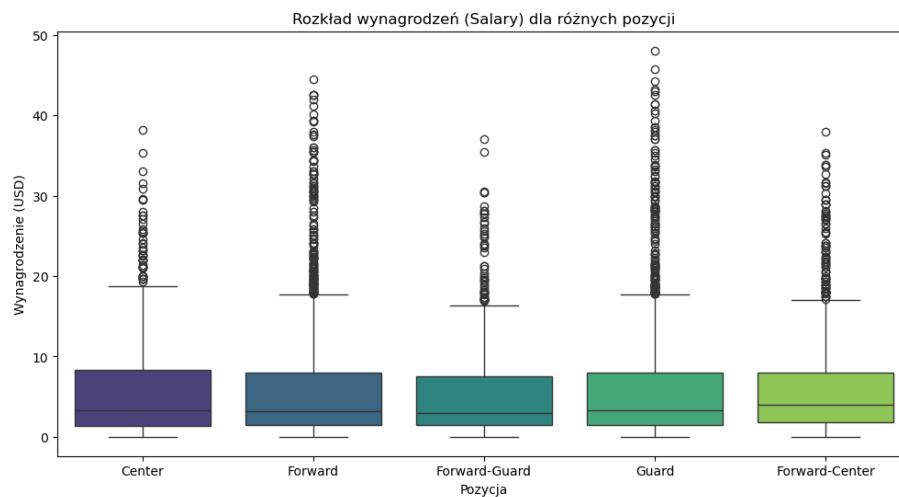
Wniosek, jaki można wyciągnąć z eliminacji wcześniejszych "gwiazd", które dominowały w zdobywaniu punktów podczas meczy, jest taki, że teraz kluczową rolę odgrywają inne czynniki. Ważnymi wskaźnikami stają się teraz zagrane minuty, doświadczenie zawodników oraz ich numer w drafcie, a nie rzucone punkty. To sugeruje, że zespoły kładą większy nacisk na równomierną dystrybucję odpowiedzialności i umiejętności między wszystkich graczy, zamiast polegać na pojedynczych gwiazdach.

Dodatkowo, istotne stają się wszystkie inne aspekty gry, takie jak zbiórki, asysty, bloki i przechwyty, obok punktów zdobywanych przez zawodników. Może to wskazywać na zmianę strategii w podejściu do gry, gdzie równowaga między różnymi

aspektami gry staje się kluczowa dla sukcesu zespołu.

3.5 Model dla poszczególnych pozycji

Pozycje w koszykówce stanowią istotny element strategii i taktyki, choć ich znaczenie może się różnić w porównaniu na przykład z piłką nożną, gdzie role są bardziej zdefiniowane. W analizie danych dotyczących zawodników NBA, zbadano modele predykcyjne dla różnych pozycji, aby lepiej zrozumieć czynniki wpływające na wynagrodzenia graczy. Sekcja ta prezentuje wyniki modeli dla rozgrywających, skrzydłowych oraz środkowych, ukazując istotne zależności pomiędzy cechami graczy a ich pensjami.



Rysunek 9: Wykres przedstawiający rozkład pensji od pozycji 4

Rozkład wynagrodzeń pokazuje, że chociaż różne pozycje w NBA mają zbliżone mediany wynagrodzeń, istnieje znaczna zmienność w zarobkach w obrębie każdej pozycji. Najwyżej opłacani zawodnicy znajdują się na każdej pozycji, co pokazuje, że każda pozycja jest ważna dla danej drużyny i żadna pozycja nie dominuje pod względem płacy.

Zarobki dla pozycji rozgrywającego wykazują największą zmienność, ale również osiągają największe wartości co może sugerować, że rola rozgrywającego w drużynie jest bardzo zróżnicowana, w porównaniu z innymi pozycjami.

Podsumowując, wynagrodzenia zawodników NBA są bardzo zróżnicowane w zależności od pozycji, jednak kluczowe znaczenie mają indywidualne cechy i wartość dla drużyny, co powoduje duże rozbieżności w zarobkach w obrębie każdej pozycji.

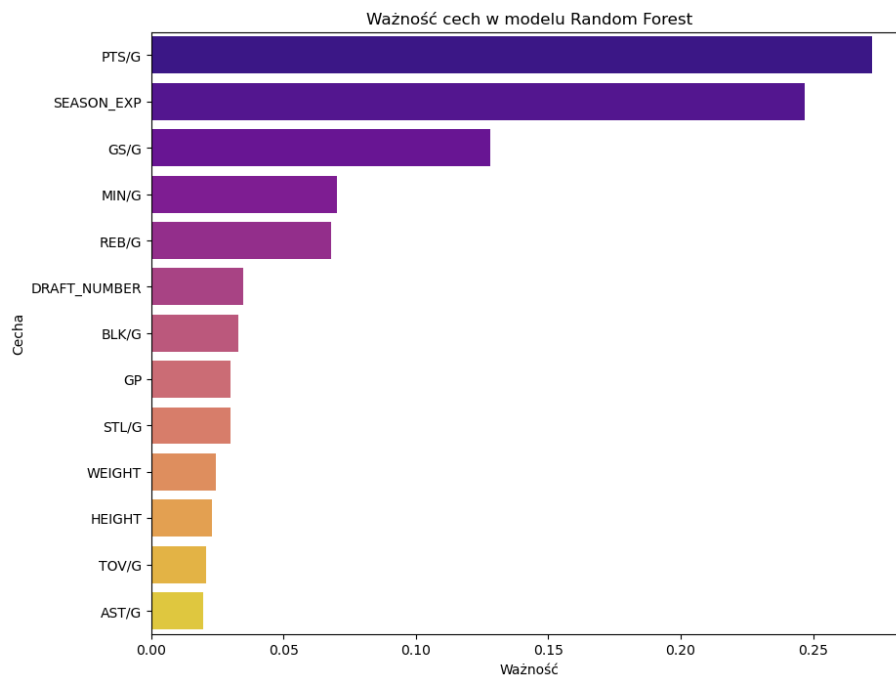
Teraz skupimy się na modelach dla danych pozycji i właśnie jakie indywidualne cechy wpływają na pensje gracza z danej pozycji

3.5.1 Rozgrywający

Jeżeli podstawimy tylko statystyki dla rozgrywających pod model RF to otrzymamy takie wyniki:

- RMSE: 3.61
- R^2 : 0.74

Więc możemy stwierdzić, że nie ma większej różnicy, niż gdy porównujemy model ze wszystkimi danymi. Jest minimalnie większy RMSE, ale prawdopodobnie wynika to z mniejszej ilości danych. Warto jednak się przyjrzeć jakie czynniki mają największe znaczenie, gdy przewidujemy pensje dla rozgrywającego:



Rysunek 10: Wykres przedstawiający wagę zmiennych w przewidywaniu pensji NBA przez model RF dla rozgrywających

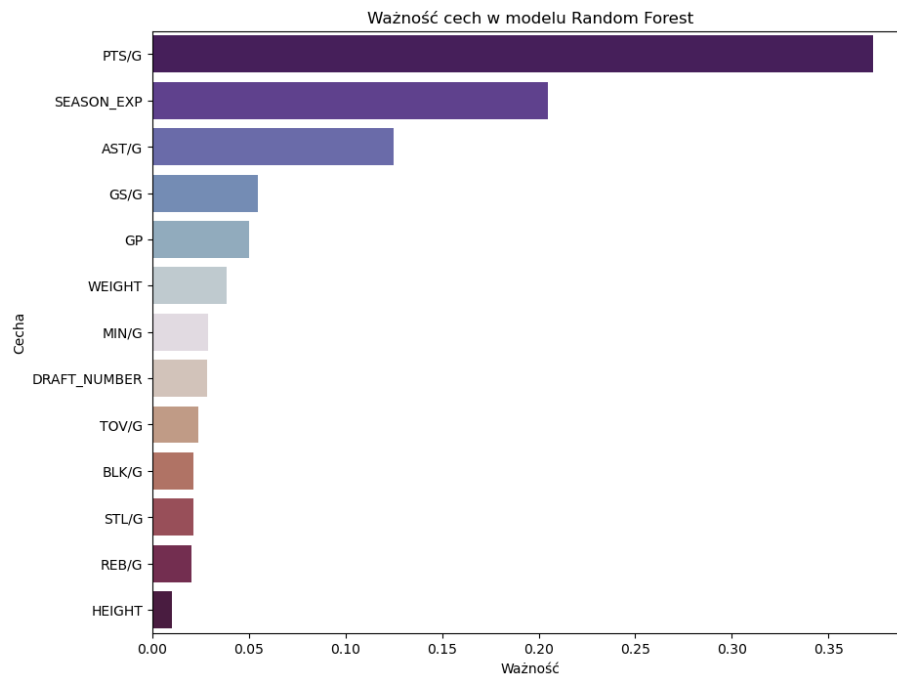
Jak możemy zauważyć istotność cech jest bardzo podobna jak dla 3.3 jedyną większą różnicą jest to, że większe znaczenie mają również zbiórki, które są rzadko spotykane u rozgrywającego

3.5.2 Skrzydłowy

Dla modelu RF z danymi tylko ze skrzydłowymi dostajemy takie wyniki:

- RMSE: 3.52
- R^2 : 0.73

Więc tak samo jak wspomnieliśmy wcześniej, błędy są bardzo podobne, różnica wynika z innej ilości danych. Natomiast jeżeli spojrzymy na ważność zmiennych w modelu RF dla tych danych:



Rysunek 11: Wykres przedstawiający ważność zmiennych w przewidywaniu pensji NBA przez model RF dla skrzydłowego

Teraz możemy zauważyć, że punkty dla skrzydłowego to przeważająco najważniejszy czynnik. Wynika to z tego, że jest to jego główna rola. Widzimy również, że

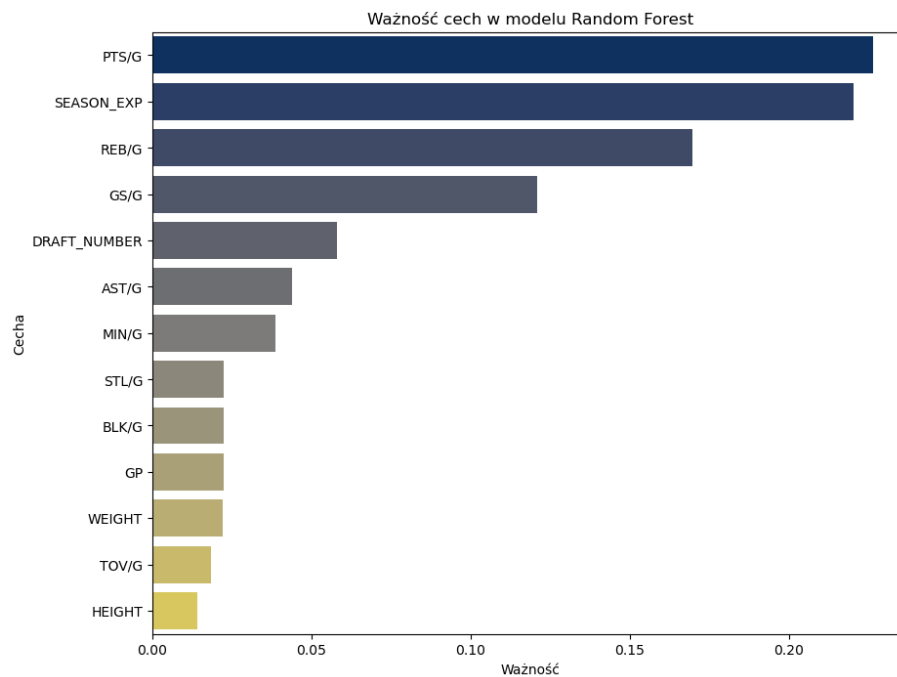
dużą rolę odgrywają asysty. Udowadnia to, że głównym zadaniem skrzydłowego jest zdobywanie lub przyczynianie się do zdobywania punktów i głównie na tej podstawie jest oceniany.

3.5.3 Środkowy

Jeżeli spojrzymy na wyniki modelu RF dla środkowego:

- RMSE: 4.07
- R^2 : 0.56

Tak słabe wyniki wynikają z tego, że jak zobaczyliśmy w 2.2 środkowych mamy najmniej, więc również wyniki będą gorsze.



Rysunek 12: Wykres przedstawiający wagę zmiennych w przewidywaniu pensji NBA przez model RF dla środkowego

Rola środkowego, różni się najbardziej z wszystkich pozycji w koszykówce. Jego głównym celem nie są punkty, a obrona i zbiórki. Widzimy, że w przeciwieństwie do wcześniejszych pozycji, doświadczenie jest prawie tak samo ważne jak punkty.

Jednak również zbiórki odgrywają jedną z bardziej znaczących ról. Tak samo ważne jest kiedy środkowy jest w pierwszym składzie ponieważ zwykle drużyna ma 2, 3 środkowych na mecz.

Analiza modeli predykcyjnych dla poszczególnych pozycji w koszykówce wykazała istotne różnice oraz podobieństwa w czynnikach determinujących wysokość pensji graczy. Mimo minimalnych różnic w wynikach pomiędzy modelami dla różnych pozycji, można zauważyć pewne stałe tendencje. Na przykład, punktacja pozostaje kluczowym czynnikiem dla każdej pozycji. Możemy jednak zauważyć pewne zależności jak dla skrzydłowych największe znaczenie ma przyczynienie się do zdobycia punktów, podczas gdy dla środkowych większe znaczenie ma doświadczenie i umiejętności obronne.

4 Wnioski

Przeprowadzona analiza pokazała, że statystyki związane z wydajnością na boisku są kluczowe dla przewidywania wynagrodzeń zawodników NBA. Chociaż cechy fizyczne mają pewien wpływ, to doświadczenie i osiągnięcia na boisku takie jak rzucone punkty, zbiórki, czy asysty głównie decydują o zarobkach zawodnika.

Model Random Forest okazał się najskuteczniejszym narzędziem do przewidywania pensji na podstawie analizowanych zmiennych, szczególnie dla zawodników w średnim przedziale zarobków. Model dla ograniczonych danych udowodnił, że gdy wyeliminujemy najlepszych zawodników to zespoły kładą na równomierną dystrybucję odpowiedzialności i umiejętności między wszystkich graczy i wtedy każdy aspekt (jak np. zbiórki, asysty, bloki) jest tak samo ważny. W węższym przedziale zarobkowym model ten bardzo dobrze sobie radził lecz niestety eliminował "gwiazdy", najbardziej popularnych zawodników o których zarobkach chcielibyśmy wiedzieć najwięcej.

Dowiedzieliśmy się również, że dla pozycji mediana zarobków jest bardzo podobna ale różne czynniki decydują o tym ile zawodnik będzie zarabiał. Dla rozgrywających głównym czynnikiem są punkty i doświadczenie, ale na przykład do środkowego dochodzi jeszcze ilość zbiórek.

Końcowo dowiadujemy się, że możemy obliczyć na podstawie statystyk średnie zarobki zawodnika ze średnim błędem o 3.44 mln dolarów a węższym przedziale nawet o 1.66 mln dolarów. Jednak do zarobków danego zawodnika dochodzi jeszcze wiele innych czynników jak np. popularność na social mediach, czy historia urazów.

Ostatecznie, wyniki te mogą być użyteczne dla menedżerów drużyn oraz agentów zawodników, pomagając w negocjacjach kontraktów oraz ocenie wartości graczy lub

dla fanów koszykówki którzy są zainteresowani losem swojego ulubionego koszykarza lub drużyny.

References

- [1] <https://www.espn.com/nba/salaries>
- [2] https://github.com/swar/nba_api
- [3] *Model choice inspired from:* <https://towardsdatascience.com/predicting-nba-salaries-with-machine-learning-ed68b6f75566>
- [4] *Plot inspired from:* <https://www.kaggle.com/code/rikdifos/nba-players-salary-prediction>