



Politechnika Wrocławska

Wydział Matematyki

Kierunek studiów: Matematyka Stosowana

Specjalność: –

Praca dyplomowa – inżynierska

## TYTUŁ PRACY DYPLOMOWEJ

Michał Ceraży

słowa kluczowe:  
tutaj podajemy najważniejsze słowa kluczowe (łącznie nie powinny być dłuższe niż 150 znaków).

krótkie streszczenie:

Tutaj piszemy krótkie streszczenie pracy (nie powinno być dłuższe niż 530 znaków).

Opiekun pracy dyplomowej	dr hab. inż. Krzysztof Burnecki	.....	.....
	Tytuł/stopień naukowy/imię i nazwisko	ocena	podpis

*Do celów archiwalnych pracę dyplomową zakwalifikowano do:\**

*a) kategorii A (akta wieczyste)*

*b) kategorii BE 50 (po 50 latach podlegające ekspertyzie)*

*\* niepotrzebne skreślić*

pieczęćka wydziałowa

Wrocław, rok 2018





Wrocław University  
of Science and Technology

Faculty of Pure and Applied Mathematics

Field of study: Applied Mathematics

Specialty: –

Engineering Thesis

## TYTUŁ PRACY DYPLOMOWEJ W JĘZYKU ANGIELSKIM

Michał Ceraży

keywords:

tutaj podajemy najważniejsze słowa kluczowe w języku angielskim (łącznie nie powinny być dłuższe niż 150 znaków)

short summary:

Tutaj piszemy krótkie streszczenie pracy w języku angielskim (nie powinno być dłuższe niż 530 znaków).

Supervisor	dr hab. inż. Krzysztof Burnecki	.....	.....
	Title/degree/name and surname	grade	signature

*For the purposes of archival thesis qualified to:\**

*a) category A (perpetual files)*

*b) category BE 50 (subject to expertise after 50 years)*

*\* delete as appropriate*

stamp of the faculty

Wrocław, 2018



# Spis treści

<b>Wstęp</b>	<b>3</b>
<b>1 Opis ligi NBA, jak wygląda sezon</b>	<b>5</b>
<b>2 Teoria, matematyka</b>	<b>7</b>
2.1 Zasada bootstrap . . . . .	7
<b>3 Metodologia, algorytmy</b>	<b>9</b>
3.1 Model uśredniony . . . . .	10
3.2 Model rywalizacji . . . . .	10
3.3 Model symulacji fazy pucharowej . . . . .	11
<b>4 Wyniki, porównanie modeli</b>	<b>13</b>
<b>5 Wnioski</b>	<b>15</b>
<b>Podsumowanie</b>	<b>17</b>
<b>Dodatek</b>	<b>19</b>



# Wstęp

Dzięki powszechnemu dostępowi do internetu i rozpowszechnieniu kultury masowej amerykańska liga koszykarska NBA zyskała popularność na całym świecie, przyciągając do siebie najlepszych graczy i masy fanów. Dzięki nieprzewidywalności i złożoności tego sportu podejmowano wiele prób przewidywania wyników rozgrywek, które często toczyły się inaczej, niż by zakładano (najlepszym tego przykładem może być sezon 2003/2004, kiedy to nisko notowani Detroit Pistons pokonali faworytów w postaci Los Angeles Lakers). Najlepsi analitycy sportowi starają się ?analizować? każdy aspekt gry i jego wpływ na sytuację na boisku, lecz nikt do tej pory nie był w stanie zaproponować skutecznego modelu opisującego przebieg rozgrywek. W niniejszej pracy inżynierskiej podjęto próbę przewidzenia rezultatów wybranego sezonu ligi NBA przy pomocy informacji o wynikach poszczególnych drużyn w poprzednich sezonach.





# Rozdział 1

## Opis ligi NBA, jak wygląda sezon

National Basketball Association (NBA) została założona 6 czerwca 1946 roku. Pierwotnie była znana jako Basketball Association of America, a swoją obecną nazwę zyskała w roku 1949, kiedy to wchłonęła rywalizującą National Basketball League. Od 2004 roku w NBA gra 30 zespołów, 29 z USA i 1 z Kanady. Liga podzielona jest na dwie konferencje po 15 drużyn, te natomiast składają się z dywizji po 5 ekip. TU WSTAW OBRAZ Z MAPKĄ!!!!!!!!!!!!!!!!!!!!. Nazwy wszystkich drużyn z podziałem na konferencje i dywizje:

Konferencja Wschodnia		
Atlantic Division	Southeast Division	Central Division
Boston Celtics	Atlanta Hawks	Chicago Bulls
Brooklyn Nets (dawniej New Jersey Nets)	Charlotte Hornets (dawniej Charlotte Bobcats)	Cleveland Cavaliers
New York Knicks	Miami Heat	Detroit Pistons
Philadelphia 76ers	Orlando Magic	Indiana Pacers
Toronto Raptors	Washington Wizards	Milwaukee Bucks

Konferencja Zachodnia		
Northwest Division	Southwest Division	Pacific Division
Denver Nuggets	Dallas Mavericks	Golden State Warriors
Minnesota Timberwolves	Houston Rockets	Los Angeles Clippers
Oklahoma City Thunder (dawniej Seattle SuperSonics)	Memphis Grizzlies	Los Angeles Lakers
Portland Trail Blazers	New Orleans Pelicans	Phoenix Suns
Utah Jazz	San Antonio Spurs	Sacramento Kings

Sezon w NBA składa się z dwóch części: zasadniczej i następującej po niej pucharowej (playoffs). W sezonie zasadniczym każda drużyna rozgrywa 82 mecze, grając z każdym innym zespołem od 2 do 4 gier. Terminarz wyznaczany jest wedle następujących reguł:

1. drużyny z różnych konferencji grają ze sobą 2 spotkania (1 na wyjeździe i 1 na własnym boisku),
2. drużyny z tej samej dywizji grają ze sobą 4 spotkania (2 na wyjeździe i 2 na własnym boisku),
3. drużyny z tej samej konferencji oraz różnych dywizji grają ze sobą 3 albo 4 spotkania (przynajmniej po jednym na wyjeździe i własnym boisku).

Mecze koszykówki nie mogą zakończyć się remisem (w razie remisu po regulaminowym czasie gry rozgrywa się dogrywki aż do wyłonienia zwycięzcy). Po zakończeniu sezonu następuje wspomniana wyżej faza pucharowa; wchodzi do niej po 8 najlepszych zespołów z każdej konferencji (w razie takiej samej ilości zwycięstw dla obu zespołów decydują mecze bezpośrednie pomiędzy nimi). W tej fazie drużyny grają ze sobą maksymalnie 7 meczów, czyli zespół, który pierwszy wygra 4 mecze, przechodzi do następnego etapu. W fazie Playoff jasno zdefiniowane są lokalizacje odgrywania spotkań — lepszy bilans zwycięstw w sezonie zasadniczym skutkuje przewagą parkietu. Seria spotkań grana jest w formacie 2–2–1–1–1, czyli mecze numer 1, 2, 5 i 7 grane są u lepszej z drużyn. Przy doborze przeciwników w tej fazie bierze się pod uwagę pozycję w tabeli konferencji: drużyna z miejsca pierwszego gra z zespołem o ósmym bilansie w danej konferencji, druga z siódmą, i tak dalej. Zwycięzca serii przechodzi do następnego etapu z czterema drużynami, po którym następują finały konferencji — najlepsze drużyny ze swoich konferencji spotykają się w finałach NBA. Dla lepszego zrozumienia systemu rozgrywek Playoff ZAMIESZCZONO DRZEWKO PONIŻEJ!!!!!!!!!!!!!!

# Rozdział 2

## Teoria, matematyka

Monte Carlo, rozkład jednostajny, bootstrap, rozkład, estymacja, metod nieparametryczne, boxplot, dystrybuanta, wartość oczekiwana? test shapiro-wilka, gęstość rozkładu, rozkład normalny, qqplot?

### 2.1 Zasada bootstrap



# Rozdział 3

## Metodologia, algorytmy

Czy znając wyniki zakończonych rozgrywek jesteśmy w stanie przewidzieć rezultaty przyszłych zawodów?

Dane, które będziemy wykorzystywać do symulacji sezonu zostały zebrane ze strony [basketballreference.com](http://basketballreference.com) i przerobione w odpowiedni sposób CZY MAM PODAWAĆ TRANSFORMACJE DANYCH, JAK DO TEGO DOSZŁO?

Na potrzeby tej pracy zebrano wyniki starć pomiędzy drużynami poczynając od sezonu 2004/2005 aż do 2017/2018. Początkowo symulowano rozgrywki w sezonie 2014/2015 w celu wybrania najlepszego modelu, a następnie skorzystano z niego, aby przewidzieć wyniki rozgrywek we wciąż trwającym sezonie 2018/2019. Posiadając ilość wygranych jednej drużyny z drugą na przestrzeni lat dokonano następujących transformacji danych: W zależności od interwału czasowego, jaki będziemy rozpatrywać, zebrano wyniki w określonych rozgrywkach (na przykład, przy wyznaczaniu wyników sezonu 2014/2015 i interwałe 5 lat, używać będziemy danych z lat 2009 do 2014). Dzięki uzyskanej w ten sposób liczbie wygranych w możemy stosunek zwycięstw do porażek dla wybranych zespołów (załóżmy, że Boston Celtics i Atlanta Hawks grali ze sobą 10 razy, Jastrzębie wygrały zaledwie 4 razy, dlatego też w starciu z Celtami jej stosunek wygranych do przegranych wynosi 0.4). Po zastosowaniu tej metody dla wszystkich zespołów uzyskamy macierz o rozmiarze 30 wierszy i kolumn zawierającą prawdopodobieństwa na wygraną z każdym zespołem w lidze.

Podczas prób symulacji dokonano intuicyjnego założenia, wedle którego największy wpływ na postawę sezonu mają rozgrywki bezpośrednio go poprzedzające. W tym celu dobrano system wag — z powodu dynamicznych zmian w lidze, najstarsze sezony otrzymują najmniejszą rangę, która stopniowo zwiększa się, im bliżej do zawodów rozpatrywanych w symulacji. Ważona ilość zwycięstw  $Z_i$   $i$ -tej drużyny  $D_i$  z  $j$ -tą drużyną  $D_j$  wynosi

$$Z_{ij} = \sum_{k=1}^n (1 + x \cdot k) \cdot R_{ijk}, \quad (3.1)$$

gdzie  $n$  to ilość sezonów, z których zaciągamy dane,  $x$  ustalona waga kolejnych rozgrywek, a  $R_{ijk}$  to wynik starć drużyny  $D_i$  z drużyną  $D_j$  w  $k$ -tym sezonie.

Podczas symulacji program przechodzi przez terminarz rozgrywek — drużyna gra z przeciwnikiem tyle razy, ile spotkań wyznaczono w rozkładzie. Algorytmy losowania opisano szczegółowo w ROZDZIALE Z MODELAMI

Z powodu losowości w rozkładzie sezonu (drużyny z jednej konferencji grające ze sobą 3 albo 4 razy w zależności od sezonu) skorzystano z gotowych terminarzy dla symulowanych sezonów.

### 3.1 Model uśredniony

Pierwszy ze stworzonych modeli polega na obliczeniu ogólnego stosunku zwycięstw do porażek dla każdej drużyny w wybranym okresie — wszystkie wygrane zespołu zostają podzielone przez łączną liczbę rozegranych spotkań, wynikiem czego jest liczba z przedziału  $[0, 1]$  określana jako  $P_i$ , gdzie  $i$  to  $i$ -ta drużyna. Algorytm symulowania wyników spotkań między drużynami wygląda następująco:

1. wstaw  $i = 1$ 
  - (a) wstaw  $j = i$
  - (b) znajdź drużyny  $D_i$  i  $D_j$
  - (c) odczytaj średnie ilości zwycięstw  $W_i$  i  $W_j$  dla drużyn  $D_i$  i  $D_j$
  - (d) wyznacz prawdopodobieństwo zwycięstwa  $W_{ij}$  przez drużynę  $D_i$  równe  $W_{ij} = \frac{W_i}{W_i + W_j}$
  - (e) w terminarzu znajdź liczbę spotkań  $S_{ij}$  pomiędzy drużynami  $D_i$  i  $D_j$ 
    - i. symuluj liczbę  $U$  z rozkładu jednostajnego  $U \sim U[0, 1]$
    - ii. jeżeli  $W_{ij} \leq U$ , to zwiększ licznik zwycięstw drużyny  $D_i$ , w przeciwnym razie zwiększ licznik zwycięstw drużyny  $D_j$
    - iii. powtórz  $S_{ij}$  razy
  - (f) wstaw  $j = j + 1$
2. wstaw  $i = i + 1$
3. jeżeli  $i > 30$ , to zakończ algorytm

Przy testowaniu skuteczności tego modelu

### 3.2 Model rywalizacji

Drugi z zaproponowanych modeli zakłada zwracanie uwagi na historyczne wyniki przeciwko konkretnej drużynie. W zawodowym sporcie niejednokrotnie można trafić na zażarte rywalizacje między dwoma klubami lub zwykłą łatwość w pokonaniu szczególnego przeciwnika. Algorytm symulowania wyników spotkań między drużynami wygląda następująco:

1. wstaw  $i = 1$ 
  - (a) wstaw  $j = i$
  - (b) znajdź drużyny  $D_i$  i  $D_j$
  - (c) odczytaj z macierzy wyników stosunek zwycięstw  $W_{ij}$  drużyny  $D_i$  przeciw drużynie  $D_j$
  - (d) w terminarzu znajdź liczbę spotkań  $S_{ij}$  pomiędzy drużynami  $D_i$  i  $D_j$ 
    - i. symuluj liczbę  $U$  z rozkładu jednostajnego  $U \sim U[0, 1]$
    - ii. jeżeli  $W_{ij} \leq U$ , to zwiększ licznik zwycięstw drużyny  $D_i$ , w przeciwnym razie zwiększ licznik zwycięstw drużyny  $D_j$
    - iii. powtórz  $S_{ij}$  razy

- (e) wstaw  $j = j + 1$
- 2. wstaw  $i = i + 1$
- 3. jeżeli  $i > 30$ , to zakończ algorytm

### 3.3 Model symulacji fazy pucharowej

Po symulacji całego sezonu, czyli 1230 spotkań, 8 najlepszych drużyn z każdej konferencji przechodzi do fazy Playoff, gdzie toczy rozgrywki zgodnie z systemem opisanym we WSTĘPIE. Na tym etapie rozgrywek symulacja spotkań różni się od części zasadniczej: zamiast jednego z zasugerowanych wcześniej modeli korzysta się ze wcześniejszej symulacji fazy zasadniczej. W celu oddania trendów panujących w wygenerowanych rozgrywkach (a mianowicie potencjalnych kontuzjach, spadkach lub zwyżkach formy), użyta zostaje jedynie informacja o ilości wygranych przed rozpoczęciem Playoffów. Algorytm symulowania tej fazy jest postaci:

1. wybierz drużyny  $D_i$  i  $D_j$
2. odczytaj symulowane ilości zwycięstw  $W_i$  i  $W_j$  dla wybranych drużyn  $D_i$  i  $D_j$
3. wyznacz prawdopodobieństwo zwycięstwa  $W_{ij}$  przez drużynę  $D_i$  równe  $W_{ij} = \frac{W_i}{W_i + W_j}$
4. wstaw liczniki zwycięstw  $Z_i = 0$  i  $Z_j = 0$ 
  - (a) symuluj liczbę  $U$  z rozkładu jednostajnego  $U \sim U[0, 1]$
  - (b) jeżeli  $W_{ij} \leq U$ , wstaw  $Z_i = Z_i + 1$ , w przeciwnym razie wstaw  $Z_j = Z_j + 1$
  - (c) powtarzaj dopóki  $Z_i = 4$  lub  $Z_j = 4$
5. jeżeli  $Z_i = 4$ , to przenieś drużynę  $D_i$  do następnego etapu, w przeciwnym razie przenieś drużynę  $D_j$

Ilości zwycięstw drużyn w kolejnych symulowanych rozgrywkach są zapisywane i zapamiętywane, podobnie jak informacje o przejściach do kolejnych faz rozgrywek pucharowych.





# Rozdział 4

## Wyniki, porównanie modeli

modelowanie okresem, sezonem, wagami, długością próbki? symulacja dla 2018-19?



# Rozdział 5

## Wnioski

gęstości symulacji powinny mieć rozkład normalny?



# Podsumowanie



# Dodatek

tabele z prawdopodobieństwami, terminarz sezonu