

# Przewidywanie wyników meczów NBA

Jan Gorycki 304051

Jakub Górka 313240

Przemysław Kacprzak 313269

Kamil Kubiak 313293

1 czerwca 2023

## 1 Wstęp

NbaPredictor to narzędzie, służące do przewidywania wyników meczów ligi NBA. Celem aplikacji jest przy obstawianiu meczy ligi na stronach bukmacherskich za pomocą podpowiedzi dotyczącej przewidywanych szans na wynik meczu.

## 2 Dane wykorzystane do stworzenia modelu

Użyto danych dostępnych na platformie kaggle: <https://www.kaggle.com/datasets/wyattowalsh/basketball>. Dane to 16 tabel z danymi związanymi z przeszłymi rozgrywkami.

## 3 Eksploracja danych

W trakcie powstawania modelu próbowane były różne zestawienia danych, te bardziej obszerne uwzględniające takie statystyki jak próby rzutów drużyn, próby rzutów za 3 punkty oraz wszystkie punkty zdobyte przez drużynę w ciągu meczu. Testowane również było kilka podejść z wyborem rekordów, które nas interesują, z jednej strony nie chcemy, żeby mecze sprzed 80 lat miały taki sam wpływ jak mecze z ostatniego miesiąca, naturalne jest, że tamte spotkania były rozgrywane przez zawodników, którzy już dawno zakończyli kariery, a drużyny, które ówczesnie dominowały często nie są w tym momencie na szczycie, ale zbyt surowe ograniczenie rekordów też powodowałoby bardzo mały zbiór danych, przez co zdecydowaliśmy się na jeszcze inną opcję, czyli wprowadzenie nowej kolumny określającej formę obu drużyn przed przystąpieniem do spotkania, dzięki temu każde spotkanie (poza 5 najstarszymi, bo tam informacje o formie są niemożliwe do skonstruowania z oczywistych powodów) możemy wziąć pod uwagę w naszym modelu, co znacząco poprawia jego działanie.

## 4 Działanie modelu

Został wykorzystany model z biblioteki tensorflow (dokładniej DNNclassifier z parametrem *hidden\_units* = [32,32]), a wybrane do niego kolumny po ekstrakcji danych i różnych podejściach to:

- informacja o tym, która drużyna była gospodarzem
- informacja o tym która drużyna była gościem
- wynik meczu
- obliczone formy obu drużyn

Model po trenowaniu umożliwił nam poprzez wybranie drużyn obliczenie prawdopodobieństw wygrania każdej z nich. Parametr *accuracy*, który był obliczany na zestawie testowym wynosił w okolicach 80%, ale nie była to informacja, która była przesadnie brana pod uwagę, a przy pracy nad modelem jako kryterium oceny kierowaliśmy się błędem względem szacowań bukmacherów, których to model uznaliśmy za idealny.

## 5 Skrypt ekstrakcji informacji o nadchodzących meczach

Aby sprawdzać jakie mecze odbędą się w najbliższym czasie przygotowano skrypt pobierający ze strony `flashscore.pl` dane na temat najbliższych spotkań. Skrypt został przygotowany w technologii puppeteer w języku JavaScript, a jego wynikiem jest plik o formacie danych:

```
team_abbreviation_home,team_abbreviation_away,Win,Loss,Date
```

gdzie Win i Loss oznaczają kurs bukmachera na wygraną i przegraną drużyny grającej u siebie.

## 6 Działanie aplikacji

Modelu można używać w formie aplikacji okienkowej. Program ten działa w bardzo prosty sposób: wybieramy drużynę, która gra u siebie oraz drugą którą gra na wyjeździe. Następnie wybieramy datę w której odbędzie się mecz. Po użyciu przycisku, program wygeneruje przewidywane kursy na zwycięstwo drużyn. Istnieje również opcja zobaczenia przewidywań dotyczących meczów, które odbywają się w niedalekiej przyszłości. Mecze te pobierane są za pomocą wspomnianego wcześniej skryptu.

## 7 Wyniki pracy

W celu zweryfikowania trafności przewidywań modelu przygotowano symulację działania napisanego programu. W tym celu wybrano 10 gier ligi NBA z ostatniego miesiąca i dokonano analizy porównawczej przewidywań modelu oraz bukmacherów. W większości przypadków przewidywania naszego modelu były zgodne z przewidywaniami modeli bukmacherów, co jest zadowalającym rezultatem. Aby sprawdzić możliwość potencjalnego zarobku na zakładach bukmacherskim przy pomocy naszego projektu przetestowaliśmy również trzy strategie stawiania zakładów

- W pierwszej strategii obstawiliśmy każdy z rozpatrywanych meczów zgodnie z przewidywaniami modelu, a wysokość zakładu zależna była od procentowej szansy na zwycięstwo. W związku z podobnymi estymacjami wyniku dla obu modeli w tym przypadku udało się zebrać kilka niewielkich wygranych i w ostatecznym rozrachunku wyjść na plus.
- Druga strategia zakładała wykorzystanie wiedzy o kursach bukmacherskich tak, aby obstawiać tylko te mecze, dla których przewidywania modeli różnią się, co przełożyłoby się na potencjalny duży zysk. W tym wariancie obstawiono tylko trzy mecze (z czego dwa poprawnie). Ostatecznie model wyszedł nieznacznie na plus.
- Trzecim pomysłem było zwiększenie ryzyka gry z poprzedniego wariantu poprzez wprowadzanie obstawiania drużyny przegranej w meczach, w której dajemy jej więcej szans na zwycięstwo niż bukmacherzy. W tym wariancie padło więcej źle obstawionych meczy, ale również większe wygrane w przypadku poprawnych predykcji - np. wygrana przy kursie 3.65 obstawiając zwycięstwo słabszej drużyny. W ostatecznym rozliczeniu model korzystający z tej strategii przyniósł straty. Niemniej jednak analiza tego wariantu przyniosła pomysły na ulepszenie techniki obstawiania na przykład poprzez wprowadzenie minimalnego progu dla różnicy estymowanych szans tak, aby nie obstawiać meczów z niewielką szansą pokonania predykcji bukmacherów.