UNIWERSYTET RZESZOWSKI

Kierunek: informatyka

Specjalność: Aplikacje Internetowe

**Serwis web gromadzący, weryfikujący i analizujący statystyki piłkarskie**

Jakub Fałowski  
nr albumu: 113605

Praca inżynierska

napisana pod kierunkiem

prof. dr inż. Marcina Ochaba

Rzeszów 2023

Spis treści

**1 Wstęp 4**

* 1. Technologie użyte w aplikacji i ich opis 4
  2. Użyte API / scrappowane serwisy internetowe i ich opis 6
  3. Cel i założenia projektu 8

1. **Moduł zawodników 9**
   1. Implementacja 9
   2. Zasada działania 11
   3. Możliwość rozwinięcia 17
   4. Przykładowe widoki 17
   5. (diagramy)
2. **Moduł klubów 17**
   1. Implementacja 17
   2. Zasada działania 20
   3. Możliwość rozwinięcia 31
   4. Przykładowe widoki 31
   5. (diagramy)
3. **Testy aplikacji 32**
   1. Predykcje wyników a rzeczywistość 32
   2. Zależność pomiędzy statystykami Sofascore a danymi z gier piłkarskich 36
   3. Testy jednostkowe i manualne

**Podsumowanie i wnioski 38**

**Bibliografia / netografia 40**

**1 Wstęp**

**1.1 Technologie użyte w aplikacji i ich opis**

**React.js**

**Next.js**

**Cheerio.js**

**Axios**

**NX**

**Mantine**

**React Router**

**Recharts**

PRZYSZŁE DEFINICJE Z KSIĄZEK

**1.2 Użyte API/serwisy**

**API Flashscore**

FlashScore API to interfejs programowania aplikacji (API), który umożliwia dostęp do danych dotyczących wyników meczów i statystyk sportowych ze strony flashscore.com. API umożliwia przetwarzanie danych w czasie rzeczywistym i wyświetlanie wyników meczów i statystyk na różnych platformach, takich jak strony internetowe, aplikacje mobilne i inne. Zapewnia dostęp do danych z wielu dyscyplin sportowych, w tym piłki nożnej, koszykówki, hokeja, tenisa i wielu innych. API umożliwia również wyszukiwanie i filtrowanie danych według różnych kryteriów, takich jak drużyna, lig, a nawet indywidualni zawodnicy. Jest często używany przez firmy sportowe, portale internetowe i aplikacje mobilne, które wymagają dostępu do aktualnych i wiarygodnych danych dotyczących wyników meczów i statystyk sportowych. Jego zalety to:

- Dostęp do danych w czasie rzeczywistym - FlashScore API umożliwia dostęp do danych w czasie rzeczywistym, co oznacza, że wyniki meczów i statystyki są aktualizowane na bieżąco.

- Dostęp do danych z wielu dyscyplin - FlashScore API umożliwia dostęp do danych z wielu dyscyplin sportowych, w tym piłki nożnej, koszykówki, hokeja i wielu innych.

- Możliwość wyszukiwania i filtrowania danych

API zostało użyte by pobierać terminarz, wyniki oraz kursy bukmacherskie na dane spotkanie w module przewidywania wyników.

**Forebet.com**

Forebet.com to strona internetowa, która oferuje prognozy i analizy dotyczące wyników meczów piłkarskich z całego świata. Strona zawiera informacje o meczach, drużynach, ligach i zawodnikach, a także statystyki i inne dane, które pomagają w przewidywaniu wyników meczów. Oferuje również system typowania wyników meczów, w którym użytkownicy mogą obstawiać wyniki meczów i rywalizować z innymi graczami.

Forebet.com jest popularną stroną internetową wśród miłośników piłki nożnej, którzy szukają informacji i analiz dotyczących wyników meczów.

Dane webscrappowane na bieżąco z powyższej strony są trzecim źródłem po mojej aplikacji i kursach bukmacherskich do przewidywania spotkań meczów.

**Fifa – futhead.com**

FIFA to seria gier piłkarskich wydawana przez firmę Electronic Arts (EA). Gra została pierwotnie wydana w 1993 roku i od tego czasu wydano wiele kolejnych wersji z ulepszeniami i dodatkami. Gra pozwala na rozgrywanie meczów piłkarskich z różnymi drużynami i ligami z całego świata. FIFA oferuje również tryb kariery, w którym gracze mogą zarządzać drużyną i rozwijać ją w trakcie wielu sezonów. FIFA jest jedną z najpopularniejszych gier piłkarskich na świecie i cieszy się dużą popularnością wśród miłośników tego sportu. Gra jest regularnie aktualizowana i ulepszana przez twórców, aby zapewnić jak najlepszą rozgrywkę i realizm.

Dane dotyczące statystyk piłkarzy, które pojawiają się w grze pobrałem za pomocą webscrappingu z portalu futhead.com. Zostały użyte jako jedno ze źródeł modułu porównywania statystyk piłkarzy

**Football Manager – fmdataba.com**

Football Manager to seria gier menedżerskich, w których gracze mogą zarządzać drużyną piłkarską i prowadzić ją przez kolejne sezony. Gra została wydana pierwotnie w 1992 roku i od tego czasu wydano wiele kolejnych wersji z ulepszeniami i dodatkami. Gra pozwala na rozgrywanie meczów piłkarskich i prowadzenie drużyny w trakcie sezonu. Gracze mogą wybierać spośród setek licencjonowanych drużyn i lig z całego świata, a także zarządzać składem drużyny, taktyką i budżetem. Football Manager oferuje również szczegółowy system transferów, w którym gracze mogą negocjować kontrakty z piłkarzami.

W mojej aplikacji atrybuty zawodników z Football Managera zostały użyte jako źródło porównawcze ze statystykami z Fify i API Sofascore

**API Sofascore**

Sofascore to internetowa platforma, która oferuje aktualne wyniki sportowe oraz statystyki z wielu różnych dyscyplin. Platforma ta prócz informacji o rezultatach, udostępnia szczegółowe statystyki dotyczące poszczególnych zawodników i drużyn. Sofascore to idealne narzędzie dla miłośników sportu, którzy chcą być na bieżąco z rozgrywkami sportowymi.

Z API Sofascore można pobrać tak szczegółowe dane jak np. liczba odbiorów, wślizgów, procent celnych podań i strzałów. Te statystyki idealnie nadają się do porównania względem atrybutów z gier.

**1.3 Cel i założenia projektu**

Projekt został stworzony do ułatwienia analizy danych piłkarskich. W internecie jest wiele informacji związanych z piłką nożną, lecz często są one sprzeczne. Aby móc je zweryfikować, porównać i zauważyć ewentualne zależności stworzyłem huba, gromadzącego wszystkie dane w aplikacji webowej. Na stronie znajdują się takie funkcjonalności jak:

- porównywanie statystyk zawodników w oparciu o trzy źródła

- wykres porównania średnich wartości atrybutów w różnych źródłach

- wykres porównania najlepszych klubów w oparciu o 11 najlepszych zawodników na podstawie wartości atrybutów zawodników

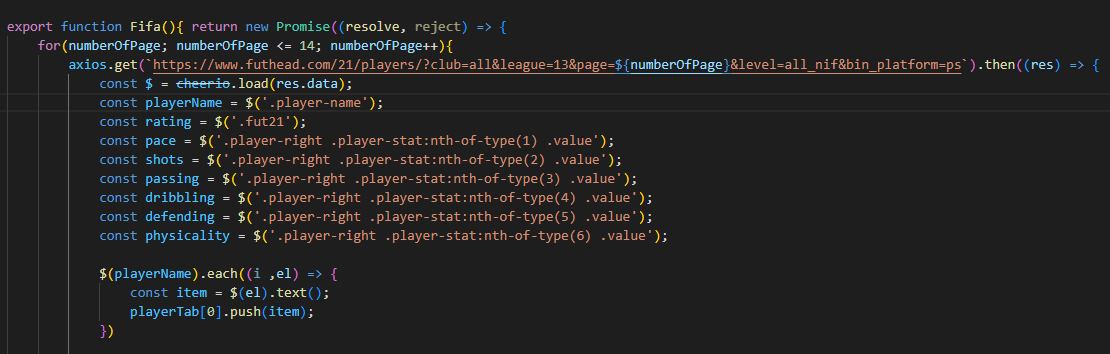
- własny system przewidywania meczy na podstawie ostatnich spotkań

- porównania różnych źródeł przewidujących najbardziej prawdopodobny wynik meczu

Za pomocą porównań atrybutów zawodników z trzech źródeł, można obiektywniej spojrzeć na umiejętności danego zawodnika. Przy okazji możliwa jest weryfikacja ewentualnej kradzieży danych bądź zależności pomiędzy wartościami opracowanymi przez researcherów dwóch gier piłkarskich. Dodatkowo trzecie źródło jakim są statystyki z sezonu na który te gry zostały przygotowane, weryfikują słuszność obydwu danych. Podobnie w funkcjonalności porównywania aplikacji do przewidywania wyników meczów. Trzy źródła mogą dawać trzy inne wyniki, lecz rzeczywistość po czasie weryfikuje, które z nich było najbliżej prawdy. Dodatkowo wiedząc jak działa stworzona przeze mnie aplikacja do predykcji, mogę obserwować wspólne zależności powodujące, że dana drużyna ma większe szanse na wygraną. Może mi to pomóc zrozumieć jak dwa inne źródła, z których pobieram dane o ich przewidywaniach, działają.

**2 Moduł zawodników**

**2.1 Implementacja**

****

Webscrapping statystyk z Fify za pomocą Cheerio i strony futhead

Obraz zawierający tekst

Opis wygenerowany automatycznie

Wyświetlanie powyższych statystyk w formie jsona na serwerze.

Obraz zawierający tekst

Opis wygenerowany automatycznie

Zdarzyły się przypadki, gdy ten sam zawodnik wystąpił dwa razy w danych pobieranych podczas scrappowania. By uniknąć tego typu błędów na stronie wrzucam do osobnej tablicy takich zawodników, aby ich kopia nie ukazywała się po złączeniu dwóch tablic

Obraz zawierający tekst

Opis wygenerowany automatycznie

Złączone zostają obydwie tablice w jedną, tak by mieć łatwiejszy dostęp do atrybutów zawodników, którzy występują w jednej i drugiej grze piłkarskiej.



Pobieranie statystyk zawodnika po jego nazwie za pomocą Sofascore API

**2.2 Zasada działania**

Jestem fanem piłkarskim, i grając od czasu do czas w gry o takiej tematyce zauważyłem, że niektórzy zawodnicy mimo świetnej dyspozycji i dobrych wyników indywidualnych nie są wystarczająco docenieni i posiadają nieadekwatne atrybuty do ich umiejętności. Zdarzało się też tak, że korzystałem z jednego z zawodników w Football Managerze, spisywał się na tyle dobrze, że miałem z nim pozytywne wspomnienia. Z tego względu zależało mi na tym by zagrać nim także w drugiej grze. Niestety różnica w jakości była duża. Teoretycznie każdy piłkarz ma swoje mocne i słabe strony, można go szczegółowo ocenić na podstawie jego dyspozycji w klubie. Jednak niestety różnice w ocenie atrybutów sprawiły, że w jednej z gier zawodnik ten nie był już tak dobry, w porównaniu do innych zawodników. Wiadomą sprawą jest to, że takie rzeczy jak dyspozycja zawodnika podczas mojej rozgrywki może zależeć od kilku czynników. Przykładowo szybkość jest szczególnie ważna dla zawodników ofensywnych, jeśli drużyna gra z kontrataku, a technika bardziej sprawdza się podczas wolnego rozgrywania akcji. Jednak te różnice wynikały czasami z samego oceniania atrybutów. Postanowiłem więc zrobić coś czego nie znalazłem w sieci, czyli aplikacje z możliwością porównywania atrybutów z obu gier oraz prawdziwej dyspozycji na podstawie statystyk ze sprawdzanego sezonu. Takie informacje jak umiejętności można odczytać z samej gry, bądź przeznaczonych do tego stron internetowych. Lecz te witryny są stworzone typowo pod graczy jednej z gier i zazwyczaj skupiają się na różne pomoce i wskazówki. Przykładowo strona futhead stworzona dla graczy Fify, skupia się i słynie z takich funkcjonalności jak używając kart zawodników wykonać dane wyzwanie, po których wykonaniu zdobywa się wirtualne pieniądze bądź paczki. Konkurentami dla witryny futheadu są takie portale jak futbin czy futwiz. W Football Managerze istnieje witryna fmdataba. Niestety nie istnieje żadne łatwodostępne API, dzięki któremu łatwo można by było te dane pobrać, więc aby móc stworzyć taką aplikacje wartości atrybutów zostały zwebscrappowane z portali, które posiadają takie informacje.

Obraz zawierający tekst

Opis wygenerowany automatycznie

Strona futhead, z której pozyskuje informacje dotyczące atrybutów z Fify

Jak widać na screenie powyżej obok nazwiska widnieją takie informacje jak:

- klub, w którym obecnie się znajduje (np. Manchester City),

- wartość oceny ogólnej(RAT- rating),

- szybkości (PAC – pace),

- strzałów (SHO – shotting),

- podań (PAS – passing),

- dryblingu (DRI – dribbling),

- obrony (DEF – defensive),

- fizyczności (PHY – physical)

Po pewnym czasie zauważyłem, że w tym źródle wśród zawodników nie ma bramkarzy. Nie zmienia to w żaden sposób sensu porównań, jedynie lekko zmniejsza szeroki już zakres bazy danych. Stwierdziłem, że dobrym rozwiązaniem będzie ograniczenie piłkarzy do jednej ligi, ponieważ wystarczy to do wyciągnięcia odpowiednich wniosków, a zwiększy miarodajność porównań. Niektóre ligi są silniejsze, a niektóre stoją na niższym poziomie, przez co strzelenie 20 bramek w polskiej lidze różni się od strzelenia 20 bramek w lidze angielskiej. Z tego względu wybrałem angielską Premier League, ponieważ według współczynniku UEFA (organizacja piłkarska, która organizuje największe europejskie turnieje takie jak Liga Mistrzów) jest ona najsilniejsza. Wyprzedza między innymi ligę hiszpańską, niemiecką, włoską czy francuską.

Obraz zawierający stół

Opis wygenerowany automatycznie

Współczynnik UEFA, na podstawie danych z portalu transfermarkt

Scrapper pozyskał informacje z wszystkich podstron i zapisał je do tablicy, po czym przekonwertowałem je do obiektów i wrzuciłem w formie JSONa do API.

Obraz zawierający tekst

Opis wygenerowany automatycznie

Jest to dosyć czytelne rozwiązanie, ponieważ każdy obiekt ma wartość name z imieniem i nazwiskiem piłkarza, rating z oceną ogólna oraz inne atrybuty. Jest to również łatwe do obsługi podczas wyświetlania na stronie.

Zdarzały się pojedyncze przypadki, gdy nazwa piłkarza była tak długa, że na powyższej stronie nazwisko było niepełne i wypełnione trzema kropkami jak na screenie poniżej. Wynikało to po prostu z takich danych w ich bazie.



W takich przypadkach musiałem ręcznie podmieniać wartość przed wrzuceniem obiektu tego piłkarza do API



Musiałem też ustawić zezwolenie na CORSy po stronie serwera, ponieważ było to wymagane to prawidłowego wyświetlania informacji na stronie

Drugim źródłem informacji była fmdataba. Tutaj ostatnie dane jakie są dostępne pochodzą z 21 edycji Football Managera. Istnieją inne strony o takiej tematyce, aczkolwiek nie ma tam tak szczegółowych danych, których chciałbym użyć do porównań. Skutkiem tego była decyzja, by atrybuty zarówno z Fify, Football Managera oraz statystyki z rozgrywek pochodziły z 2021 roku. Dodatkowo chciałem by dane z rzeczywistej postawy były niejaką weryfikacją wcześniejszych danych, czy rzeczywiście ten piłkarz zasługiwał na taką a nie inną ocenę. Potrzebna była więc edycja wcześniejsza, tak by statystyki poszczególnych zawodników mogły być porównywane na przekroju przyszłego całego sezonu.

Obraz zawierający stół

Opis wygenerowany automatycznie

Na portalu do Football Managera podobnie jak w przypadku futheada, na stronie widnieje widok tabeli z nazwiskiem i atrybutami piłkarzy. Znajdują się tam zawarte informacje o:

- klubie (Club),

- ocenie ogólnej(Ove – overall), statystyka taka sama jak w Fifie,

- ataku(Att – attacking), odpowiednik strzałów w Fifie,

- defensywie(Def – defending), odpowiednik w Fifie,

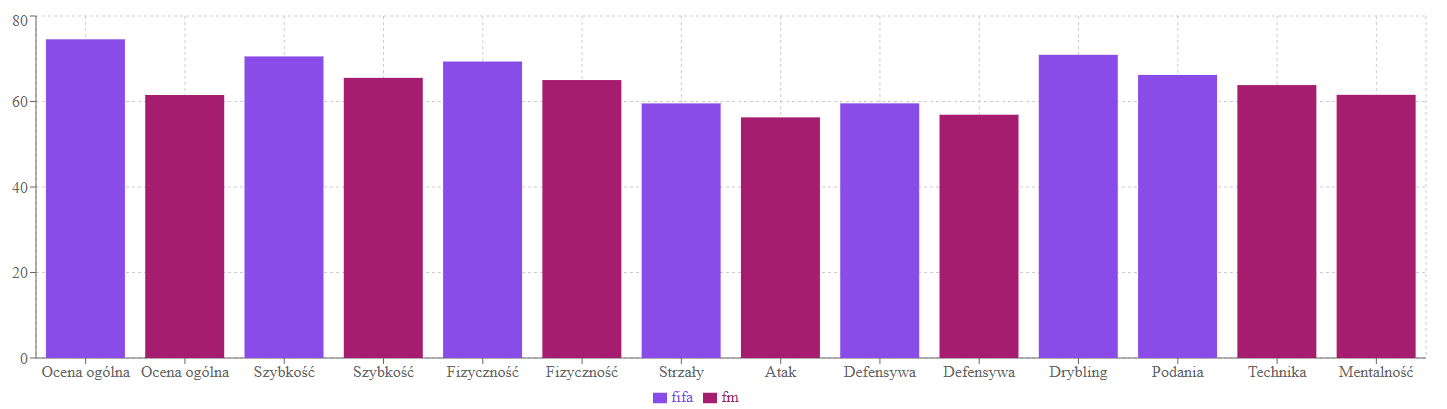
- technice(Tec – technical), składowa dryblingu i podań w Fifie,

- mentalności(Men – mental), nie ma tego typu statystyki w Fifie. Wynika to z tego faktu, że w Football Managerze gracz zostaje trenerem piłkarskim i zarządza swoimi zawodnikami, którzy mogą podejmować różne decyzje ze względu na swój charakter czy inteligencje boiskową. W Fifie to gracz podejmuje decyzje, ponieważ on bezpośrednio decyduje o zachowaniach boiskowych za pomocą kontrolera, a nie tylko i wyłącznie taktyka przez niego przydzielona,

- fizyczności(Phy – physical), odpowiednik w Fifie,

- szybkości(Spe – speed), odpowiednik w Fifie

Zakres atrybutów w Football Managerze jest inny niż w Fifie. W symulatorze trenera wynosi on 1-20 a nie 1-100. Nie jest to jednak problemem, ponieważ podczas porównywania, w której grze zawodnik jest lepszy zwracam uwagę na to który jest w rankingu w stosunku do innych zawodników. Przykładowo Raheem Sterling zajmuje 1 miejsce pod względem szybkości w Football Managerze, a 4 w Fifie. Można w ten sposób wysunąć wnioski, że pod tym względem w pierwszej grze został bardziej doceniony. Zakładam również, że 18 w FMie to 90/100 w Fifie, ze względu na to że to po prostu 90% względem maksymalnego atrybutu.



Jedynym problemem w tym założeniu mogłoby być to, że patrząc tym sposobem średnia ocena ogólna w Fifie to 74.55, a w Football Managerze 61,52. Przez to wartości w Fifie są zazwyczaj wyższe niż w FMie. Jednak jest to kwestia indywidualnej opinii, czy ważniejsze jest to by zawodnik był jak najlepszy względem maksymalnej wartości czy lepszym względem innych piłkarzy. Dodatkowo jednym z kryteriów przy porównywaniu jest procent w stosunku do najlepszej oceny tego konkretnego atrybutu.

Gdy już mamy dane zebrane z jednego i drugiego źródła to można je porównywać. Liczba ich różni się, z różnych względów. Przede wszystkim zespoły mają szeroką kadrę, która składa się nie tylko z zawodników grających na co dzień w pierwszym zespole. Kluby posiadają również drugą drużynę i juniorów, którzy w razie kontuzji, zawieszenia, bądź po prostu woli trenera uzupełniają skład. Przede wszystkim Football Manager zwraca dużą uwagę również na tych młodszych zawodników. Dla wielu graczy obserwowanie rozwoju nastoletniego zawodnika w dłuższym okresie czasu potrafi być satysfakcjonujące. Zwiększanie się umiejętności po zdobytym doświadczeniu działa bardzo dobrze w tej grze. Aplikacja została stworzona w taki sposób, że dany zawodnik jest wyświetlany w niej, tylko w momencie gdy pojawi się w obydwu grach. Potencjalni odbiorcy witryny zainteresowani będą najsławniejszymi zawodnikami, lub chociaż graczami z pierwszej drużyny swojego klubu. Jeśli nie ma kogoś w obydwu tytułach to oznacza prawdopodobnie, że jest nieznaczący w kontekście analizowanych rozgrywek. Dodatkowo do analizy dołączają się dane statystyczne z rozgrywek pozyskane z Sofascore. Te dane pokazane są w sposób bardziej poglądowy i stworzone do własnej analizy. Niektóre statystyki takie jak średnia ocena powinny być zawsze miarodajne i dobrze odzwierciedlać formę piłkarza względem oceny ogólnej z gier. Jednak niektóre atrybuty takie jak „ilość faulów na danym zawodniku”, które zostały sklasyfikowane przeze mnie w kolumnie szybkość są sporo większe w przypadku zawodników atakujących. Sofascore co prawda posiada bardzo dużo cennych informacji, które mogą służyć do analityki, aczkolwiek posiada wady, przez które ciężkie byłoby stworzenie jednoznacznego rankingu pod względem całościowych umiejętności w obronie czy ataku. Wartości poszczególnych danych można porównać procentowo w stosunku do najlepszej oceny, co posiadają również wartości z Fify i Football Managera.

OCENA OGÓLNA:

- średnia ocena – domyślna ocena zawodnika przed meczem wynosi 6.6. W momencie gdy zawodnik strzeli bramkę, wykona dokładne podanie itd. otrzymuje dodatnie punkty do początkowej średniej. To samo się tyczy ujemnych punktów w przypadku złych zagrań. Oczywiście po strzelonej bramce zawodnik otrzymuje większy bonus punktowy niż po dokładnym zagraniu. Te dane odnośnie mnożnika nie są szeroko dostępne i są oceniane subiektywne przez człowieka odpowiedzialnego od tego w serwisie SofaScore. Portal ten jest jednym z popularniejszych źródeł piłkarskich, a najlepsza 11 kolejki ligowej, która jest wystawiana na podstawie średniej oceny, jest szeroko rozpowszechniana w mediach społecznościowych

SZYBKOŚĆ:

- minięcia zawodników na mecz – ta statystyka mówi ile razy zawodnik wyprzedził swojego przeciwnika z piłką przy nodze. Najlepsze wartości osiągają ofensywni szybcy zawodnicy. Umiejętność szybkiego biegania, które przekłada się na wymijanie przeciwników jest bardzo ważną zaletą, ponieważ pomaga drużynie w szybkim przedostaniu się pod bramkę przeciwnika, bądź znalezienia wolnej przestrzeni, która pozwoli na skuteczne dośrodkowanie piłki w pole karne przeciwnika.

- ilość faulów na danym zawodniku – ten parametr wiąże się w dużym stopniu z powyższym, ponieważ do nieczystego zagrania dochodzi zazwyczaj w momencie, gdy zawodnik wymknie się przeciwnikom spod kontroli i nie mogą oni za nim nadążyć. Większość przewinień to faule taktyczne, czyli takie w których drużynie faulującej opłaca się by piłka została zatrzymana na rzucie wolnym, a zawodnicy broniącej drużyny zdążą wrócić się na swoje pozycje, bądź takie w których wolniejszy zawodnik chce odebrać piłkę, ale ze względu na jego spóźnienie, jego noga trafia w nogi przeciwnika.

FIZYCZNOŚĆ:

- wygrane pojedynki powietrzne – największy wpływ na to czy zawodnik wygra pozycje w momencie, gdy piłka jest w powietrzu ma wzrost, skoczność i siła. Są to atrybuty, które jak i reszta wartości opisywanych statystyk, składają się na podgrupę z danej kategorii w obydwu grach.

- wygrane pojedynki z piłką na trawie – siła i zwinność to wartości, które mają duże znaczenie podczas pojedynków odbywających się na niskiej wysokości.

ATAK:

- liczba bramek – informacje o ilości zdobytych bramek przez tego zawodnika w przekroju całego sezonu

- liczba strzałów – ilość bezpośrednich prób zdobycia bramek

DEFENSYWA:

- odbiory – ilość sytuacji w których dany zawodnik przejmie piłkę przeciwnikowi i bezpośrednio po tym się z nią utrzyma

- wślizgi – statystyka mówiąca o tym ile razy zawodnik próbuje odebrać piłkę za pomocą ruchu ślizgowego po ziemi

- wybicia - ilość sytuacji w których dany zawodnik przeszkodził przeciwnikowi w utrzymaniu się przy piłce, lecz po manewrze nie znajduję się przy niej

TECHNIKA:

- asysty – liczba podań po których kolega z drużyny zdobędzie bramkę

- kluczowe podania – ilość podań które pośrednio prowadzi do niebezpiecznej sytuacji

- stworzone niebezpieczne sytuacje – ilość ostatnich podań przed niebezpieczną sytuacją

- procent celnych podań – rozkład procentowy ilości podań do kolegi z drużyny w stosunku do wszystkich podań

Dodatkowo prócz stworzonej ogólnej tabelki porównawczej pomiędzy zawodnikami, istnieją wykresy w których można uzyskać bardziej zaawansowane informacje o atrybutach. W jednym z nich można porównywać średnią statystykę danego atrybutu pomiędzy Fifą, a Football Managerem. W pozostałych widnieje ranking według najlepszych klubów, gdzie siła drużyny równa się sile najlepszych 11 zawodników z danego klubu.

**2.3 Możliwość rozwinięcia**

Jestem zadowolony z efektów jakie jest w stanie zdziałać strona. Jednak dostrzegam ewentualny krok, który mógłbym wykonać jeśli chciałbym rozszerzyć możliwości w razie nadmiaru czasu. Przede wszystkim oceny z gier piłkarskich są poddawane statystykom jedynie z przyszłego roku. Mógłbym pozyskiwać informacje z API także z poprzednich paru lat by móc szczegółowiej określić w jaki sposób zawodnik doszedł do takiej oceny. Możliwa też by była dodatkowa statystyka pokazująca, który zawodnik najbardziej zasłużył na progres ocen względem roku poprzedniego.

**2.4 Poszczególne widoki**

Screeny strony internetowej

**3 Moduł klubów**

**3.1 Implementacja**

**Obraz zawierający tekst

Opis wygenerowany automatycznie**

Funkcja pobierająca z API ostatnie mecze dwóch drużyn, których wynik meczu będzie przewidywany. Zmienna tablicowa ekstraklasaTab została stworzona z powodu wadliwości API. Z założenia w jsonie każdy numer tablicy był przydzielony do osobnych rozgrywek - ligi krajowej, pucharu, lig europejskich itd. Jednak z nieznanych przyczyn niektóre mecze ligowe były w osobnej tablicy. Wszystkie miały taką samą nazwę w atrybucie NAME. Rozwiązaniem okazało się zsumowanie wartości wszystkich tablic o takiej nazwie.

Obraz zawierający tekst

Opis wygenerowany automatycznieFunkcja pobierająca z API wysokość kursów bukmacherskich

Obraz zawierający tekst, wewnątrz, ekran, zrzut ekranu

Opis wygenerowany automatycznie

Wyświetlanie danych z forebetu

Obraz zawierający tekst

Opis wygenerowany automatycznie

Obraz zawierający tekst

Opis wygenerowany automatycznie

W API data przesyłana jest jako liczba, 1654293600 to 4 czerwca 2022 roku - północ. Zauważyłem to na podstawie jednego z meczów. Ta liczba to czas w sekundach od mniej więcej 1650 roku. Jednak zależało mi na wyświetlaniu czasu rozgrywania meczów w formie daty. Aby pozyskać prawidłowy czas wyznaczam różnice liczby podanej w API od znanej wyżej daty i konwertuje tą liczbę na datę rozgrywania spotkania.

**3.2 Zasada działania**

Wynik meczu jest najważniejszym elementem w piłce nożnej. Każdy kibic lubi jak jego drużyna gra efektownie i prezentuje ofensywny styl, aczkolwiek to zwycięstwa gwarantują lepszą przyszłość. Dzięki dobrym wynikom można sobie zagwarantować awans do wyższej ligi co przekłada się na konkurowanie z lepszymi drużynami. Lepsze drużyny to mecze na większym poziomie, więcej fanów na trybunach bądź większe wpływy z telewizji, które można wydać na jeszcze większy rozwój klubu. Dlatego też predykcja tychże wyników jest bardzo popularna co wykorzystują firmy bukmacherskie. To źródło w teorii powinno być najbardziej rzetelne. Zakłady bukmacherskie zarabiają duże pieniądze na swoich ofertach i muszą inwestować również spore środki, by ten system przynosił zyski. Zauważyłem, że często na przewidywania przed meczem ma wpływ nie tylko dyspozycja i siła lecz także reputacja danej drużyny. Moja opinia jest taka, że ten argument nie ma zbyt dużego znaczenia i jest to „dziura”, którą można wykorzystać w typowaniu. Jako przykład mogę wymienić mecz Legii Warszawa z Radomiakiem Radom w 19 kolejce sezonu 2021/2022. Drużyna z Warszawy do momentu rozegrania meczu zdobyła 18 punktów w 18 meczach i plasowała się na 15 pozycji, a forma w jej ostatnich meczach nie zachwycała. W opozycji do najbardziej znanego klubu w Polsce, drużyna Radomiaka, dla której był to pierwszy sezon w najwyższej klasie rozgrywkowej. Mimo małego doświadczenia zajmowali oni 4 miejsce z 32 punktami na koncie. Firmy bukmacherskie oferowały 2 złote i 20 groszy za każdą postawioną złotówkę na Legie, a na Radomiaka 3.55 złotych. Oznacza, że dosyć dużym faworytem była Legia. Mecz ostatecznie zakończył się wynikiem 3-0 dla zespołu z Radomia. Aby zbadać moją hipotezę i sprawdzić jak to działa w praktyce, na podstawie wcześniejszych wyników, statystyk oraz opinii ekspertów postanowiłem zrobić własną aplikacje do predykcji. Aby przewidywania były jak najbardziej zbliżone do prawidłowych postanowiłem skupić się na jednej konkretnej lidze. Statystyki pokazują jasno, że takie czynniki jak ilość remisów czy bramek ,są zależne od kraju i poziomu rozgrywkowego. Na ilość remisów wpływa w dużej mierze ilość bramek. Jeżeli wartość ta znajduję się w zakresie 0-2 to 2 z możliwych 6 wyników oznacza podział punktów. W przypadku 3-5 goli, zaledwie 1 z 15 rezultatów jest remisowych.

0: **0-0**

1: 1-0, 0-1

2: **1-1**, 0-2, 2-0

3- 3:0, 0:3, 2:1, 1:2

4– 4:0, 0:4, 3:1, 1:3**, 2:2**

5 – 5:0, 0:5, 4:1, 1:4, 3:2, 2:3

W niektórych państwach panuje filozofia ofensywna i piłkarzom zależy na dużej zdobyczy bramkowej, gdyż to wpływa na atrakcyjność meczy i przyciąga więcej kibiców. W innych częściej gra się bardziej zachowawczo, ponieważ to często pomaga w uzyskiwaniu dobrych wyników, co się przekłada na miejsce w tabeli. W dodatku niektóre narodowości słyną z tego, że są nadprzeciętne dobre w danej dziedzinie. Przykładowo zawodnik z Brazylii kojarzy się z dobrym dryblingiem, szybkością i wieloma walorami ofensywnymi, a włoskie kluby z przykuwaniem dużej uwagi do taktyki.

Obraz zawierający stół

Opis wygenerowany automatycznie

źródło: <https://www.eprogresja.pl/statystyki/pilkarskie/remisy_w_ligach/?fbclid=IwAR1ozhjSmX2-S89VQwz7uBJOgr0GpcpU-tAlZeSc_muXt_MQJ0h_VYHLae0>

Według danych portalu eprogresja.pl z dnia 2 grudnia 2022 roku najczęściej remis pada w bahrańskiej Premier League, a częstotliwość takiego wyniku wynosi 40%. Po drugiej stronie Indyjska Super League, gdzie ten współczynnik jest równy 9.52%. Jak widać różnica zależna od ligi jest kolosalna.

Wartości najbardziej popularnych lig:

- angielska Premier League: 23.29%

- hiszpańska La Liga: 23.57%

- niemiecka Bundesliga 22.96%

- francuska Ligue 1: 26.67%

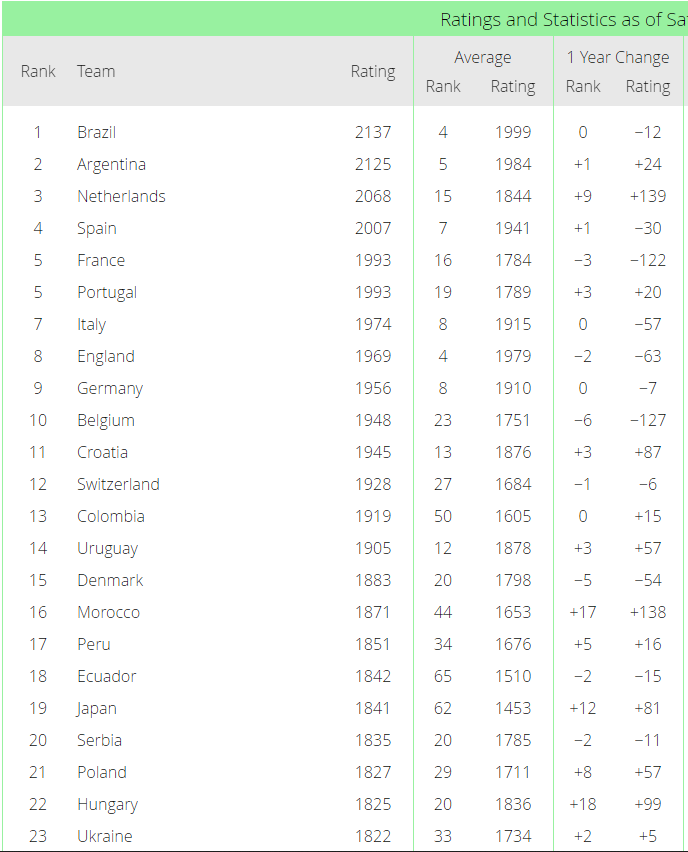
- włoska Serie A: 24%

- polska Ekstraklasa: 23.03%

- polska 1 liga: 31.48%

Zdecydowałem się na analizowanie polskiej Ekstraklasy, ze względu na sympatie i największą wiedzę dotyczących źródeł spośród tych rozgrywek.

Pierwszym zadaniem aplikacji jest pobranie informacji dotyczących ostatnich meczy w Ekstraklasie. Aby je uzyskać używam API Flashscore. Za jego pomocą tworze ocenę siły drużyn, w podobny sposób jak ranking ELO. Ranking ten jest stosowany do określenia mocy reprezentacji narodowych w piłce nożnej. W przypadku piłki kobiecej nawet ranking FIFA, czyli oficjalnej federacji zarządzającej mistrzostwami świata, opiera się o to kryterium. W przypadku reprezentacji narodowych stosuje się go na podstawie 30 lub więcej spotkań. Taka pula meczy wynika z dużej liczby narodowości mogących grać między sobą i możliwości gry przez dłuższy okres czasu wyłącznie ze słabszymi zespołami.



Ranking ELO męskich drużyn narodowych na dzień 3 grudnia 2022 roku według portalu eloratings.net

Podczas rozgrywek ligowych w trakcie jednego sezonu trwającego około rok czasu, drużyny mierzą się z wszystkimi pozostałymi przeciwnikami, a ich drużyn wynosi zazwyczaj od 12 do 20. Często zdarza się, że dyspozycja danej drużyny zmienia się bardzo szybko i wyniki sprzed roku nie są adekwatne do obecnej sytuacji w klubie. Takim przykładem jest sytuacja Zawiszy Bydgoszcz w latach 2013-2015. Sezon 2013/2014 był by dla nich bardzo udany, zajęli oni 8 miejsce oraz wygrali Puchar Polski, dzięki czemu grali w kwalifikacjach do europejskich pucharów. Rok później zajęli oni przedostatnie 15 miejsce, przez co spadli na niższy poziom rozgrywkowy. Wpływ na tak gwałtowną zmianę pozycji mają w dużym stopniu roszady kadrowe. Dobre rezultaty zespołu wynikają nierzadko z nadprzeciętnej dyspozycji jednego czy dwóch piłkarzy, którzy w trakcie okienka transferowego przechodzą do innych, lepszych klubów. Stąd też moja teza o nacisk przy przewidywaniu wyników na bieżącą formę. Postanowiłem więc do punktacji brać uwagę przekrój ostatnich 15 meczy, ponieważ jest to reprezentatywna ilość danych, gdyż zespół mierzy się przez ten okres z praktycznie całą ligą. Rywalizuje zarówno z lepszymi i słabszymi przeciwnikami i nie ma mowy o tym, że dobra dyspozycja była spowodowana jedynie sprzyjającym terminarzem. Z drugiej strony mecz, który odbywał się rok czy dwa lata temu nie ma wpływu na przewidywanie wyniku z obecnego okresu i nie zakłóca obrazu podczas predykcji. Zmiana trenera i zawodników, która w polskiej lidze zdarza się często, może sprawić, że nawet po takim okresie jak rok wyniki są zupełnie odmienne w stosunku do dzisiejszych, zarówno pod względem zdobyczy punktowej, jak i bilansu bramkowego.

Obraz zawierający tekst

Opis wygenerowany automatycznie

Źródło: <https://pl.wikipedia.org/wiki/Ranking_Elo_(pi%C5%82ka_no%C5%BCna)>, szukanie alternatywnego źródła np. w książkach

W rankingu ELO indeksem ważności oznaczana jest ranga meczu. W przypadku spotkania podczas mistrzostw świata waga jest sporo większa niż w meczu towarzyskim. Jeśli chodzi o spotkania ligowe, każdy mecz jest równie ważny, ponieważ niezależnie od przeciwnika zdobycz punktowa w razie zwycięstwa jest taka sama. W tabeli, która weryfikuje siłę drużyn, liczy się suma wszystkich punktów i to od niej zależy, które drużyna zajmie miejsce w tabeli, czy się utrzyma w lidze, bądź awansuje. W mojej aplikacji chce zweryfikować ostatnią formę danej drużyny, a w ocenieniu tego pomagają ostatnie mecze, szczególnie te które się odbyły ostatnio. Na podstawie otrzymywanych wyników metodą prób i błędów, postanowiłem ustalić rangę za pomocą następującego liniowego wzoru.

K = 2 – (X/7.5), gdzie

K – indeks ważności meczu

X – numer meczu

Nie komplikowałem obliczeń z wstawianiem spodziewanego wyniku meczu do obliczania rankingu, ze względu na to, że w obszarze 15 kolejek kluby mierzą się z całym przekrojem ligi. Rywalizują zarówno z lepszymi drużynami, jak i tymi które w teorii powinny być na dnie tabeli. Dodałem za to taki parametr jak gra na własnym boisku. W analizie statycznej Ekstraklasy portalu 90minut.pl z ostatniego sezonu 2021/2022 można znaleźć informacje związane z grą u siebie.

*W 306 meczach zdobyto 818 goli, co daje średnią 2,673 goli na mecz. Gospodarze zdobyli 459 goli (56%), a goście 359 goli (44%) (...) Gospodarze wygrali 135 meczów (44%), padły 84 remisy (27%), a goście wygrali 87 meczów (28%).*

Stosunek zwycięstw gości do gospodarzy wynosi 64,(4)%. W lidze „każdy gra z każdym” zarówno na swoim jak i wyjazdowym terenie. Widać więc tu sporą zależność wyniku od miejsca w którym spotkanie jest rozgrywane. Około 1/3 wyników spotkań jest przechylane na rzecz gospodarzy ze względu na swoje boisko, kibiców i brak konieczności odbywania podróży. Jednak nie jest to jednoznaczne, często jest to kwestia indywidualna. W sezonie 2021/2022 aż 15 na 18 zespołów zdobyło więcej punktów u siebie względem meczów na wyjeździe. Niektóre zespoły wręcz były uzależnione od dobrych wyników na własnym boisku. Dobrym przykładem jest 40 zdobytych punktów Lechii Gdańsk u siebie i tylko 17 na obcym boisku. Podobnie Wisła z dorobkiem 35 punktów w Płocku i 13 oczek na wyjazdach. Zdarzają się jednak sytuacje, w których drużynom lepiej się gra na innym boisku, niż tym gdzie formalnie są gospodarzem. W takiej sytuacji jest Warta Poznań, która rozgrywa swoje spotkania domowe w Grodzisku Wielkopolskim, ze względu na brak własnego stadionu, spełniającego wymogi licencyjne na najwyższym poziomie rozgrywkowym. W sezonie 2021/2022 stosunek punktów domowych Warty do wyjazdowych wynosił 18:24. Po rundzie jesiennej kolejnego sezonu, czyli po rozegraniu połowy spotkań, ta różnica jest sporo większa i wynosi 6:19. Mając na uwadze powyższe statystyki postanowiłem, że prócz punktów mających określić formę w ostatnich 15 meczach do rankingu klubu będzie wliczała się też forma zespołu w miejscu gdzie będzie się odbywał mecz. Jeśli drużyna gra kolejny mecz u siebie, 1/3 punktów w rankingu będzie zależna od ostatnich 5 spotkań u siebie. Analogicznie w przypadku spotkań wyjazdowych. Ogólny wzór na obliczanie punktów w rankingu, który będzie miał zakres procentowy, czyli wartości od 0 do 100 jest następujący:

PP = ((2\*(PO/45)+(PT/15))/3)\*100, gdzie

PP – punkty procentowo,

PO – punkty ogólne, czyli liczba punktów w ostatnich 15 meczach pomnożona razy indeks ważność meczu k(wzór powyżej, uśredniona wartość równa 1), 45 to maksymalna wartość po,

PT – punkty terenowe, czyli liczba punktów w ostatnich 5 meczach u siebie, bądź na wyjeździe w zależności od tego gdzie dana drużyna gra, 15 to maksymalna wartość pt

Mamy więc informacje o ilości punktów drużyny w rankingu. Lecz celem aplikacji nie jest tylko i wyłącznie wygenerowanie siły drużyny, lecz wynik w spotkaniu z drużyną przeciwną. Aby próbować przewidywać jakim wynikiem zakończy się spotkanie potrzebne są informacje o tym która drużyna wygra spotkanie, jaka będzie różnica bramkowa i ile bramek padnie. W piłce nożnej dosyć często dochodzi do remisów i trzeba je też brać pod uwagę przy predykcjach. Przy obliczaniu tylu działań mało prawdopodobne jest, że siła obu drużyn będzie identyczna. W celu rozwiązania tego problemu komunikowałem się z mgr. Piotrem Klimkiem, który specjalizuje się w matematycznych obliczeniach i wykresach dotyczących piłki nożnej. Posiada on liczną grupę obserwujących na serwisie społecznościowym Twitter i jest uważany za eksperta w swojej dziedzinie. Dodatkowo biorę pod uwagę liczby z poprzedniego sezonu. Skoro w ostatnim sezonie liczba remisów wynosiła 23.03% to w obecnym stosunek powinien mieć w miarę podobną wartość i co czwarte bądź piąte przewidywanie powinno wskazywać na remis. Wzory, które po konsultacjach i analizie najlepiej pasowały do podziału sił drużyn A i B na procenty A, B i remisu wynoszą następująco:

Jeśli H >= 60:

HW = 50 + H-60

DW = 30 - (H-60)/2

AW = 20 - (H-60)/2

Jeśli H > 50 i H < 60:

HW = 30 + (H-50)\*2

DW = 40 - (H-50)

AW = 30 - (H-50)

Jeśli H <= 50 i H > 40:

HW = 30 - (A-50)

DW = 40 - (A-50)

AW = 30 + (A-50)\*2

Jeśli H <= 40:

HW = 20 – (A-60)/2

DW = 30 - (A-60)/2

AW = 50 + A-60, gdzie

H – procentowy stosunek siły drużyny gospodarzy do siły obu drużyn, H równe 100 - A

A – procentowy stosunek siły drużyny gości do siły obu drużyn, A jest równe 100 - H

HW – prawdopodobieństwo na zakończenie spotkania zwycięstwem gospodarzy

DW - prawdopodobieństwo na zakończenie spotkania remisem

AW - prawdopodobieństwo na zakończenie spotkania zwycięstwem gości

Przykładowe rozkłady:

55-45: 40-35-25,

59-41: 48-31-21,

70-30: 60-25-15

Mając parametry HW, DW i AW można określić potencjalnego zwycięzcę spotkania. Wiadomą sprawą jest, że czym większa wartość danej statystyki tym większe prawdopodobieństwo na danego zwycięzcę. Wynika to z różnicy siły pomiędzy drużynami. Skoro drużyna A jest sporo lepsza od drużyny B, to można przypuszczać, nie tylko że zdobędzie 3 punkty, ale również że różnica bramkowa będzie trochę większa. Z analizy statystycznej Ekstraklasy wynika, że

***Częstość wyników:*** *1-1 - 46 razy, 1-0 - 35 razy, 2-1 - 25 razy, 0-1 - 24 razy, 1-2 - 20 razy, 0-0 - 19 razy, 2-0 - 17 razy, 3-1 - 16 razy, 2-2 - 15 razy, 3-0 - 13 razy, 0-2 - 13 razy, 1-3 - 10 razy, 4-1 - 8 razy, 0-3 - 5 razy, 0-4 - 5 razy, 3-2 - 5 razy, 4-0 - 5 razy, 2-3 - 4 razy, 3-3 - 4 razy, 4-2 - 4 razy, 5-0 - 3 razy, 3-4 - 2 razy, 4-3 - 2 razy, 2-4 - 2 razy, 5-1 - 1 raz, 1-6 - 1 raz, 0-5 - 1 raz, 5-3 - 1 raz,*

***Najwięcej goli (8) padło w meczu:***[*Legia Warszawa 5-3 Górnik Zabrze*](http://www.90minut.pl/mecz.php?id_mecz=1677841)

***Najwyższe zwycięstwo gospodarzy padło w meczach:***[*Lech Poznań 5-0 Wisła Kraków*](http://www.90minut.pl/mecz.php?id_mecz=1677625)*,*[*Raków Częstochowa 5-0 Jagiellonia Białystok*](http://www.90minut.pl/mecz.php?id_mecz=1677575)*,*[*Lech Poznań 5-0 Bruk-Bet Termalica Nieciecza*](http://www.90minut.pl/mecz.php?id_mecz=1677743)

***Najwyższy remis padł w meczach:***[*Cracovia 3-3 Stal Mielec*](http://www.90minut.pl/mecz.php?id_mecz=1677642)*,*[*Jagiellonia Białystok 3-3 Piast Gliwice*](http://www.90minut.pl/mecz.php?id_mecz=1677672)*,*[*Stal Mielec 3-3 Lechia Gdańsk*](http://www.90minut.pl/mecz.php?id_mecz=1677693)*,*[*Cracovia 3-3 Lech Poznań*](http://www.90minut.pl/mecz.php?id_mecz=1677733)

***Najwyższe zwycięstwo gości padło w meczach:***[*Wisła Kraków 0-5 Śląsk Wrocław*](http://www.90minut.pl/mecz.php?id_mecz=1677667)*,*[*Radomiak Radom 1-6 Zagłębie Lubin*](http://www.90minut.pl/mecz.php?id_mecz=1677843)

Wynika z tego, że 222 mecze, które nie zakończyły się remisem zakończyły się następującą różnicą bramek:

- wygrane 1 bramką: 117 (52,7%)

- wygrane 2 bramkami: 63 (28,4%)

- wygrane 3 bramkami: 26 (11,7%)

- wygrane 4 bramkami: 11 (4,9%)

- wygrane 5 bramkami: 5 (2,3%)

Dopasowując to do trzytorowych procentów na dany wynik (HW – wygrana gospodarzy, DW – remis, AW – wygrana gości) najbardziej zbliżone wyniki wychodziły w momencie gdy zależności procentów do różnicy bramkowej były następujące:

WP < DW+10 – różnica 1 bramki

WP >= DW+10 i WP < DW+25 – różnica 2 bramek

WP >= DW+25 i WP < DW+40 – różnica 3 bramek

WP >= DW+40 – różnica 4 bramek, gdzie:

WP – procent na drużynę, która według trzytorowych przewidywań ma największe szanse na wygraną

DW – procent na remis w tym meczu

Następnie obliczam ilość strzelonych bramek przez oba zespoły w ostatnich 15 kolejkach. Suma wszystkich goli wpływa na stopień bramkostrzelności obu zespołów:

Stopień 1 – poniżej 2 goli

Stopień 2 – 2.0-2.33

Stopień 3 – 2.34-2.66

Stopień 4 – 2.67-2.99

Stopień 5 – powyżej lub równe 3 gole

W zależności od stopnia bramkostrzelności obu ekip i przewidywanej różnicy pomiędzy zespołami najbardziej prawdopodobny wynik jest podawany przez aplikacje.

Kolejnym źródłem są kursy od bukmacherów. Jeśli kurs wynosi 2.0 to za każdą postawioną złotówkę klient dostanie dodatkową złotówkę, czyli podwaja swoją kwotę. Z matematycznego punktu widzenia szansa na wygraną w takich sytuacjach wynosi 50%. W momencie gdy kurs wynosi 4.0, wskazuje to na szanse 25%. Wzór na obliczanie szans wynosi wtedy 1/(wysokość kursu). Z reguły suma szans na zwycięstwo gospodarzy, gości oraz remis wynosi 105-110%. Wynika to z tego, że większe prawdopodobieństwo według tej metody oznacza mniejszą możliwą wygraną dla klienta. A zakłady bukmacherskie są konstruowane w taki sposób, aby móc zarabiać na swoich klientach. Aby móc czytelniej porównywać dane powyższe procenty zostały podzielone w taki sposób by całość wynosiła 100%. Zakłady bukmacherskie w Polsce obejmuje 12% podatek. Żeby móc sprawdzić jakie minimalne prawdopodobieństwo powinno być by na powyższy mecz opłacane było stawianie swoich pieniędzy stworzona została kolumna „opłacalne prawdopodobieństwo”. Jest to prawdopodobieństwo na dany wynik pomnożone przez 88% (kwota, która nie jest objęta podatkiem). Ze względu na to, że od kursów zależny jest zarobek bukmacherów, to źródło powinno być najbardziej prawidłowe. Jednak gdy uda się znaleźć porównywalny wyznacznik, ze względu na rangę sprawy stanie się on bardzo cenny.

Ostatnim źródłem są dane z witryny forebet.com. Ta strona przewiduje nie tylko procent szans na zwycięstwo danej drużyny, ale również prognozuje najbardziej prawdopodobny wynik przez co stanowi dobrą alternatywę do analiz. Podobnie jak moja aplikacja opiera ona się na matematyce, lecz często daje inne wyniki ze względu na sposób liczenia.

Podstrona z analizami konkretnego meczu znajduje się po przejściu ze strony, gdzie znajduje się terminarz i można wybrać konkretny mecz. Ustaliłem by do wyboru były spotkania w zakresie od dnia dzisiejszego do przyszłego tygodnia. Powodem tego jest, że w rundzie mecze odbywają się raz na tydzień. Z uwagi na to kursy na mecz nie są dostępne u bukmacherów kilka tygodni wcześniej, podobnie jak z przewidywaniami portalu forebet. Również w wypadku mojej aplikacji byłoby to niezbyt przydatne, gdyż przez okres przykładowo miesiąca mogłoby się sporo zmienić i moje predykcje mogłyby odwrócić się o 180 stopni. Problemem przy tworzeniu sortowania meczów po dacie, był sposób w jaki termin meczu został podany w API. Zamiast daty pojawiała się liczba, przykładowo 1654293600. Jednak skoro można było pozyskać informacje jakie drużyny będą grać w konkretnym meczu, to można było sprawdzić na internecie, kiedy mecz pomiędzy tymi dwoma drużynami będzie się odbywał. Na podstawie tych obserwacji można było wywnioskować, że 3600 oznacza godzinę różnicy pomiędzy meczami. Za pomocą tego stworzyłem funkcję kompilująca liczbę z API na datę wyświetlaną w aplikacji.

**3.3 Możliwe rozwinięcia**

Aplikacja pobiera większość informacji potrzebnych do generowania po stronie z API Flashscore. Jest to darmowe API, lecz po przekroczeniu danej ilości zapytań w miesiącu, aby aplikacja działała prawidłowo na jednym kluczu potrzebna jest zapłata. W przypadku indywidualnej pracy nie mam problemów z liczbą zapytań, a w razie potrzeby mogę podmienić klucz z innego konta w kodzie. Lecz w momencie gdybym chciał opublikować stronę do szerszego grona powstałby większy problem. Dane z API można pobierać raz przed każdą kolejką i generowanie zapytania za każdym kliknięciem w podstronę nie jest potrzebne. W takich sytuacjach skorzystanie z takiego oprogramowania jak Redis wydaje się dobrym rozwiązaniem. W razie takiego wariantu, cenną funkcjonalnością, byłoby zapisywanie przewidywań z wcześniejszych meczy w bazie danych, co umożliwiłoby obliczanie na bieżąco zysku i strat podczas korzystania z konkretnych sposobów typowania. Dodatkowo, mógłbym rozszerzyć aplikacje o więcej popularnych lig. Tak jak wspomniałem w jednym z poprzednich podpunktów wynik często jest zależny od rodzaju ligi ze względu na styl gry, który przekłada się na ilość remisów i liczbę bramek. Aczkolwiek nie byłoby większym problemem, by pobierać informacje o aktualnym procencie remisów za pomocą webscrappingu z portalu eprogresja.net. Dodatkowo istnieje możliwość włączenia statystyk z meczu (takich jak posiadanie piłki, strzały, niebezpieczne ataki) do skryptu obliczającego najbardziej popularny rezultat. Te dane dostępne są w API Flashscore i być może pomogłyby w lepszej predykcji, aczkolwiek wymagałoby to głębszej analizy i eksperymentów nad nowym modelem.

**3.4 Poszczególne widoki**

Screeny strony internetowej

**4 Testy**

**4.1 Predykcje wyników a rzeczywistość**

Jako próbę badawczą wziąłem pod lupę 17 kolejkę sezonu 2022/2023 w Ekstraklasie. Kolejka oznacza wszystkie mecze grane przez dany weekend w lidze. W niej wszystkie zespoły rozgrywają po jedno spotkanie. 17 kolejka była ostatnim cyklem meczy przed zakończeniem rundy jesiennej.

**Warta Poznań – Stal Mielec**

|  |  |  |  |
| --- | --- | --- | --- |
| **Faktyczny rezultat** | **Bukmacher** | **Moja aplikacja** | **Forebet** |
| **1-1 / X** | 1-0 / 1 | **1-1 / X** | 0-2 / 2 |

**Wisła Płock - Cracovia**

|  |  |  |  |
| --- | --- | --- | --- |
| **Faktyczny rezultat** | **Bukmacher** | **Moja aplikacja** | **Forebet** |
| **1-0 / 1** | 1-1 **/ 1** | 1-2 / 2 | 1-1 / X |

**Korona Kielce – Widzew Łódź**

|  |  |  |  |
| --- | --- | --- | --- |
| **Faktyczny rezultat** | **Bukmacher** | **Moja aplikacja** | **Forebet** |
| **0-1 / 2** | 1-1 / **2** | 0-3 **/ 2** | 0-2 / **2** |

**Górnik Zabrze – Miedź Legnica**

|  |  |  |  |
| --- | --- | --- | --- |
| **Faktyczny rezultat** | **Bukmacher** | **Moja aplikacja** | **Forebet** |
| **0-3 / 2** | 1-0 / 1 | 4-2 / 1 | 2-0 / 1 |

**Jagiellonia Białystok – Lech Poznań**

|  |  |  |  |
| --- | --- | --- | --- |
| **Faktyczny rezultat** | **Bukmacher** | **Moja aplikacja** | **Forebet** |
| **1-2 / 2** | 1-1 / **2** | **1-2 / 2** | **1-2 / 2** |

**Zagłębie Lubin – Raków Częstochowa**

|  |  |  |  |
| --- | --- | --- | --- |
| **Faktyczny rezultat** | **Bukmacher** | **Moja aplikacja** | **Forebet** |
| **1-2 / 2** | 0-1 / **2** | 0-2 **/ 2** | **1-2 /2** |

**Radomiak Radom – Pogoń Szczecin**

|  |  |  |  |
| --- | --- | --- | --- |
| **Faktyczny rezultat** | **Bukmacher** | **Moja aplikacja** | **Forebet** |
| **1-2 / 2** | 1-1 / **2** | 2-2 **/** X | 2-2 / X |

**Lechia Gdańsk – Piast Gliwice**

|  |  |  |  |
| --- | --- | --- | --- |
| **Faktyczny rezultat** | **Bukmacher** | **Moja aplikacja** | **Forebet** |
| **1-3 / 2** | 1-1 / 1 | 1. / 1 | 0-0 / X |

**Śląsk Wrocław – Legia Warszawa**

|  |  |  |  |
| --- | --- | --- | --- |
| **Faktyczny rezultat** | **Bukmacher** | **Moja aplikacja** | **Forebet** |
| **0-0 / X** | 1-1 / 2 | 1-2 / 2 | 1-3 / 2 |

Wygrana gospodarzy oznacza 1, remis X, a zwycięstwo gości 2. Wyniki które zgadzają się z rzeczywistością, są podkreślone i pogrubione. Przewidywane typy bukmachera brałem na podstawie najmniejszego kursu. Zazwyczaj jest to wynik 1-0 bądź 1-1. Wynik 1 bramkowy pojawia się zazwyczaj w momencie gdy jedna z drużyn wyraźnie jest lepsza od drugiej. Remisowy przewidywany jest, gdy spotkanie w teorii powinno być bardziej wyrównane. Zdarzają się takie sytuację, jak np. powyżej w meczu Lechia – Piast, gdzie najbardziej prawdopodobnym wynikiem według tego źródła, jest wygrana Lechii, a dokładny wynik prognozowany to remis 1-1. Fakt że bukmacherzy prognozują większość meczy głównie tylko na te dwa wyniki to nie przypadek. W piłce nożnej statystycznie pada około 2,5 gola w zależności od ligi. W momencie, gdy padnie dużo bramek wynik jest trudniejszy do przewidzenia, gdyż możliwych wyników z dużą ilością bramek jest sporo więcej. Widać to dobrze w wyżej cytowanych statystykach, z podsumowania sezonu 2021/2022 prowadzonego przez portal 90 minut. Wyniki te są najczęściej występującymi rezultatami, spośród 306 meczy aż 105 kończyło się wynikiem 1-0 bądź 1-1.

Zakładając, że badawczo pewien gracz postawiłby na zakładach bukmacherskich po 10 walut na a) zwycięzcę meczu i b) dokładny wynik u bukmachera zyskałby (od zysków odliczony został już podatek):

**Najmniejszy kurs bukmachera:**

a) wygrana 1.05 przy wkładzie 90, czyli 1.17% zysku

wygrana 10.39 na meczu Wisła Płock – Cracovia

wygrana 9.62 na meczu Korona Kielce – Widzew Łódź

wygrana 7.25 na meczu Jagiellonia – Lech Poznań

wygrana 2.76 na meczu Zagłębie – Raków

wygrana 11.03 na meczu Radomiak – Pogoń

strata 40.00 na pozostałych meczach

b) strata 90, czyli 100% wkładu

**Moja aplikacja**

a) strata 12.33 przy wkładzie 90, czyli 13.7% na minusie

wygrana 19.04 na meczu Warta Poznań – Stal Mielec

wygrana 9.62 na meczu Korona Kielce – Widzew Łódź

wygrana 7.25 na meczu Jagiellonia – Lech Poznań

wygrana 2.76 na meczu Zagłębie – Raków

strata 50.00 na pozostałych meczach

b) wygrana 35.05 przy wkładzie 90, czyli 38.9% zysku

wygrana 43.42 na meczu Warta Poznań – Stal Mielec

wygrana 61.63 na meczu Jagiellonia Białystok – Lech Poznań

strata 70 na pozostałych meczach

**Forebet**

a) strata 40.37 przy wkładzie 90, czyli 43.63% na minusie

wygrana 9.62 na meczu Korona Kielce – Widzew Łódź

wygrana 7.25 na meczu Jagiellonia – Lech Poznań

wygrana 2.76 na meczu Zagłębie – Raków

strata 60 na pozostałych meczach

b) wygrana 60.62 przy wkładzie 90, czyli 67.36% zysku

wygrana 68.99 na meczu Zagłębie Lubin – Raków Częstochowa

wygrana 61.63 na meczu Jagiellonia Białystok – Lech Poznań

strata 70 na pozostałych meczach

Opcje, które w tej kolejce okazały się opłacalne:

1. Typowanie dokładnego wyniku na podstawie witryny forebet.com: 2/9, +67,36%
2. Typowanie dokładnego wyniku na podstawie mojej aplikacji: 2/9, +38.9%
3. Typowanie zwycięzcy patrząc na najmniejszy kurs: 5/9, +1.17%

Opcje, które w tej kolejce okazały się nieopłacalne:

1. Typowanie dokładnego wyniku patrząc na najmniejszy kurs: 0/9 -100%
2. Typowanie zwycięzcy na podstawie witryny forebet.com: 3/9, -43.63%
3. Typowanie zwycięzcy na podstawie mojej aplikacji: 4/9, -13.7%

Prawidłowa predykcja jest mocno zależna od szczęścia. W analizowanej kolejce było sporo takich przypadków meczów, w których mało brakowało, by przewidywany wynik przez moją aplikacje bądź serwis forebetu się sprawdził. Zawodnik Legii Warszawa nie wykorzystał rzutu karnego, a następnie strzelił gola, który został anulowany przez system VAR, przez co mecz zakończył się jedynie bezbramkowym remisem. Pogoń Szczecin strzeliła bramkę w 90 minucie, która zaważyła o jej zwycięstwie, a obie aplikacje twierdziły przed meczem, że będzie remis. Z drugiej strony dzięki bramkom w końcowych minutach Warty, Lecha i Widzewa wcześniejsze predykcje się spełniły.

Podsumowując, mimo braku gwarancji lepszego typowania przez matematyczne kalkulatory korzystanie z dodatkowych narzędzi do analizy przewidywania meczu może znacząco pomóc w trafnej predykcji. Szczególnie duży wpływ ma wtedy, gdy próbujemy przewidzieć konkretny wynik, a nie tylko zwycięzcę. Wynika to ze stosunku opłacalności ryzyka do prawdopodobieństwa.

**4.2 Zależność pomiędzy ocenami z SofaScore, a Fifą / Football Managerem**

Postanowiłem sprawdzić oceny ogólne z gier najlepszych zawodników z Premier League i porównać je do jakości występów w przyszłym sezonie. Jakość występów definiuje jako średnią ocena z sezonu w Sofascore. Ta witryna wystawia oceny za każdy mecz na podstawie szczegółów, takich jak ilość odbiorów, celne podania itp. Można te dane analizować w szczegółowy sposób i każdy atrybut przyrównać w mojej aplikacji, lecz chciałbym skupić się na tej najważniejszej wartości. Pozwoli to określić w której grze oceny zostały nadane w lepszy sposób. Aby porównanie było jak najbardziej przejrzyste w każdej z kategorii, wartością będzie jakość piłkarza w stosunku do reszty, czyli procent zawodników od których jest lepszy bądź tak samo dobry. W tabeli niżej najlepszych 10 zawodników z każdej gry, którzy rozegrali minimum 450 minut w sezonie.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Imie i nazwisko | Fifa | FM | Sofascore | Bardziej miarodajne źródło |
| Bernardo Silva | 98.17 | 100 | 73.74 | Fifa |
| Kevin De Bruyne | 100 | 99.77 | 100 | Fifa |
| Sadio Mane | 99.77 | 99.77 | 97.26 | brak |
| Mohamed Salah | 99.77 | 99.77 | 93.15 | brak |
| Harry Kane | 98.86 | 99.77 | 99.77 | FM |
| Andrew Robertson | 98.17 | 99.77 | 84.47 | Fifa |
| Bruno Fernandes | 98.17 | 99.77 | 99.54 | FM |
| Sergio Aguero | 99.09 | 98.17 | 71 | FM |
| Ngolo Kante | 98.86 | 98.17 | 78.77 | FM |
| Fabinho | 98.17 | 98.17 | 73.06 | brak |
| Raheem Sterling | 98.86 | 87.21 | 60.27 | FM |

W powyższym porównaniu wygrywa Football Manager, który trafniej przewidział większą liczbę (pięciu) zawodników. Fifa zrobiła to trzykrotnie, tak samo często pozycja w rankingu zawodników była identyczna. Po liczbach widać, że niektórzy zawodnicy, którzy mieli być tymi najlepszymi, spisywali się co najwyżej dobrze i researcherzy gier nie trafili dobrze z niektórymi ocenami. Na podstawie tak dużych różnic tak wielu zawodników można wysnuć tezę, że wielu zawodników dostaje wysoką ocenę w oparciu o dobrą reputacje. Ona też się nie bierze znikąd, bo żeby zawodnik był znany w pozytywny sposób, musi grać w dobrym klubie i na dobrym poziomie jakiś okres czasu wcześniej. Jednak można zauważyć niechęć do szybkich i gwałtownych zmian w oparciu o formę z ostatniego czasu. Aby zobaczyć która gra szybciej zauważyła możliwy talent i wykryła możliwość dobrej dyspozycji w sezonie, w którym tytuł był dedykowany stworzyłem tabelę najlepszych zawodników według sofascore, którzy nie byli w TOP 10 obu gier.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Imie i nazwisko | Fifa | FM | Sofascore | Bardziej miarodajne źródło |
| Jack Grealish | 83.33 | 98.17 | 99.32 | FM |
| Gareth Bale | 91.10 | 87.21 | 99.09 | Fifa |
| James Rodriguez | 87.90 | 98.17 | 98.86 | FM |
| Mason Mount | 83.33 | 98.17 | 98.63 | FM |
| Riyad Mahrez | 95.66 | 98.17 | 98.17 | FM |
| Trend Alexander-Arnold | 98.17 | 98.17 | 97.95 | brak |
| Rodri | 95.66 | 87.21 | 97.72 | Fifa |
| Antonio Rudiger | 86.53 | 87.21 | 97.49 | FM |
| James Ward-Promise | 78.54 | 87.21 | 97.26 | FM |
| James Maddison | 86.53 | 87.21 | 96.80 | FM |

Podobnie jak wcześniej, bardziej miarodajnym źródłem okazał się Football Manager, który zwyciężył 7 razy. Dwukrotnie bliżej przewidywań była Fifa, a raz dany zawodnik był na takiej samej pozycji w rankingu.

**Podsumowanie i wnioski**

Za pomocą aplikacji porównującej statystyki zawodników można zauważyć, że zarówno w jednej, jak i drugiej grze atrybuty oceny ogólnej, czyli najważniejszej statystyki wyglądają podobnie. Przykładowo trójka Sadio Mane, Mohamed Salah i Virgil van Dijk, która jest w topce jeśli chodzi o zawodników z Premier League, ma takie same wartości ocen. Virgil van Dijk złapał poważną kontuzje i grał bardzo mało w analizowanym sezonie, lecz pozostali zawodnicy w rzeczywistości spisywali się dobrze. Znalazło się paru piłkarzy, którzy byli niżej oceniani, a zaliczyli ciut lepszy sezon według sofascore, lecz jest to akceptowalna różnica. Mimo paru podobieństw można zauważyć jednak, że zdarzają się też takie sytuacje w których statystyki jednej z gier bardziej faworyzują jednego zawodnika. Tak dzieje się przykładowo z Bernardo Silvą. W Football Managerze jest on stawiany za najlepszego zawodnika w całej lidze, a w Fifie zajmuje 9 miejsce w rankingu. W weryfikacji boiskowej okazało się, że to osoby odpowiedzialne za Fife miały racje. Kevin De Bruyne, który był stawiany przez nich w roli lidera tych rozgrywek, rzeczywiście posiadał najlepszą średnią ocenę ogólną w serwisie sofascore po całym sezonie. Dodatkowo Bernardo Silva, będący faworytem Football Managera zaliczył przeciętny sezon. Algorytmy portalu sofascore oceniły jego postawę gorzej od ponad 26 procent piłkarzy. Można też dojść do takich wniosków, że przewidywanie dokładnej dyspozycji piłkarza jest bardzo trudno. Prócz czołowej piątki, tak zwanych gwiazd ligowych, większość zawodników uważanych za jednych z najlepszych, zawodziła. Ich obecna wysoka ocena prawdopodobnie jest skutkiem świetnej gry w poprzednich sezonach. Mało który piłkarz potrafi być w świetnej formie przez parę lat. Aczkolwiek duża ocena ogólna może wynikać częściowo z wyrobionej sobie reputacji na temat danego piłkarza. Obydwie gry mogą celowo wolno obniżać notowania zawodników. W razie nagłego powrotu do szczytowej formy zawodnika, niska ocena w grze mogłaby powodować zła reakcję ze strony fanów. A w razie niedocenienia zawodnika, przed którym dopiero najlepsze sezony, sprawa jest bardziej zrozumiała. Ogólnie rzecz biorąc trudno jest znaleźć przesłanki o kradzieży danych jednego ze źródeł. Być może występuje nieznaczna inspiracja, lecz tego nie da się jednoznacznie stwierdzić i uniknąć. Jeśli chodzi o prawdziwość danych, tutaj sprawa wygląda indywidualnie. W przypadku jednego zawodnika gra może przewidzieć jego dyspozycje w danym sezonie bardzo dobrze, w innym przypadku predykcja oceny nie będzie spójna z jakością występów piłkarza. Jako całość oceniam spójność danych solidnie, ale nie bez wad. Mimo iż Football Manager ustala wartości w zakresie od 1 do 20, a Fifa od 1 do 99 to uważam źródło danych z Football Managera za minimalnie bardziej jakościowe od Fify. Zauważyłem również, że w danych z symulatora trenera piłkarskiego wartość szybkości ma mniejszy wpływ na wysokość oceny ogólnej.

Jeśli chodzi o aplikacje porównującą przewidywanie wyników to tutaj również dane są w miarę podobne do siebie. Najbardziej spośród trzech źródeł wyróżniają się kursy bukmacherskie. Działają one trochę inaczej ponieważ biorą pod uwagę, takie rzeczy jak nieobecność zawodników pauzujących, przewidywania graczy i inne możliwe tajne informacje, o których nie wie przeciętny człowiek. Jednak to jest system, który nie jest idealny i w niektórych sytuacjach może gorzej sprawdzać się od innych źródeł. Kolejnymi źródłami są witryna forebet oraz moja aplikacja. Przewidywane wyniki są często podobne, lecz zauważyłem jedną zależność. Rezultaty różnią się, gdy spotykają się drużyny, dla których miejsce rozgrywania meczu jest ważnym czynnikiem. Tak było w przypadku meczu Warta Poznań – Stal Mielec. Drużyna z Poznania bardzo słabo sobie radzi na własnym stadionie. Dla Stali statystycznie teren nie wpływa w dużym stopniu na wynik spotkania. Jeśli chodzi o ogólną formę to wygląda to podobnie i oba zespoły są blisko siebie w tabeli. Mój algorytm działa w takim sposób, że 2/3 siły drużyny stanowi jej obecna forma ogółem, a 1/3 to forma na terenie na jakim gra. Taka proporcja została stworzona na podstawie danych statystycznych z poprzedniego sezonu z portalu 90 minut. W serwisie forebet czynnik formy w miejscu rozgrywania meczu musi być większy. Nie jestem w stanie wydobyć dokładnego wzoru, bo to by było ekstremalnie ciężkie, ale przewiduje, że ten współczynnik musi wynosić około 50%. Wnioskuje tak ponieważ obydwie drużyny pod względem formy czy też innych czynników, które mogłyby być brane pod uwagę takich jak np. ostatnie mecze między obiema drużynami, idą mniej więcej na równi. Przedmeczowe przewidywania forebetu wyświetlają 26% szans na zwycięstwo Warty, 28% na remis i 47% na zwycięstwo co daje w przybliżeniu prawdopodobnie 40-60 na korzyść Stali, gdyby mecz nie mógł zostać zakończony remisem. Stosunek ilości punktów Stali na wyjeździe w tym sezonie do sumy ilości punktów Stali na wyjeździe i punktów u siebie Warty wynosi 70%. Gdyby zsumować 50-50 formy ogólnej z 70-30 formy zależnej od terenu taką samą wagą to średnia wyszłaby 60-40, czyli taka jak przypuszczalna wartość sił obydwu drużyn przed rozkładem bez możliwości remisu. Analiza procentów witryny forebet w oparciu o tą tezę przy innych spotkaniach pozwala wysunąć podobne wnioski. Strona ta jest na pewno bardziej zaawansowana obliczeniowo od mojego algorytmu, lecz własna metoda skupienia się na jednej lidze, może sprawić, że osobisty skrypt będzie dawał równie dobre, lub nawet lepsze przewidywania. Patrząc na testowanie wszystkich źródeł z rzeczywistością można wysnuć wnioski, że przewidywanie dokładnych wyników za pomocą matematycznych algorytmów jest bardziej opłacalne od typowania zwycięzcy spotkania. Wynika to z atrakcyjności kursów i niskiego prawdopodobieństwa u bukmacherów na konkretny wysokobramkowy rezultat. Oczywiście jest to często kwestia szczęścia bo żaden system nie jest w stanie idealnie przewidywać wynik każdego meczu.

**Bibliografia / netografia**

<https://www.futhead.com/21/players/?level=all_nif&league=13&club=all&bin_platform=ps>

<https://fmdataba.com/21/l/2613/premier-league/best-players/>

https://www.transfermarkt.pl/statistik/5jahreswertung

<https://www.eprogresja.pl/statystyki/pilkarskie/remisy_w_ligach/>

<https://twitter.com/pklimek99>

<https://eloratings.net/>

<http://img.90minut.pl/misc/stats-ekstraklasa-2021-22-wiosna.html>

<https://www.oddsportal.com/soccer/poland/ekstraklasa/results/>

https://pl.wikipedia.org/wiki/Ranking\_Elo\_(pi%C5%82ka\_no%C5%BCna)