

Projekt: Football progress analysis - Aplikacja do analizy piłkarzy na podstawie danych z ich występów oraz ceny rynkowej

Skład zespołu: Krzysztof Bieniasz, Mateusz Nabywaniec, Michał Przybycień

1. Motywacja

W dzisiejszych czasach na temat jakichkolwiek rozgrywek sportowych zbierane są ogromne ilości danych. Dodatkowo w wielu sportach występuje zaangażowanie wielkich inwestorów, którzy chcą jak najbardziej optymalnie wykorzystywać swoje pieniądze. Jednym z głównych sposobów inwestycji w tego typu branży jest zakup nowych zawodników. Pojawia się jednak pytanie - kto jest wart wydania pieniędzy? Niektórzy chcą ufać swojemu "instynktowi" przy tego typu zakupach, inni wolą postawić na nieco bardziej solidne metody wyboru. Z tego właśnie powodu chcemy stworzyć narzędzie, które będzie pozwalało użyć wiedzy matematycznej oraz nowoczesnych metod eksploracji danych w procesie wyboru kandydatów do zakupu. Analizując historię oraz statystyki zawodników, chcemy w wyniku otrzymać zestawienie, którzy piłkarze mają największą szansę być udaną inwestycją w kolejnych latach.

2. Klient

Nasza aplikacja jest skierowana do klubów piłkarskich, które chcą wzmocnić skład na rynku transferowym. Aplikacja pomagalaby w wyszukiwaniu obiecujących zawodników, których wartość na rynku transferowym nie jest jeszcze wysoka jak na jakość ich gry oraz perspektywy rozwoju. Poprzez wykorzystywanie naszego produktu działy skautingu będą w stanie szybciej od konkurencji zdecydować się na zakup obiecującego zawodnika.

3. Użyte dane.

W projekcie zdecydowaliśmy się na użycie danych pochodzących z dwóch źródeł. Pierwszym z nich jest repozytorium <https://github.com/vaastav/Fantasy-Premier-League>. Repozytorium zawiera szczegółowe dane odnośnie występów piłkarzy angielskiej Premier League z ostatnich 5 sezonów. Dane są zapisane w postaci plików csv. Fantasy Premier League to gra w której użytkownik buduje swój skład z prawdziwych zawodników występujących w lidze angielskiej. Zawodnicy otrzymują punkty na podstawie aktualnych występów w sezonie. Celem gry jest dobranie składu, aby zdobyć jak najwięcej punktów. Dane z FPL zawierają statystyki dotyczące poszczególnych zawodników takie jak: minuty rozegrane, liczba asyst, strzelonych goli, czystych kont, współczynniki influence, creativity, thread i wiele innych.

Drugim źródłem danych jest strona <https://www.transfermarkt.pl/>. Zawiera ona szczegółowe dane o transferach w świecie piłki nożnej, a także wyceny wartości piłkarzy wraz z danymi historycznymi. Aby pozyskać dane stworzyliśmy scraper ściągający dane dla zawodników w występujących w Premier League w latach 2016-2021.

4. Funkcjonalności:

- Raport na temat danego piłkarza
- Prognoza rozwoju danego piłkarza
- Wylistowanie piłkarzy z największą możliwością rozwoju spełniających dane kryteria (pozycja na boisku, wiek)
- Porównanie dwóch piłkarzy na zasadzie head-to-head

5. Struktura bazy danych

Dane z FPL nie wymagały dużych modyfikacji, jako że zostały odpowiednio przygotowane w repozytorium.

Dane z transfermarktu wymagały wstępnego przygotowania. Przykładowo należało ujednolicić kwoty dotyczące wartości piłkarzy, sparsować podane daty.

W celu wygodnego korzystania z danych pochodzących z dwóch źródeł, zdecydowaliśmy się stworzyć bazę danych sqlite. Jej użycie pozwala na wczytanie danych do dataframów w prosty sposób korzystając z zapytań.

Fantasy Premier League:

players - identyfikator zawodnika z danych z Fantasy Premier League oraz imię i nazwisko
player_statistics - dane dotyczące zawodnika w danym sezonie pochodzące z Fantasy Premier League

Transfermarkt:

players_transfermarkt_singles - podstawowa tabela z danymi piłkarza zebranych z transfermarkt, do niej odnoszą się pozostałe tabele z tego obszaru.

clubs - lista klubów występujących w Premier League w latach 2016 - 2021 wraz z podanym rokiem występowania

player_values_transfermarkt - dane dotyczące wartości zawodnika w momencie w przeszłości, kiedy była szacowana jego wartość

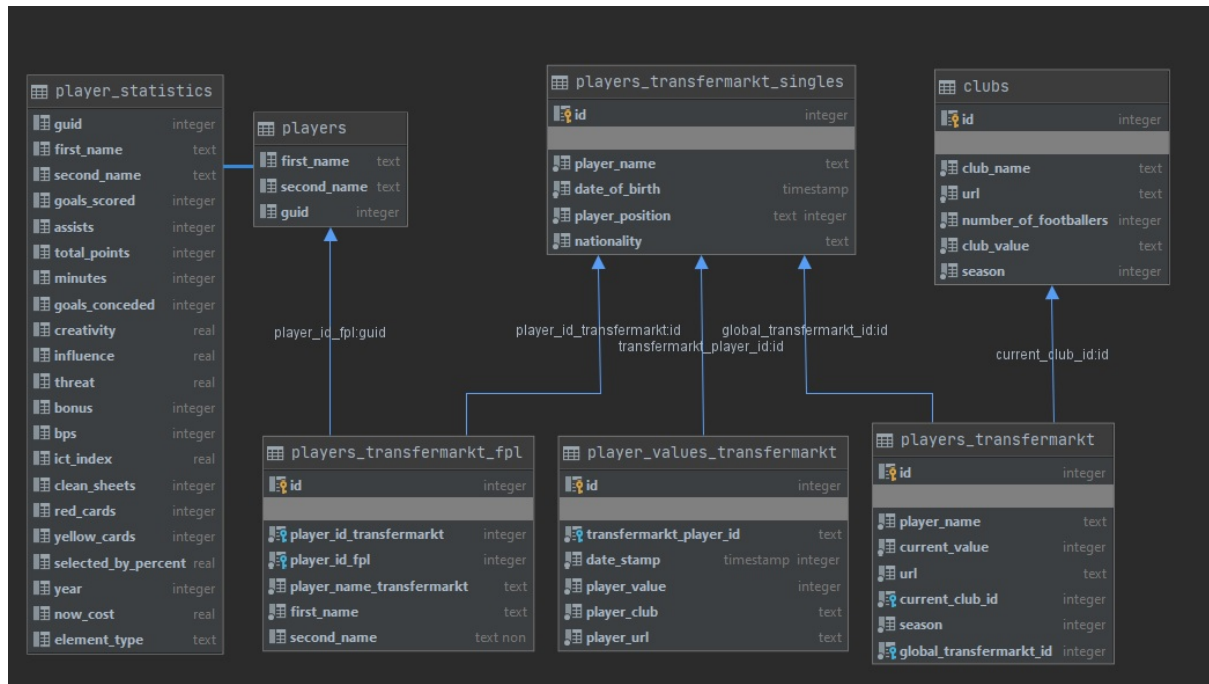
players_transfermarkt - lista zawodników występujących w Premier League w latach 2016 - 2021 wraz z podanym rokiem występowania oraz klubem w danym sezonie

Połączenie danych:

players_transfermarkt_fpl - tabela łącznikowa pozwalająca na korzystanie i łączenie danych zebranych z FPL oraz portalu transfermarkt. Połączenie było wykonane na podstawie imienia i nazwiska zawodnika w dwóch źródłach. Zdarzały się często przypadki, że pisownia była różna albo imię miało inną formę, dlatego to połączenie musiało nastąpić w bardziej nadzorowany sposób.

Ogółem posiadamy kompletne dane o 1300 piłkarzach, gdzie mamy pełną synchronizację danych między dwoma źródłami. Ogółem ze strony FPL mamy informacje o statystykach 1400 graczy, a ze strony transfermarkt o 1700. Mogą się zdarzyć drobne przypadki, gdzie brak sparowania danych wynika z różnych pisowni, ale przede wszystkim te rozbieżności wynikają z tego, że dane pochodzą z dwóch niezależnych źródeł. Dane z Fantasy Premier League zostały już wcześniej zmodyfikowane, zanim je wykorzystaliśmy.

Posiadamy również blisko 30 tysięcy archiwalnych wycen piłkarzy - zazwyczaj każdy piłkarz posiada 2-3 wyceny w ciągu roku.



Baza danych zawiera dosyć rozbudowane informacje i może również zawierać informacje, które zostaną wykorzystane przy rozbudowywaniu aplikacji. Przykładowo aktualnie informacja o klubie jest tylko dodatkową informacją, a może służyć jego atrybut, który też wpływa na wartość zawodnika.

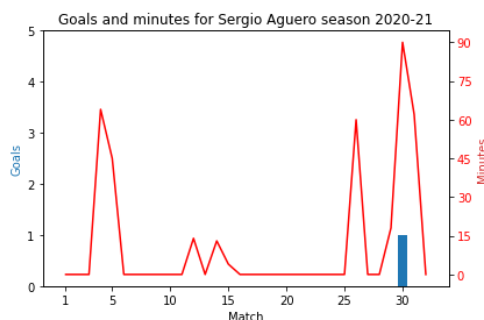
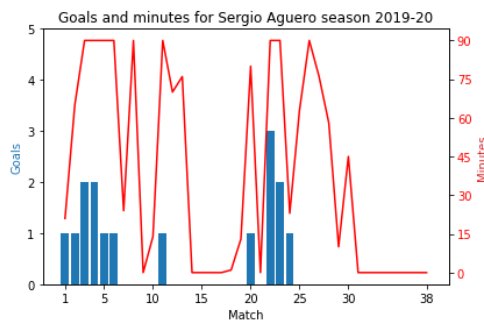
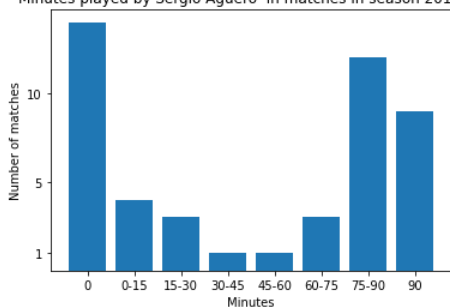
6. Podstawowe analizy

W początkowej fazie naszego projektu dosyć intensywnie analizowaliśmy zebrane dane. Nie wszystkie analizy zostały wykorzystane w ostatecznym produkcie, ale przedstawiamy je tutaj jako dowód wstępnej analizy danych.

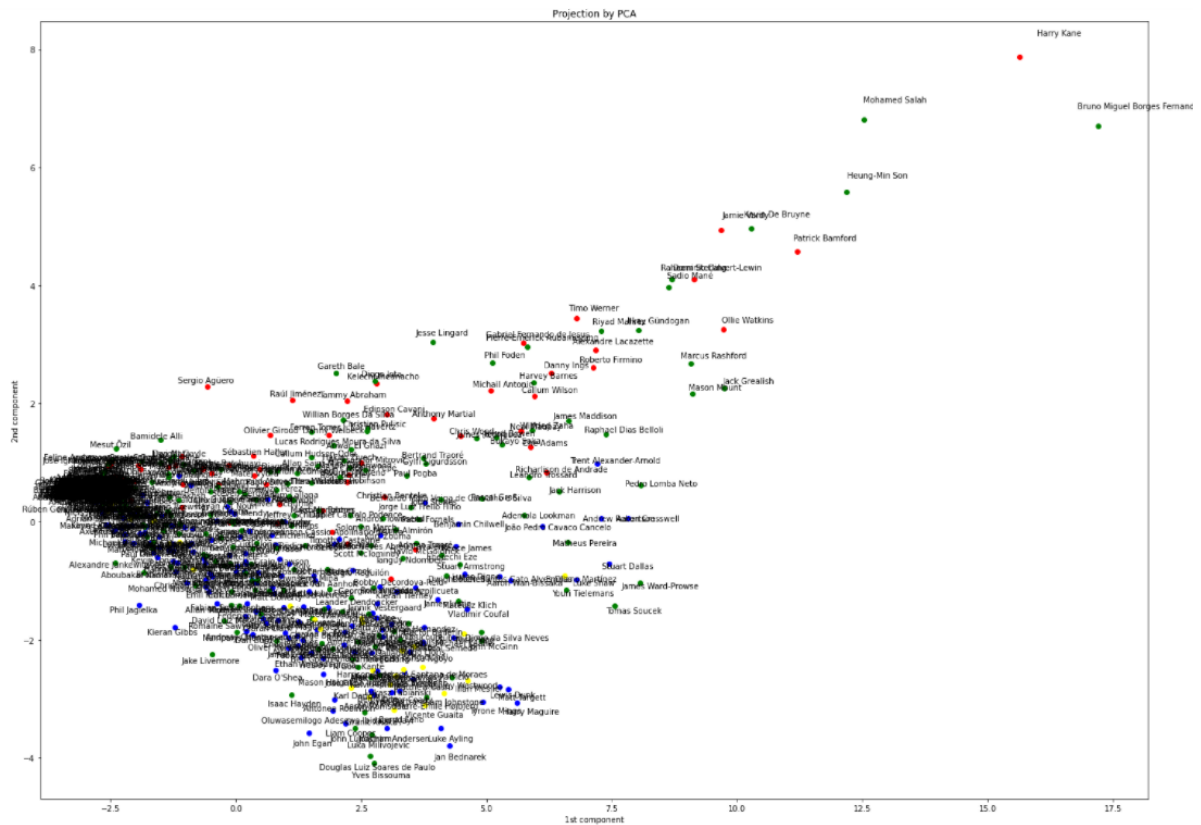
Wyświetlanie statystyk danego piłkarza

Statistics for Sergio Aguero in season 2019-20
Minutes played: 1449
Average minutes per match: 38
Goals scored: 16
Assists: 6
Minutes for goals : 90
Minutes for assist : 241
Structure of minutes played in match: {'0': 14, '0-15': 4, '15-30': 3, '30-45': 1, '45-60': 1, '60-75': 3, '75-90': 12, '90': 9}

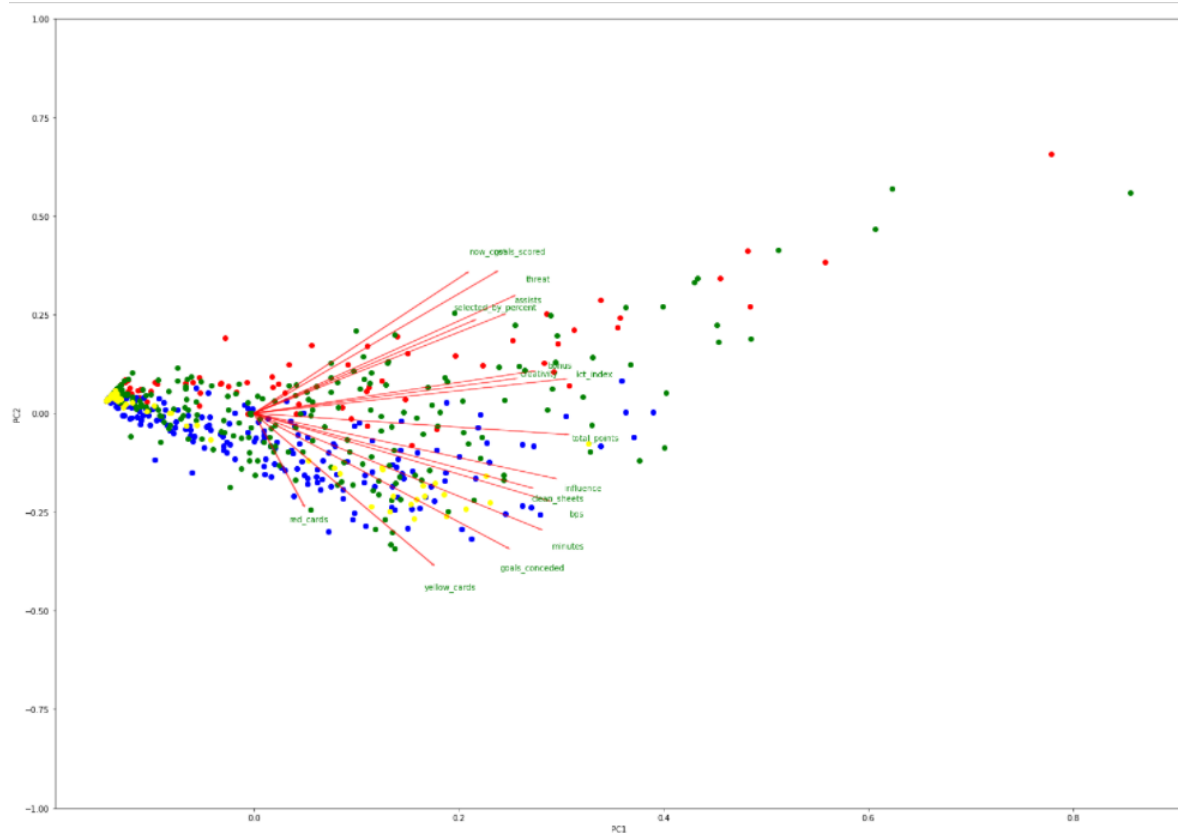
Minutes played by Sergio Aguero in matches in season 2019-20



Dokonaliśmy analizy przy użyciu PCA, aby stwierdzić które statystyki są najbardziej ze sobą powiązane. Na wykresach czerwonym kolorem zostali zaznaczeni napastnicy, zielonym - pomocnicy, niebieskim - obrońcy, a żółtym bramkarze.



Wizualizacja przy użyciu biplota



Jak można zauważyć dzięki PCA widzimy jakie które cechy są ze sobą powiązane. Widzimy że w celu predykcji wartości zawodników będziemy musieli użyć innych cech w zależności od pozycji.

7. Wyszukiwanie perspektywicznych zawodników.

Naszym pomysłem aby wyszukać obiecujących piłkarzy było zastosowanie regresji liniowej w celu wyznaczenia oczekiwanej wartości piłkarza.

W omawianym zadaniu, zmienną objaśnianą jest wartość zawodnika ('player_value') pozyskana z Transfermarkt, natomiast zmiennymi objaśniającymi będą statystyki zawodnika w danym sezonie.

Z powodu różnic związanych z charakterystyką piłkarzy grających na danych pozycjach dokonaliśmy podziału piłkarzy w zależności od pełnionej na boisku funkcji. Piłkarzy podzieliiliśmy na:

- napastników (forwards)
- skrzydłowych (wingers)
- pomocników (midfielders)
- bocznych obrońców (back-defenders)
- środkowych obrońców (center defenders)
- bramkarzy (goalkeepers).

Dla każdej z grup piłkarzy braliśmy inne cech jako zmienne objaśniające. W celu optymalnego doboru tych cech dokonaliśmy wyznaczenia współczynników korelacji Pearsona między wartością zawodnika, a jego statystykami z FPL.

Przykładowo dla napastników:

```
player_value      1.000000
influence         0.575337
bps              0.572236
ict_index         0.560568
goals_scored      0.555728
threat            0.541196
total_points      0.540146
assists           0.521259
clean_sheets      0.517021
bonus             0.515530
creativity         0.508586
selected_by_percent 0.470258
minutes           0.432320
goals_conceded    0.255672
yellow_cards      0.229661
year              0.077850
age               0.056457
red_cards         0.035967
player_id_fpl     -0.143684
guid              -0.143684
player_id_transfermarkt -0.240256
id                -0.240615
Name: player_value, dtype: float64
```

Zgodnie z zasadami, zdecydowaliśmy się na wzięcie kilku cech, dla których współczynnik korelacji był większy niż 0,5.

Napastnicy: ['bps', 'influence', 'goals_scored', 'ict_index', 'total_points']

Skrzydłowi: ['threat', 'goals_scored', 'ict_index', 'total_points', 'influence', 'bonus']

Pomocnicy: ['ict_index', 'assists', 'creativity', 'selected_by_percent', 'influence', 'total_points']

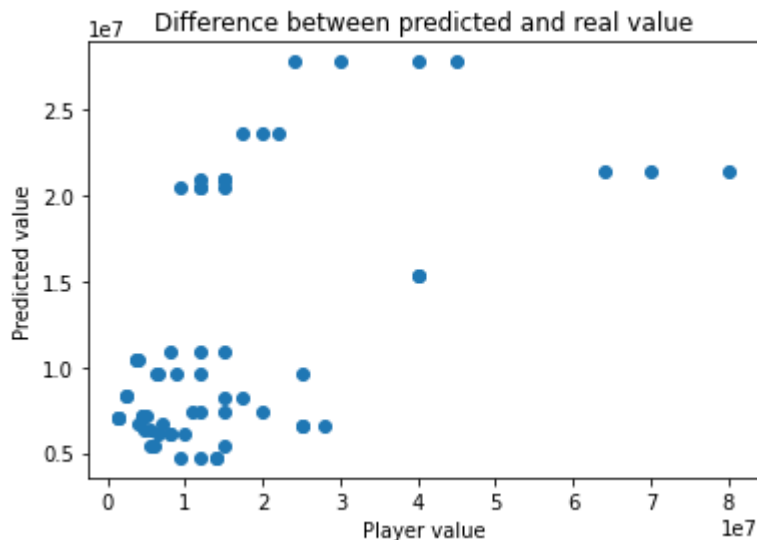
Boczni obrońcy: ['creativity', 'assists', 'bonus']

Środkowi obrońcy: ['clean_sheets', 'total_points', 'bps']

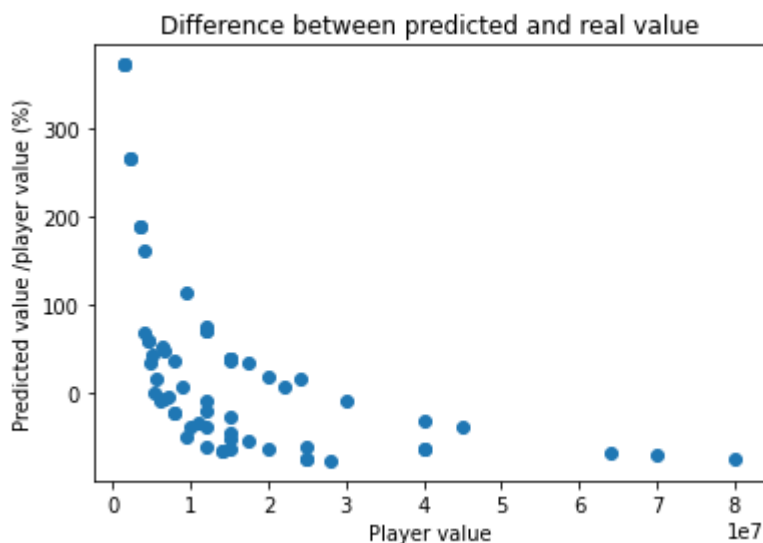
Bramkarze: ['clean_sheets', 'total_points', 'minutes']

Dla poszczególnych zawodników dokonaliśmy predykcji ich wartości. Wyznaczyliśmy także różnicę między wartościami tych zawodników z transfermarktu a przewidywaną wartością.

Wykres poniżej przedstawia różnicę między przewidywaną a aktualną wartością napastników poniżej 24 z sezonu 2020/21.



Kolejny wykres przedstawia różnicę między przewidywaną a aktualną wartością napastników poniżej 24 z sezonu 2020/21 wyrażoną w procentach w stosunku do aktualnej wartości zawodników.



Wyszukiwanie perspektywicznych zawodników polega na wybraniu tych, których wartość przewidywana była większa od wartości aktualnej.

8. Przewidywanie wartości zawodników w przyszłości.

Przewidywanie wartości zawodników jest skomplikowanym zadaniem. Powodem jest to, że forma młodych piłkarzy jest często zmienna, zawodnicy rozwijają się także w różnym tempie. Zdarza się że młody zawodnik osiąga pewien poziom, którego z różnych przyczyn nie potrafi przeskoczyć. Przewidywanie wartości zawodników w przyszłości jest zatem zadaniem trudnym.

Nasz pomysł wynikał z posiadanych danych. Ponieważ posiadamy dane o poszczególnych zawodnikach Premier League z ostatnich 5 sezonów, to dla piłkarza możemy wyszukać podobnych zawodników z ubiegłych lat. W naszej bazie przechowujemy także komplet danych związanych z rozwojem wartości danego piłkarza.

Postanowiliśmy zatem, że dla konkretnego zawodnika wyszukamy podobnych zawodników z podanego sezonu i sprawdzimy jak rosła wartość zawodników od tego sezonu przez 2 lata. Jako przewidywaną wartość bierzemy najwyższą wartość osiągniętą przez podobnych do naszego piłkarza zawodników od danego sezonu przez 2 lata.

9. Produkt końcowy.

Wyszukiwanie perspektywicznych napastników z sezonu 2020/21

Perspective players	Head to head	Closest players	Player details			
<div><div>Search</div><div>forwards</div><div>▼</div><div>2020</div><div>24</div></div>						
Name	Age	Value	Club	Nationality	Predicted value	
1. Kelechi Iheanacho	24	15000000	Leicester City	Nigeria	20900000	Show details
2. Dominic Calvert-Lewin	23	24000000	Everton FC	England	27800000	Show details
3. Fabio Silva	18	9000000	FC Porto	Portugal	9600000	Show details
4. Keinan Davis	22	2300000	Aston Villa	England	8400000	Show details
5. Che Adams	24	12000000	Southampton FC	Scotland	20400000	Show details
6. Michael Obafemi	20	4500000	Southampton FC	Ireland	7100000	Show details
7. Neal Maupay	24	22000000	Brighton Hove Albion	France	23600000	Show details
8. Aaron Connolly	20	4000000	Brighton Hove Albion	Ireland	6700000	Show details
9. Josh Maja	22	6500000	FC Girondins Bordeaux	Nigeria	9600000	Show details
10. Tyler Roberts	21	3600000	Leeds United	Wales	10400000	Show details
11. Oliver McBurnie	24	8000000	Sheffield United	Scotland	10900000	Show details
12. Andi Zeqiri	21	1500000	Brighton Hove Albion	Switzerland	7100000	Show details
13. Oliver Burke	23	5500000	Sheffield United	Scotland	6400000	Show details

Znalezienie podobnych zawodników do podanego

Perspective players

Head to head

Closest players

Player details

Search

2020

2017

Fabio

Silva

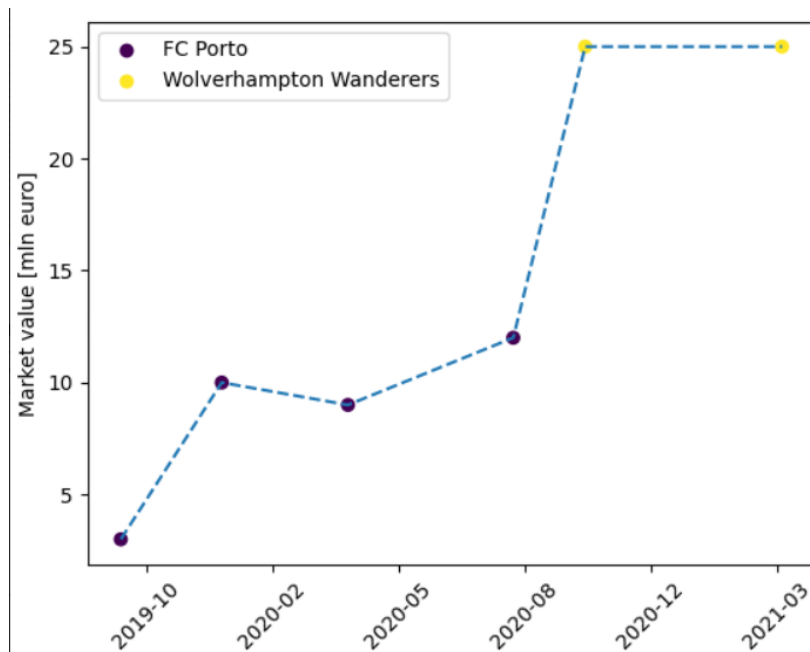
Fabio Silva predicted value: 26000000

Name	Age	Value	Club	Nationality	Predicted value	
1. Kelechi Iheanacho	20	7000000	Manchester City	Nigeria	16400000	Show details
2. Andy Carroll	27	10000000	West Ham United	England	23600000	Show details
3. Erik Lamela	24	26000000	Tottenham Hotspur	Argentina	16800000	Show details
4. Hal Robson-Kanu	27	2000000	Reading FC	Wales	8000000	Show details
5. Jermain Defoe	34	6000000	Sunderland AFC	England	24500000	Show details

Wyświetlanie statystyk zawodnika

Perspective players	Head to head	Closest players	Player details
<div>Fabio Silva 2020</div>			
Show player: Fabio Silva Back			
first_name	Fabio		
second_name	Silva		
date_of_birth	2002-07-19 00:00:00		
player_position	Centre-Forward		
nationality	Portugal		
goals_scored	3		
assists	1		
total_points	48		
minutes	866		
goals_conceded	21		
creativity	117.8		
influence	151.8		
threat	466		
bonus	3		
bps	131		
ict_index	73.3		
clean_sheets	0		
red_cards	0		
yellow_cards	2		
selected_by_percent	1		
date_stamp	2020-10-13 00:00:00		
player_value	25000000		
player_club	Wolverhampton Wanderers		
age	18		
predicted_value	9600000		
predicted_value_diff	600000		
predicted_value_diff_percent	6.666666667		
transfermarkt_url	https://www.transfermarkt.pl/fabio-silva/profil/spieler/505653		

Zmiana wartości zawodnika:



Porównanie head to head:

Perspective players	Head to head	Closest players	Player details
first_name	Fabio	Erk	
second_name	Silva	Lamela	
date_of_birth	2002-07-19 00:00:00	1992-03-04 00:00:00	
player_position	Centre-Forward	Right Winger	
nationality	Portugal	Argentina	
goals_scored	3	2	
assists	1	4	
total_points	48	55	
minutes	866	826	
goals_conceded	21	12	
creativity	117.8	283.1	
influence	151.8	204.8	
threat	466	364	
bonus	3	1	
bpa	131	186	
att_index	73.3	84.7	
clean_sheets	0	4	
red_cards	0	0	
yellow_cards	2	4	
selected_by_percent	1	0.1	
date_stamp	2020-10-13 00:00:00	2017-06-28 00:00:00	
player_value	25000000	26000000	
player_club	Wolverhampton Wanderers	Tottenham Hotspur	
age	18	23	
predicted_value	9600000	13700000	
predicted_value_diff	600000	-11300000	
predicted_value_diff_percent	6.666666667	-45.2	
transfermarkt_url	https://www.transfermarkt.pl/fabio-silva/profil/spieler/305653	https://www.transfermarkt.pl/erik-lamela/profil/spieler/111630	

Fabio

Silva

2020

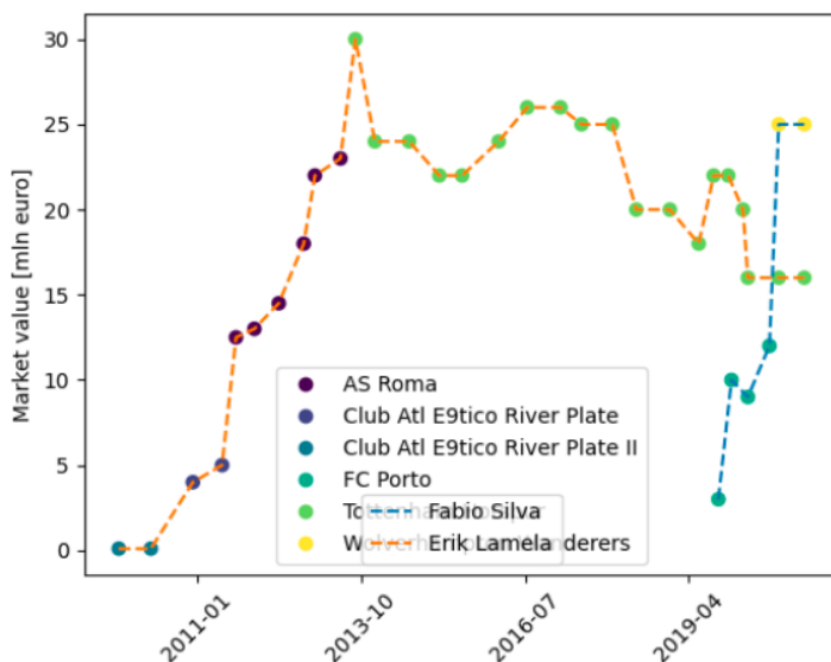
Erk

Lamela

2017

Compare

Porównanie rozwoju wartości dwóch zawodników:



10. Wnioski

- przydatnym narzędziem jest metoda PCA - pozwoliłem nam wyodrębnić istotne cechy dla piłkarzy z różnych pozycji
- w odniesieniu do rzeczywistych "odczuć" człowieka funkcja znajdowania podobnych piłkarzy do danego zwraca całkiem rozsądne rekordy - tutaj wyniki też są czasami ciekawe, bo nie pomyślelibyśmy o jakimś piłkarzu, że jest podobny do innego, ale jednak jak popatrzymy na dane to zaczyna to nabierać sensu.
- udało nam się stworzyć narzędzie, które z punktu widzenia dziedziny zwraca rozsądne dane dotyczące wartości piłkarza. Warto tutaj dopowiedzieć, że specyfika rynku transferowego jest taka, że kwota transferu nie zawsze odpowiada bezpośrednio wartości rynkowej piłkarza. Wchodzą tutaj jeszcze takie zależności jak pozostała długość kontraktu, zapowiedzi piłkarza (na przykład ogłoszenie, że będzie chciał kontynuować karierę w lepszym klubie) czy też umiejętności podbicia ceny przez menedżera zawodnika lub klub
- wadą naszego rozwiązania jest niewystarczająca liczba danych, są piłkarze np. defensywni pomocnicy, którzy nie wyróżniają się dużymi zdobyczami punktowymi w FPL, przez co ich wartość jest niższa niż w rzeczywistości. Brakowało nam takich danych jak np. procent wygranych pojedynków, celność podań, liczba podań
- problemem jest także to, że w rzeczywistości nie tylko statystyki zawodników wpływają na jego wartość lecz także wartość wizerunkowa, występy w innych rozgrywkach (np. Liga Mistrzów) z których nie mieliśmy danych

11. Propozycje rozwoju

Nasz projekt stwarza duże możliwości rozwoju. Myśleliśmy nad następującymi rozwiązaniami:

- dodanie wyszukiwania perspektywicznych piłkarzy z Championship (2 stopień rozgrywek w Anglii).
- poszerzenie danych o dane z innych lig europejskich
- wykorzystanie innego źródła danych o statystykach meczowych zawodników - na przykład istnieją strony z płatnymi statystykami, które są jednolite dla wielu lig.