1. Opis rzeczywistego problemu:

Kto z nas nie chciałby zgarnąć fortuny na zakładach bukmacherskich? Nasz projekt ma to ułatwić poprzez przewidzenie przed sezonem końcowej tabeli ligowej (w tym przypadku Ekstraklasy). Z naszego projektu można korzystać w dwojaki sposób:

- przez typerów jako wskaźnik na kogo postawić, kto jest "pewniaczkiem", a kto może być tegoroczną niespodzianką
- przez bukmacherów aby na podstawie naszych predykcji ustawili odpowiednie kursy na odpowiednie drużyny

Danymi wejściowymi naszego projektu będą:

- wyniki drużyny w poprzednich sezonach (czy są z sezonu na sezon coraz lepsi, słabsi czy oscylują wokół tych samych pozycji)
- niezależna ocena składów przeprowadzona poprzez kanadyjskich ekspertów(oceny zaczerpnięte z gier z serii FIFA)
- statystyki zespołów z ostatnich lat (bramki strzelone, stracone itp.)

Dane te będą z odpowiednią wagą uwzględniane w predykcji wyników. Projekt ten jest związany z zagadnieniem obliczania prawdopodobieństwa i predykcji przyszłości.

2. Możliwe rozwiązania:

- **AIrsenal: A machine learning manager for Fantasy Premier League** (https://www.turing.ac.uk/news/airsenal) - algorytm wykorzystujący proces uczenia maszynowego używany do przewidywania jakości występy danego zawodnika w danym meczu. Został stworzony w celu zwiększenia szans na wygrane tzw. Fantasy Premier League (https://fantasy.premierleague.com/), czyli wirtualnego odpowiednika najwyższej klasy rozgrywkowej na wyspach, gdzie gracze mogą wcielić się w rolę managera własnej drużyny składającej się z zawodników wybieranych spośród klubów występujących w PL w danym sezonie. Istnieje również modyfikacja/rozszerzenie **AIresnal**, które na bazie występów poszczególnych zawodników(składowych) szacuje występy całego zespołu w kontekście poszczególnych meczy lub całego sezonu.

  Metoda ta opiera się na nauczeniu modelu na podstawie statystyk z trzech ubiegłych sezonów oraz aktualnych statystyk z sezonu obecnego dwóch statystyk: poziom drużyny, poziom zawodnika. W modelu zastosowano podejście bayesowskie na podstawie “Dixon & Coles (1997)”. W modelu tym każdy zespół posiada dwie zdolności: do ataku i do obrony, a także parametr zwiększający prawdopodobieństwo domowego zwycięstwa danej drużyny. Na podstawie danych odnośnie drużyn model wylicza szansę na zwycięstwo danej drużyny i prawdopodobny wynik. Po czym mając dane zawodników określają jaka jest szansa, że konkretny zawodnik coś w tym meczu zrobi (strzeli, asystuje, dostanie kartkę itp.) co odpowiadało będzie konkretnym zdobytym przez niego punktom. Na podstawie tych danych algorytm jest w stanie wybrać optymalną jedenastkę na daną kolejkę.

  Zalety tej techniki

  - stosunkowo łatwe i skuteczne

Wady tej techniki:

- potrzebna duża wiedza o każdym zawodniku, której nie ma dla nowo wytranswerowanych graczy
- potrzebna ciągła aktualizacja danych wejściowych

- **Tradycyjne Metody Statystyczne:**
  - Determination of the time series prediction model
    - **Metoda "Bez Zmian"**, gdzie założeniem jest to, że wyniki zawodów sportowych w okresie T + 1 są takie same jak w okresie T
    - **Metoda zmian proporcjonalnych**, w której uważa się, że wyniki zawodów sportowych zmieniają się o pewien procent w czasie. Wyrażenie jest następujące:

$$C_{t+1}^{A} = C_t\left(1 + \frac{C_t - C_{t-1}}{C_{t-1}}\right).$$

    - **Metoda Moving Average Model**. Średnia z obserwowanych wartości w ostatnich kilku okresach jest używana jako wartość przewidywana w okresie predykcji. Wyrażenie jest następujące:

$$C_{t+1}^{A} = \frac{C_t + C_{t-1} + \cdots + C_{t-n+1}}{N} (t \geq N).$$

    - **Weighted Moving Average Model**. Różne wagi są nadawane w zależności od czasu, w którym poszczególne dane są oddalone od okresu predykcji. Ogólnie uważa się, że im bliżej czas jest od okresu predykcji, tym większą wagę należy nadać, ponieważ ostatnia wartość predykcji ma silniejszą zdolność przewidywania i dokładność.
    - **Exponential smoothing model**. Jest to specjalna metoda średniej ważonej, która wykorzystuje średnią ważoną wartości poprzedniej obserwacji i wartości przewidywanej jako przewidywaną wartość następnego okresu.
  - **Stochastic Time Series Model**. Służy do poznania trendu wyników rywalizacji sportowej z określonym czynnikiem metodą regresji. Zmienna niezależna w modelu może być dowolnym z czynników wpływających na wyniki rywalizacji sportowej. Zazwyczaj używa się czasu jako zmiennej niezależnej.

Zalety metod stochastycznych:

- Duża prostota implementacji i działania
- Na podstawie dotychczasowych występów przewiduje co będzie dalej
- potrzebne są tylko surowe i ogólnodostępne dane

Wady:

- Ciężko przewidzieć występy zespołów / zawodników, którzy dopiero co pojawili się w danym sporcie
- nie uwzględnianie czynników takich jak kontuzje, zmęczenie itp.

- **Artificial Neural Network Algorithm**
  Algorytm ten opiera się o zagadnienie sieci neuronowych. Uczenie sztucznej sieci neuronowej odbywa się pod warunkiem, że znany jest tryb wejściowy i idealny tryb wyjściowy. Jej kompleksowy błąd często przyjmuje sumę kwadratów błędów.
  - **BP Neural Network**
    Oblicza wartości wyjściowe hidden layer i output layer nodes. Kalkuluje ich błędy. Poprawia wagi oraz wartości
  - **BP Algorithm Improvement:**
    - **Additional Momentum Value Method**, opiera się ona na metodzie propagacji wstecznej, dodając do każdej zmiany wagi i wartości progowej wielkość zmiany dynamicznej proporcjonalną do ostatniej wagi i wartości progowej oraz generując nowe zmiany wagi i wartości progowej zgodnie z metodą propagacji wstecznej.
    - **Adaptive Learning Rate Method**, inaczej dostosowanie szybkości uczenia się na podstawie reguł samoadaptacji jest korzystne dla skrócenia czasu.

  Zalety:
  - potrafi przewidzieć przyszłe występy z dość dużą dokładnością
  - stosunkowo duża pewność predykcji (jeśli w sporcie można mówić o pewności)
  - model cały czas się uczy i jest skuteczniejszy

  Wady:
  - skomplikowane i złożone zagadnienie
  - musimy znać tryb wejściowy i idealny tryb wyjściowy
  - potrzeba bardzo duża wiedza o danych i modelu

3. Wybrane rozwiązanie:

- Opis ogólny:

W projekcie wykorzystamy własny pomysł na rozwiązanie tego problemu (bazujące na innych rozwiązaniach - głównie z metod stochastycznych). Danymi wejściowymi będzie tabela zawierająca dane takie jak oceny z fify (na sezon obecny oraz trzy poprzednie), a także statystyki zespołów z ostatnich trzech lat (bramki strzelone, stracone oraz punkty). Są to dane powszechnie dostępne jednak trzeba je odpowiednio uszeregować w tabeli w pliku csv, który będzie wejściem naszego algorytmu.

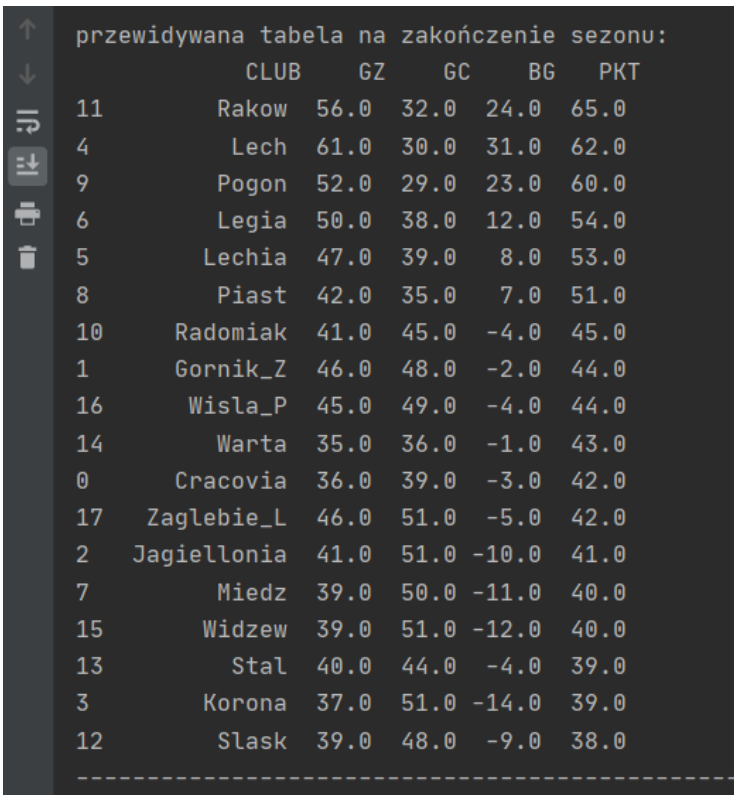
- Opis naszej metody

Nasza metoda będzie polegała na sprawdzeniu czy “moc” danego zespołu wzrosła czy zmalała w porównaniu do lat poprzednich i na podstawie tego przewidzenie czy poradzą sobie lepiej czy gorzej niż w poprzednich sezonach (więcej czy mniej punktów) co doprowadzi do ustalenia końcowej tabeli.
Pierwszą rzeczą, którą wykonujemy jest sprawdzenie czy mamy jakieś braki w danych. Jako, że taka rzecz na 100% ma miejsce (dla beniaminków na pewno nie ma statystyk z poprzedniego) to takiej drużynie przyznajemy najmniejszą średnią ilość punktów oraz najmniejszą średnią ocenę. Co do

uzupełniania ilości bramek postępujemy podobnie minimalna średnia ilość bramek straconych i maksymalna straconych.
Kolejno na podstawie danych z poprzednich lat dokonujemy predykcji: uśredniamy statystyki zespołów w tych latach jako prawdopodobne statystyki na ten sezon (średnia ważona najbardziej liczy się zeszły sezon najmniej najdawniejszy), a także uczymy algorytm dla jakich ocen danego zespołu rozkładały się jego statystyki w ostatnich latach.
Po czym obliczamy współczynniki przyrostu “mocy” ogólnej, ataku oraz obrony (na zasadzie moc ogólna z tego sezonu przez średnia moc ogólna z poprzednich lat). Dzięki temu wiemy czy klub się wzmocnił czy osłabił.
Końcowym etapem jest korekcja, a więc przeskalowanie średnich statystyk konkretnego zespołu przez wyliczone przez nas współczynniki oraz posortowanie danych wyjściowych według kolejności najpierw według ilości punktów, potem bilansu bramek, a końcowo wedle goli strzelonych. Taka końcowa tabela będzie naszymi danymi wyjściowymi

- **Test i potencjalne problemy:**
  Potencjalnym problemem będzie mała wiedza o beniaminkach, które mogą być dość mocno niedocenione, przez co nasze przewidywania odnośnie ich miejsca w tabeli mogą być przestrzelone(zaniżone). Faktem jest też, że jest to tylko i aż sport, a zatem wszystko może się zdarzyć: np. drużyna, która w zeszłym sezonie ledwo się utrzymała oraz sprzedała najlepszego piłkarz może nagle wygrać ligę (jak Leicester City w Premier League w sezonie 2015/2016). Testu dokonaliśmy dla naszej polskiej rodzimej, “ukochanej” Ekstraklasy.

4. Demo :



```
przewidywana tabela na zakończenie sezonu:
             CLUB    GZ    GC    BG   PKT
11         Rakow  56.0  32.0  24.0  65.0
4           Lech  61.0  30.0  31.0  62.0
9          Pogon  52.0  29.0  23.0  60.0
6          Legia  50.0  38.0  12.0  54.0
5        Lechia   47.0  39.0   8.0  53.0
8          Piast  42.0  35.0   7.0  51.0
10      Radomiak  41.0  45.0  -4.0  45.0
1       Gornik_Z  46.0  48.0  -2.0  44.0
16       Wisla_P  45.0  49.0  -4.0  44.0
14         Warta  35.0  36.0  -1.0  43.0
0       Cracovia  36.0  39.0  -3.0  42.0
17   Zaglebie_L   46.0  51.0  -5.0  42.0
2   Jagiellonia   41.0  51.0 -10.0  41.0
7          Miedz  39.0  50.0 -11.0  40.0
15        Widzew  39.0  51.0 -12.0  40.0
13          Stal  40.0  44.0  -4.0  39.0
3         Korona  37.0  51.0 -14.0  39.0
12         Slask  39.0  48.0  -9.0  38.0
-----------------------------------------------
```