

Maciej Kaczanowski¹

Instytut Systemów Informatycznych, Wydział Cybernetyki
Wojskowa Akademia Techniczna, Warszawa, Polska

Komputerowe wspomaganie identyfikacji talentów w sporcie **Computer based sport talent identification**

Streszczenie

W niniejszej pracy przedstawiono koncepcje systemu do komputerowego wspomaganie identyfikacji talentów w sporcie. W koncepcji założono wykorzystanie metody badania podobieństwa opartej na optymalizacji wielokryterialnej oraz nadzorowanym algorytmie klasyfikacji z obszaru uczenia maszynowego: lesie drzew decyzyjnych. Dane do budowy wzorców dyscyplin sportowych pozyskano z publikacji (Santos, Dawson, Matias i in 2014). Dane te również posłużyły do wygenerowania testowego zestawu danych sportowców do przeprowadzenia eksperymentów badawczych. Badania przeprowadzono w autorskim programie oraz w środowisku chmurowym Microsoft Azure Machine Learning Studio. Przeprowadzone eksperymenty wykazały, że analizowane metody z powodzeniem można zastosować w procesie identyfikacji talentów sportowych.

Słowa kluczowe: identyfikacja talentów sportowych, optymalizacja wielokryterialna, uczenie maszynowe, nauka o danych

Abstract

This paper presents the concept of the computer decision support system for talent identification in sport. In this concept the use of two methods was assumed: pattern recognition based on multicriteria optimization and machine learning supervised classification algorithm: decision forest. The data for sport disciplin patterns has obtained from publication (Santos, Dawson, Matias et al. 2014). This data also has been used to generate test data sets to research purposes. The researches were carried out in author's application and in the cloud environment Microsoft Azure Machine Learning Studio. The results show that both methods can be used with success to talent identification in sport.

Keywords: talent identification in sport, multicriteria optimization, machine learning, data science

¹ mgr inż. Maciej Kaczanowski, WAT, maciej.kaczanowski@wat.edu.pl

1. Wstęp

Talent powszechnie rozumiany jest jako wrodzone zdolności pozwalające wykonywać pewne czynności powyżej przeciętnego poziomu, a także ponadprzeciętny potencjał do dalszego doskonalenia się w tych czynnościach. W literaturze pojęcie talentu jest podejmowane w różnych kontekstach. Za talent można uznać zestaw wybitnych, jednokierunkowych lub wielokierunkowych uzdolnień ludzi. Klasyfikację definicji talentu można oprzeć na takich kryteriach jak: wyjątkowość uzdolnień, ich związek z celami czy rodzaj uzdolnień. Kryterium wyjątkowości opiera się na założeniu, że talent jest rzadkim dobrem, nie powszechnym, jego podaż jest zatem mała (Pocztowski 2008). Przeciwnym biegunem jest rozumienie talentu jako wysokiego poziomu kluczowych dla organizacji lub stanowiska kompetencji. W tym rozumieniu wszyscy pracownicy organizacji mogą zostać uznani za talenty, jeśli proces rekrutacji zakłada selekcję opartą na wysokich standardach klasyfikacji. Definiowanie talentu staje się tożsame z pojęciem pracownika o wysokim potencjale (Woińska i Szmidt 2005).

The Conference Board w raporcie „Integrated and Intergative Talent Management” zaproponowała następującą definicję: „Talent to każda osoba mogąca w sposób znaczący wpłynąć na obecne i przyszłe osiągnięcia swojej organizacji.” (Morton L. 2004). Pozyskiwanie do systemu szkolenia osób utalentowanych w jak najmłodszym wieku i w jak największej liczbie jest trudne, a błędy popełnione na tym etapie mogą przynieść straty czasowe i finansowe. Zarządzanie talentami według definicji T. Listwana to „zbiór działań odnoszących się do osób wybitnie uzdolnionych, podejmowanych z zamiarem ich rozwoju i sprawności oraz osiągania celów organizacji” (Listwan T. 2005). W sporcie proces ten można podzielić na cztery etapy: wykrywanie, identyfikacja, selekcja i rozwój. Osoby zaangażowane w ten proces muszą, choć w stopniu podstawowym posiadać wiedzę z obszarów takich jak: biomechanika, anatomia, fizjologia oraz psychologia.

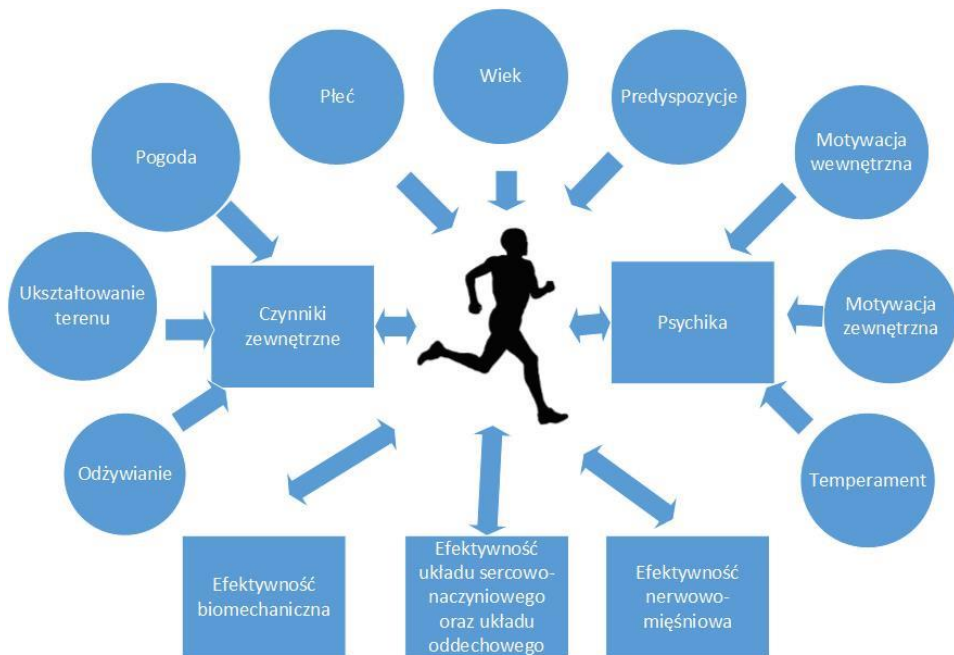
Pojęcie talentu, również w sporcie, zostało szeroko opisane w pracy Łukasza Panfila (Panfil 2012). Zostały tam przytoczone różne aspekty definicji talentu. Samo pojęcie talentu Ł. Panfil definiuje jako pewną „sprzyjającą kombinację”, która łączy w sobie w zasadzie wszystkie aspekty funkcjonowania człowieka – od genetycznych po środowiskowe. Korzystny układ ich będzie owocował talentem w takiej bądź innej dziedzinie². Autor pracy podkreśla, że, przygotowanie do uprawiania sportu na najwyższym poziomie, w odróżnieniu od innych dziedzin, zajmuje dużo więcej czasu i często rozpoczyna się już w dzieciństwie. Według Ł. Panfila w sporcie odpowiednie zdiagnozowanie talentu i później rozwijanie go jest jednym z głównych czynników późniejszego sukcesu na najwyższym poziomie. Podkreśla jak ważne jest wspieranie talentu przez: rodzinę, rówieśników czy też szkołę, a także społeczne, ekonomiczne i polityczne organizacje (Panfil 2012).

Na początku jedną z kluczowych decyzji jest wybór odpowiedniej dyscypliny. Po wyborze dyscypliny kolejnym etapem jest dobór treningu, który powinien być spersonalizowany i prowadzić do osiągnięcia jak najlepszej formy i

² <http://mlodziezowyfutbol.pl/poradnik-sukcesu-pilkarza/dla-rodzica/poradnik-mf-talent-pilce-noznej-identyfikowanie-rozwoj-cz/>

wyników sportowca. Wyniki uzyskiwane w trakcie treningów powinny być również brane pod uwagę i stanowić kolejny etap weryfikacji kandydata.

Na osiągi sportowca ma wpływ wiele czynników, część z nich została przedstawiona na rysunku 1. Wszystko to czyni proces identyfikacji i zarządzania talentami dobrym do zastosowania komputerowego systemu wspomagania decyzji.



Rysunek 1. Czynniki mające wpływ na osiągi sportowca.

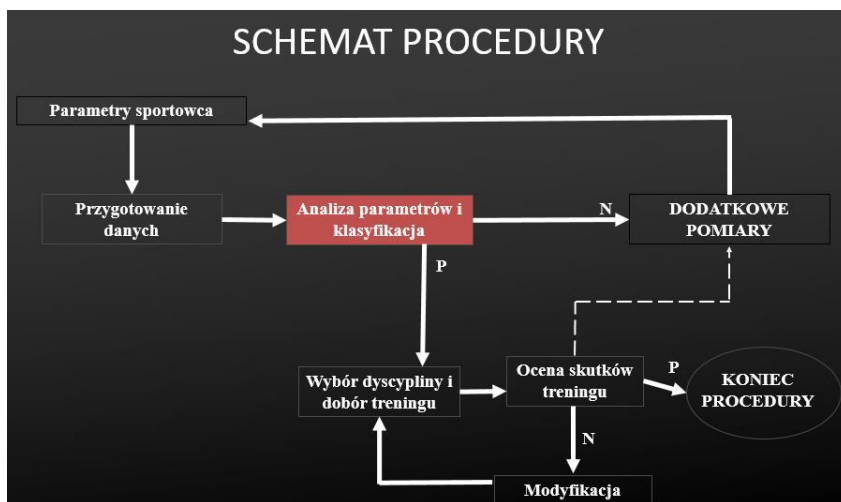
Źródło: opracowanie własne.

W dostępnych modelach identyfikacji, wyboru i rozwoju talentów w sporcie (Bar-Or's, Gimbel's, Jones & Watson's, Harre's, Bompa's) [3] w pierwszej kolejności zbiera się różne parametry, które następnie poddaje się analizie. W zależności od modelu można podzielić je na: cechy morfologiczne, fizjologiczne, psychologiczne i wydajnościowe (ang. morphological, physiological, psychological, performance). Większość parametrów można pozyskać w sposób nieinwazyjny za pomocą ogólnodostępnych sensorów i różnych testów. Prawidłowa analiza tych danych jest podstawą do dalszych czynności zdefiniowanych w poszczególnych modelach. W obszarze tym można z powodzeniem zastosować komputerowe wspomaganie podejmowania decyzji, które w znacznym stopniu przyczyni się do skrócenia całkowitego czasu potrzebnego do identyfikacji talentów i zwiększy skuteczność całego procesu.

2. Schemat procedury

Poniżej została przedstawiona propozycja procedury w trakcie, której zostaną przeanalizowane parametry zebrane od sportowców i na tej podstawie dokonany zostanie wybór odpowiedniej dyscypliny oraz dobór treningu:

- (1) Pomiar parametrów kandydata.
- (2) Przygotowanie danych do analizy. Jeden z modułów komputerowego systemu wspomaganie identyfikacji talentów sportowych powinien wspierać użytkownika w przypadkach, gdy dane są niepoprawnie sformatowane, zanieczyszczone, posiadają rekordy odstające itp.
- (3) Przetworzenie danych i poddanie ich analizie. W przypadku zbyt małej ilości wymaganych danych system powiadomi sportowca o potrzebie wykonania dodatkowych pomiarów.
- (4) Na podstawie wyników wybór dyscypliny i treningu weryfikacyjnego.
- (5) Wizualizacja wyników oraz rekomendacje dotyczące modyfikacji treningu lub zmiany dyscypliny.
- (6) Iteracyjnie wykonywana ocena skutków i odpowiednia modyfikacja rekomendacji.



Rysunek 2. Schemat procedury dla systemu do wspomaganie identyfikacji talentów w sporcie

Źródło: Opracowanie własne.

3. Metoda i organizacja przeprowadzania pomiarów

Rozwój technologii mobilnych i Internetu rzeczy (ang. internet of things) dał ogromne możliwości w świecie sportu w obszarze zbierania i analizy różnych parametrów wydajnościowych. Sportowcy mogą korzystać z zegarków sportowych, aplikacji na urządzeniach mobilnych - posiadających wbudowane sensory, liczników kadencji, pulsometrów itp.. Dokładność tych urządzeń pozwala na zebranie wiarygodnych wartości parametrów wydajnościowych potrzebnych w proponowanej procedurze. Do przeprowadzenia ich pomiarów należałoby wykorzystać gotowe lub sformułować nowe próby testowe. Testy powinny być przeprowadzane dla wszystkich osób w warunkach jednakowych czynników zewnętrznych. W przeciwnym przypadku należałoby w całej procedurze uwzględnić ich wpływ na otrzymane wyniki.

Cechy morfologiczne można mierzyć różnymi sposobami. Jednym z nich jest tzw. zestaw Martina. Posługując się taśmą antropometryczną, antropometrem, cyrklem kabłąkowym należy dokonać stosownych pomiarów, które można rozszerzyć o pomiary fałdomierzem oraz wagą. Innym sposobem, który warto zautomatyzować jest analiza fotografii kandydatów, którą można przeprowadzić za pomocą głębokich sztucznych sieci neuronowych. Następnie posługując się metodą Wiliama Sheldona w modyfikacji Heath i Cartera warto rozbudować parametry o ocenę typologiczną, która klasyfikuje ludzi pod względem budowy ciała na trzy kategorie: endomorficzny, mezomorficzny i ektomorficzny.

Liczbę cech można powiększyć przy pomocy nowoczesnych metod opisanych w [7]:

- (1) Wywiad z zawodnikiem.
- (2) Cechy psychiczne zebrane za pomocą kwestionariuszy NEO-FFI, EAS i CISS.
- (3) Pomiar zawartości wody w organizmie.
- (4) Pomiar spoczynkowej przemiany materii za pomocą ergospirometru.
- (5) Morfologia krwi.
- (6) Stężenie we krwi różnych enzymów i związków: cholesterolu, triglicerydów, glukozy, kreatyny, mocznika, kwasu moczowego, alfa-amylazy trzustkowej, hemoglobiny, potasu, fosfatazy alkalicznej, bilirubiny, kinazy kreatynowej.
- (7) Stężenie mleczanu we krwi.
- (8) Stężenie amoniaku we krwi.
- (9) Stężenie kreatyniny w moczu.
- (10) Stężenie pochodnych puryn we krwi.
- (11) Pomiar gazometrii krwi.
- (12) Pomiar objętości i szybkości przepływu powietrza metodą spirometryczną.
- (13) Wydolność tlenowa submaksymalna i maksymalna.
- (14) Temperatura powierzchni ciała.
- (15) Analiza ruchu za pomocą MyoMotion, OptimEye, BioHarness,
- (16) Aktywność bioelektryczna mięśni.
- (17) Zdolności szybkościowo-siłowe mierzone za pomocą SmartJump.

- (18) Testy koordynacyjne przeprowadzane w oparciu o czujniki świetlne np. FitLight Trainer.

4. Matematyczne modelowanie kandydatów

Każdą osobę można opisać w danym czasie za pomocą wektora cech antropometrycznych, składu budowy ciała i wydajnościowych, zawierającego takie składowe jak: wzrost, waga, suma fałdów skórnych, wskaźnik masy ciała, obwody kończyn, maksymalna wydolność tlenowa $VO_2\max$, stężenie kwasu mlekowego we krwi BLa, wzrost tętna HR, poziom tlenu we krwi, poziom nawodnienia itd.. Wektor ten uzyskuje się, przeprowadzając odpowiednie pomiary. Pozwala on określić, w jakiej formie aktualnie znajdują się dana osoba. W każdej dyscyplinie sportowej $d \in D = \{1, \dots, D\}$, bazując na wiedzy eksperckiej można utworzyć repozytorium wzorców określających pożądane wartości parametrów. Trzeba jednak pamiętać, że istotnymi cechami, które należy wziąć pod uwagę jest wiek kandydata, po którym przyporządkowuje się go do kategorii wiekowej $k \in K$ oraz płeć $s \in \{b, g\}$. Drugim ważnym założeniem jest to, że dyscypliny mogą posiadać różne zestawy parametrów potrzebnych do określenia czy dana osoba wyróżnia się na tle innych kandydatów. Biorąc powyższe pod uwagę niech będą dane:

$$N^d = \left\{ N_1^d, \dots, N_l^d, \dots, N_L^d \right\} \subset N_k^s, d \in D, k \in K, s \in \{b, g\} - \text{zbiór numerów}$$

parametrów dla danej dyscypliny sportowej, gdzie:

N_k - zbiór numerów wszystkich parametrów z repozytorium dla kategorii wiekowej $k \in K$.

$L(d)$ - liczba parametrów wydajnościowych dyscypliny sportowej $d \in D$.

Cześć z zebranych parametrów będzie miała większe znaczenie dla danej dyscypliny niż pozostałe, więc dla każdego z nich, ekspert powinien określić stopień ważności:

$$\alpha \begin{pmatrix} d \\ n_l \end{pmatrix} \in [0, 1], \quad d \in N$$

Po zebraniu parametrów kandydata x należącego do zbioru wszystkich planowanych do przebadania kandydatów X w danej kategorii wiekowej $k \in K$, możemy opisać go w następujący sposób:

$$N^x = \left\{ N_1^x, \dots, N_l^x, \dots, N_L^x \right\} - \text{zbiór numerów parametrów kandydata}$$

$x \in X_k^s, k \in K, s \in \{b, g\}$, gdzie:

L - liczba parametrów w repozytorium.

W artykule w celach demonstracyjnych ograniczono się do cech antropometrycznych oraz składu budowy ciała. Nie uwzględniono w nim również wieku kandydatów i dla wszystkich dyscyplin zastosowano ten sam zestaw cech.

Opisywane podejście w przyszłości zostanie rozbudowane o pominięte aspekty oraz planuję się dodać parametry wydajnościowe oraz psychologiczne.

5. Wzorce dyscyplin sportowych

W celu demonstracyjnym do utworzenia bazy wzorców dyscyplin wykorzystano dane referencyjne opublikowane przez autorów publikacji [2]. W pracy tej do pomiarów wzrostu i wagi wykorzystano 898 sportowców (264 kobiet i 634 mężczyzn). Parametry antropometryczne zostały zmierzone na 798 sportowcach (240 kobietach i 558 mężczyznach). Celem pracy było określenie referencyjnych wartości antropometrycznych i budowy ciała sportowców. Dane zebrano za pomocą badań densytometrycznych metodą DXA (dwuenergetyczna absorpcja rentgenowska). Służą one do określenia gęstości i mineralizacji kości, ustalenia składu ciała (zawartości tłuszczu, masy mięśniowej i składników mineralnych). Metody pomiarów składu ciała człowieka ze względu na złożoność podzielono na pięć poziomów: 1. Atomowy, 2. Molekularny, 3. Komórkowy, 4. Tkanki i narządy, 5. Całe ciało. W tym przypadku przy pomiarach skupiono się na dwóch poziomach: a) całe ciało - dokonano pomiarów antropometrycznych, b) molekularny - badania wykonane metodą densytometryczną DXA. Wyniki opublikowano w postaci tabel, w których rozróżniono płeć sportowca. W kolumnach tabeli przedstawiono percentyle 0.05, 0.25, 0.50, 0.75, 0.95, dla każdego z nich najmniejszą, estymowaną i największą wartość. W wierszach zapisano poszczególne dyscypliny. Sposób estymacji został dokładnie opisany w omawianym artykule.

Wzorce dyscyplin sportowych zbudowano w oparciu o dane z obydwu grup: „całe ciało” oraz „poziom molekularny” tylko dla mężczyzn z uwagi na brakujące dane referencyjne dla kobiet. Do wzorca każdej dyscypliny zapisano przeciętną (medianę) estymowaną wartość. Mediana (percentyl 0.50) oznacza, że wartość badanej cechy u 50% sportowców danej dyscypliny nie przekraczała tej wartości. Oryginalne dane zostały zapisane w formacie pdf, co znacznie utrudniło wykorzystanie ich w dalszych pracach. Proces wydobycia danych z pdf wymagał konwersji do docx, z którego dokonano eksportu do xlsx po czym każdy parametr zapisano w oddzielnym pliku csv.

Parametry dla wzorców z grupy „całe ciało”, które wykorzystano w demonstratorze:

BW - Waga (ang. Body weight) (kg);

H - Wzrost (ang. Height) (cm);

BMI - Wskaźnik masy ciała (ang. Body mass index) (kg/m²);

SSS - Miara fałdów skórnych w siedmiu miejscach (ang. Sum of seven skinfolds) (mm);

SAPS - Miara fałdów skórnych na kończynach (ang. Sum of appendicular skinfolds) (mm);

SAS - Miara fałdów skórnych na ramionach (ang. Sum of arm skinfolds) (mm);

SLS - Miara fałdów skórnych na nogach (ang. Sum of leg skinfolds) (mm);

STS - Miara fałdów skórnych tułowia (ang. Sum of trunk skinfolds) (mm);

AC - Obwód ramienia (ang. Arm circumference) (cm);

AMC - Obwód mięśnia ramienia (ang. Arm muscle circumference) (cm);

HC - Obwód bioder (ang. Hip circumference) (cm);

TMC - Obwód mięśnia uda (ang. Thigh muscle circumference) (cm);

CMC - Obwód mięśnia łydki (ang. Calf muscle circumference)(cm);

Tabela 1. Wzorce dyscyplin sportowych

| Dyscyplina | STS | BW | CMC | SAPS | AMC | HC | SLS | H | SAS | SSS | BMI | TMC |
|-------------------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|-------|
| Lekkoatletyka | 36.50 | 74.03 | 35.90 | 20.42 | 27.32 | 95.45 | 11.93 | 181.6 | 9.05 | 47.62 | 22.39 | 52.00 |
| Koszykówka | 59.90 | 81.76 | 36.28 | 33.94 | 27.16 | 99.29 | 20.19 | 190.3 | 14.03 | 72.65 | 22.52 | 51.08 |
| Gimnastyka | 38.34 | 66.44 | 33.49 | 27.34 | 27.12 | 92.24 | 16.12 | 169.8 | 11.87 | 59.12 | 23.00 | 46.34 |
| Korfbal | 54.28 | 72.73 | 35.62 | 27.02 | 25.93 | 93.96 | 15.83 | 180.0 | 12.31 | 62.43 | 22.32 | 47.39 |
| Pięciobój | 41.28 | 69.56 | 34.56 | 25.83 | 27.13 | 93.83 | 15.08 | 176.2 | 11.65 | 57.27 | 22.30 | 51.73 |
| Inne sporty walki | 51.47 | 70.47 | 34.45 | 30.44 | 27.10 | 94.66 | 18.89 | 175.6 | 12.00 | 62.89 | 22.77 | 50.57 |
| Wioślarstwo | 49.18 | 78.52 | 34.66 | 27.10 | 28.00 | 96.66 | 16.09 | 182.4 | 11.49 | 60.88 | 23.51 | 50.77 |
| Piłka nożna | 47.60 | 73.90 | 35.18 | 25.01 | 27.97 | 95.75 | 15.04 | 176.4 | 10.72 | 58.81 | 23.68 | 51.93 |
| Pływanie | 43.39 | 72.18 | 34.38 | 26.42 | 30.91 | 94.23 | 16.38 | 179.6 | 10.36 | 56.93 | 22.29 | 48.71 |
| Tenis | 63.98 | 71.56 | 34.13 | 31.11 | 25.93 | 93.88 | 18.27 | 177.0 | 13.52 | 67.85 | 22.75 | 47.82 |
| Triathlon | 44.78 | 66.17 | 34.20 | 23.19 | 26.38 | 90.54 | 13.98 | 175.6 | 9.61 | 50.26 | 21.44 | 47.75 |
| Siatkówka | 54.99 | 89.46 | 36.79 | 27.98 | 31.18 | 102.82 | 17.28 | 193.4 | 11.45 | 70.00 | 23.84 | 53.44 |
| Wrestling i Judo | 55.15 | 71.56 | 34.37 | 26.73 | 29.04 | 93.33 | 16.24 | 172.6 | 10.69 | 59.58 | 23.96 | 49.71 |

Źródło: Opracowanie własne na podstawie danych referencyjnych z pracy [2].

6. Badanie podobieństwa

Jedną z metod, którą można z powodzeniem zastosować w omawianym przypadku jest metoda wieloaspektowego badania podobieństwa przedstawiona w pracach [4], [5] i [6]. Prace te dotyczą modelowania stanu zdrowia pacjenta, aczkolwiek metoda w nich przedstawiona może również znaleźć zastosowanie w obszarze identyfikacji sportowych talentów. Wektor parametrów może posłużyć w systemie opartym na wieloaspektowym badaniu podobieństwa do określenia odpowiedniej dyscypliny sportowej.

Na potrzeby artykułu do badań podobieństwa pomiędzy stanem kandydata, a wzorcem dyscypliny sportowej wykorzystano modele pajęczynowe. Wyróżnia się kilka typów tych modeli: punktowe, gwiazdowe, radarowe i wiatrakowe. Ich wspólną właściwością jest to, że każda cecha ma swoją własną oś wychodzącą z centralnego punktu. W niniejszym opracowaniu zastosowano dwa typy modeli: modele radarowe i wiatrakowe. Te pierwsze powstają przez łączenie liniami wszystkich wartości tej samej serii. W drugich również wartości jednej serii łączone są liniami, ale każda wartość jest umieszczona na dwóch osiach tak, aby powstały trójkąty równoramienne przylegające do siebie bokami z jednym wierzchołkiem wspólnym. Autorska aplikacja po wprowadzeniu danych kandydata i wzorców dyscyplin sportowych wykonuje odpowiednie wykresy. Na wykresach

pajęczynowych przedstawione są: stan kandydata, wzorzec dyscypliny sportowej, które nanoszone są na jeden wspólny wykres. Pozwala to wizualnie ocenić stopień podobieństwa, na przykład, gdy kształty i wymiary wykresów są zbliżone to stopień jest wysoki. Aplikacja wylicza również charakterystyki jakościowe wykresów takie jak: promień, pole, liczba boków, długość obwodu, środek ciężkości i dla wspólnego wykresu: różnice promieni, część wspólną pól powierzchni, różnice liczby boków, różnicę długości obwodu, odległość środków ciężkości, odległość wektorów. Skala podobieństwa wykresów określona została na podstawie części wspólnej pól powierzchni oraz odległości środków ciężkości.

Dysponując tymi danymi możemy określić odległość kandydata od wzorców zdefiniowanych dla każdej dyscypliny sportowej.

Symbolami $p^*(d)$ i $s^*(d)$ oznaczymy wzorce dyscypliny sportowej nr $d \in D$ w aspekcie części wspólnej pól powierzchni i odległości środków ciężkości wykresu pajęczynowego.

Symbolami $o_1^*\left(N(x), p^*(d)\right), d \in D$ i $o_2^*\left(N(x), s^*(d)\right), d \in D$

oznaczymy odpowiednio część wspólną pól powierzchni i odległość środków ciężkości wykresów pajęczynowych kandydata i wzorca.

Następnie wykorzystując teorie optymalizacji wielokryterialnej i teorię punktu idealnego $y = \begin{pmatrix} * \\ * & * \\ y_1, y_2 \end{pmatrix}$ definiujemy zadanie w postaci:

$(W, o(d), \bar{R})$, gdzie:

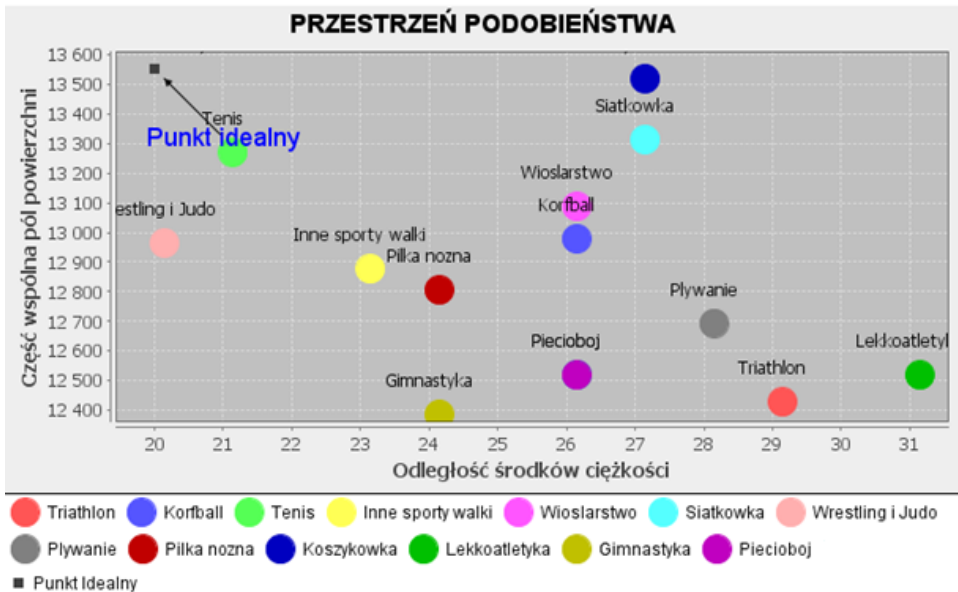
W - zbiór dyscyplin wstępnie wybranych dla kandydata;

$o(d)$ - wektorowa funkcja odległości kandydata od wzorca dyscypliny sportowej nr $d \in D$;

\bar{R} - model preferencji decydenta;

Wyznaczamy najbardziej prawdopodobną dyscyplinę i budujemy ranking dyscyplin do dalszej weryfikacji:

$$y_1 = \min o_1^*(d), \quad y_2 = \min o_2^*(d)$$



Rysunek 3. Przestrzeń podobieństwa wygenerowana przez autorską aplikację

Źródło: Opracowanie własne.

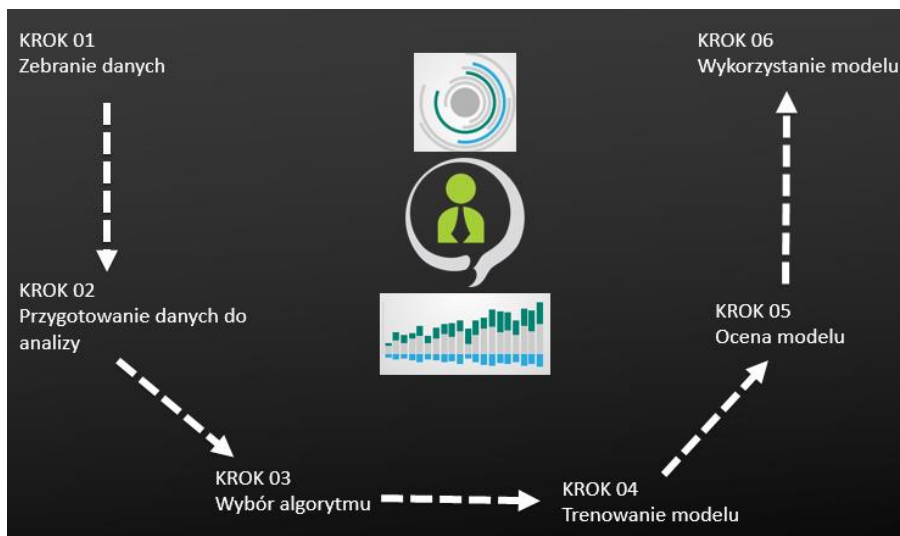
Na rysunku 3 w zbiorze pareto (najbardziej prawdopodobnych dyscyplin) znalazły się: tenis, koszykówka oraz wrestling i judo.

7. Uczenie maszynowe

Innym podejściem, które warto skonfrontować z zadaniem omawianym w niniejszej publikacji jest jeden z obszarów sztucznej inteligencji, a mianowicie uczenie maszynowe (ang. machine learning). Dziedzina ta w ostatnim czasie stała się popularna i bardzo szybko się rozwija. Główny podział metod uczenia maszynowego wyróżnia uczenie nadzorowane i nienadzorowane. Pierwszy sposób polega na tym, że system dostaje przykłady poprawnego działania, na których powinien się wzorować w swojej bieżącej pracy. Uczenie nienadzorowane (bez nauczyciela, czasem nazywane samo-uczeniem) tworzy samodzielnie model wymagany do rozwiązania zadania nie potrzebując do tego danych o pożądanym wyniku. Omówienie wszystkich algorytmów i technik wchodzących w skład tej dziedziny nawet w podstawowym zakresie wymagałoby poświęcenia setek stron, dlatego też w niniejszym opracowaniu przedstawiono problem klasyfikacji wieloklasowej, którym jest przyporządkowywanie kandydatów na podstawie ich parametrów do odpowiedniej dyscypliny sportowej. Kluczowa dla metod i technik z tego obszaru jest liczba dostępnych danych. Im więcej danych zostanie poddanych analizie, tym lepsze osiągnięte zostaną rezultaty.

Eksperyment przeprowadzono w środowisku chmurowym Microsoft Azure Machine Learning Studio przy wykorzystaniu algorytmu lasu drzew decyzyjnych

(ang. Decision Forest). Przygotowanie komponentu uczącego się do systemu komputerowego wspomagania identyfikacji sportowych talentów zostało podzielone na kilka kroków przedstawionych na rysunku 4.



Rysunek 4. Przygotowanie komponentu uczącego się do systemu komputerowego wspomagania identyfikacji talentów w sporcie

Źródło: Opracowanie własne.

- (1) Zebranie danych – pierwszy i jeden z najważniejszych procesów, który przełoży się, na jakość całego rozwiązania. Zgodnie z zasadą „śmieci na wejściu, śmieci na wyjściu” to, jakimi danymi zostanie zasilony system przełoży się na jego ogólną efektywność. Niewłaściwe dane źródłowe mogą się już na początku przeprowadzonych eksperymentów objawiać niską entropią lub brakiem korelacji między zmiennymi wejściowymi, a zmienną wyjściową.
- (2) Przygotowanie danych – przed przetworzeniem danych należy je na początku w odpowiedni sposób przygotować. W tym kroku dokonywana jest agregacja, ocena, obróbka i wzbogacanie danych.
- (3) Wybór algorytmu – zastosowany algorytm przełoży się na szybkość i poprawność wykonywanego przez system przetwarzania. Krok ten zostanie wielokrotnie powtórzony w celu porównania kilku algorytmów.
- (4) Trenowanie modelu - w tym kroku przy wykorzystaniu danych archiwalnych odpowiednio podzielonych na zbiory, dokonywana jest nauka modelu, aby w przyszłości miał on zastosowanie dla nowych danych napływających na wejście systemu.
- (5) Ocena modelu – jest to krok warunkujący zakończenie całego procesu przygotowania modelu. W typowych eksperymentach uczenia maszynowego po ocenie następuję wielokrotne powtórzenie poprzednich kroków mające na celu poprawę, jakości modelu, aż do uzyskania zadowalającego poziomu. W użytym środowisku moduł oceny dla klasyfikatorów wieloklasowych zwraca średnie

miary, jakości oraz macierz pomyłek, na której przedstawione są informacje, ile procent obiektów zostało zaklasyfikowanych do poszczególnych klas. Jedną z podstawowych miar, jakości klasyfikatorów jest trafność klasyfikacji mówiąca o tym, jaki jest stosunek liczby obiektów poprawnie zaklasyfikowanych do liczby wszystkich obiektów.

- (6) Wykorzystanie modelu – w przypadku pożądanego rezultatu najlepszy model zostanie wprowadzony do eksploatacji. Po tym kroku należy monitorować jego działanie i ewentualnie dokonywać aktualizacji przy użyciu nowych danych.

Zbiór danych na potrzeby eksperymentu wygenerowano przy użyciu języka Python. Z każdego parametru z danych referencyjnych [2] wybrano najmniejszą i największą wartość, a następnie za pomocą modułu pandas utworzono DataFrame składający się z trzech kolumn. Zestaw testowy został rozbudowany o poniższe cechy względem poprzedniego eksperymentu:

- (1) APFM - Appendicular fat-mass;
- (2) APBMC- Appendicular bone mineral content;
- (3) SBMC- Subtotal bone mineral content;
- (4) SBFMP- Subtotal fat mass%;
- (5) SBFM- Subtotal fat mass;
- (6) SBFFM- Subtotal fat-free mass;
- (7) SLST- Subtotal lean soft tissue;
- (8) WBFT- Whole body fat mass;
- (9) WBBMC- Whole-body bone mineral content;
- (10) WBBMD- Whole-body bone mineral density;
- (11) WBFFMI- Whole-body fat mass index;
- (12) WBFFM- Whole-body fat-free mass;
- (13) WBLST- Whole-body lean soft tissue;

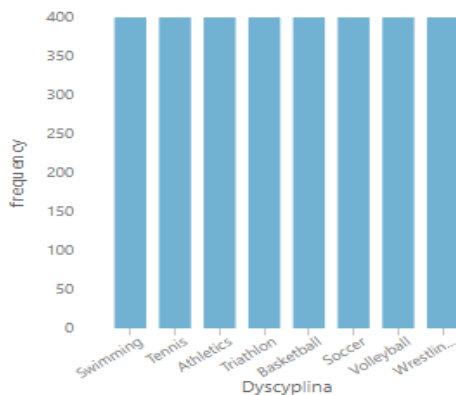
, dla następujących ośmiu dyscyplin sportowych wygenerowano 3200 przypadków, czyli po 400 dla każdej:

- (1) Lekkoatletyka (ang. Athletics);
- (2) Koszykówka (ang. Basketball);
- (3) Piłka nożna (ang. Soccer);
- (4) Pływanie (ang. Swimming);
- (5) Tenis (ang. Tennis);
- (6) Triathlon;
- (7) Siatkówka (ang. Volleyball);
- (8) Wrestling and Judo;

ID,Dyscyplina,STS,BN,CNC,SAPS,AKC,HC,SL5,H,SAS,SSS,BMI,THC,nBLST,SBFHM,APBMC,nBBND,nBMC,SBFH,nBFT,nBFFH,APFH,nBFFHM,SLST,nSBK,SBFHP
 1,Swimming,82.55,77.58,38.6,45.53,8.47,102.79,3.13,164.45,2.98,44.7,26.4,54.3,61.07,54.57,1094.47,1.11,1714.4,12.66,6.46,75.25,5.14,23.2,44.52,1985.91,5.83
 2,Swimming,10.33,66.74,30.09,28.24,13.74,93.56,26.49,180.14,6.96,87.5,19.96,44.84,41.59,30.57,1243.33,1.42,3517.21,3.56,12.34,44.64,4.95,20.36,35.55,05,2191.16,5.33
 3,Swimming,57.3,53.97,35.03,18.53,34.87,95.29,27.69,184.34,4.72,34.88,18.87,52.12,70.61,54.06,1376.85,1.24,2208.0,11.83,20.24,27.98,2.86,17.44,41.17,1732.31,15.29
 4,Swimming,46.78,72.24,38.65,18.95,52.55,89.76,17.95,163.64,14.29,103.65,24.71,46.62,55.14,78.63,772.03,1.44,2823.07,10.6,6.13,99.23,66.6,5.22,22.09,67.01,2130.87,20.27
 5,Swimming,17.78,61.31,31.17,33.0,38.62,92.82,25.95,180.63,18.18,47.89,21.16,42.18,65.5,65.64,1460.25,0.97,1862.67,11.86,9.92,48.27,2.97,18.47,39.34,1203.14,5.58
 6,Swimming,12.16,57.2,39.32,8.5,11.27,83.2,15.33,182.27,10.73,19.15,19.48,39.06,41.4,49.28,732.03,1.34,1833.59,9.61,9.81,61.46,6.91,17.25,63.97,1808.46,2.9
 7,Swimming,43.66,84.4,30.55,9.55,39.61,66.81,30.56,181.04,11.89,97.15,25.13,59.23,47.38,40.79,1634.11,0.98,2683.82,4.84,7.87,25.75,5.54,21.75,44.81,1823.76,20.7
 8,Swimming,45.05,80.19,35.09,44.91,35.15,88.2,3.5,194.26,5.05,29.19,18.18,44.2,51.22,77.32,1671.29,1.45,1724.49,5.66,5.55,21.96,6.39,17.93,53.98,1562.2,17.4
 9,Swimming,54.03,55.61,32.7,35.02,36.42,98.45,6.92,179.55,9.45,79.67,26.25,49.03,46.48,58.62,853.83,1.35,1853.36,7.69,10.96,79.05,3.55,17.51,54.73,1540.9,4.32
 10,Swimming,44.68,80.36,55.10,49.7,86.100,27,20,42,186,2,16,35,70,04,23,28,39,57,60,77,61,42,1782,53,1,32,323,57,9,87,10,0,54,64,7,75,21,22,72,11,1448,19,20,56
 11,Swimming,72.46,79.74,30.24,36.3,9.65,97.13,23.63,192.15,14.26,46.59,20.15,41.88,72.16,78.48,1165.13,1.07,2707.88,14.88,11.02,33.84,8.32,15.82,71.24,1708.84,20.87
 12,Swimming,22.02,51.5,90.52,22.88,15.57,102.01,13.39,181.27,12.3,71.68,18.0,46.84,63.26,77.36,1354.64,1.09,2088.53,11.52,9.4,41.57,8.53,16.62,41.7,1880.1,18.95
 13,Swimming,72.48,54.11,30.47,10.68,8.34,95.76,2.08,166.11,12.99,99.78,24.35,48.76,44.88,53.11,992.03,1.01,2072.18,12.75,16.82,80.0,9.21,19.1,63.62,2045.39,9.01
 14,Swimming,24.52,60.22,31.56,37.02,6.7,92.5,23.77,184.92,6.83,31.67,24.01,51.11,65.2,63.21,1986.17,0.95,3066.13,18.09,8.53,44.16,5.21,23.58,66.93,2115.58,6.02
 15,Swimming,4.46,89.47,31.41,29.13,50.6,84.88,14.01,198.27,14.64,16.46,22.83,44.22,41.89,50.43,736.45,1.18,1854.52,0.81,16.1,25.84,4.96,19.46,66.25,2077.89,16.87
 16,Swimming,42.56,89.42,38.1,27.04,10.42,85.36,12.89,184.35,8.54,39.87,22.1,40.69,66.14,43.15,1935.82,0.94,2082.21,18.06,3.77,20.41,6.79,23.77,67.86,1807.45,3.78
 17,Swimming,63.65,90.83,38.05,27.42,37.43,93.66,30.37,164.19,16.18,93.59,23.16,49.71,44.88,57.98,1468.93,1.03,2039.02,18.23,8.41,74.53,9.13,18.41,62.95,1542.44,20.2
 18,Swimming,37.46,61.77,30.31,22.13,40.07,96.51,17.51,166.33,3.96,54.83,22.5,48.54,69.74,47.9,1310.51,1.32,1699.31,8.5,7.3,70.87,8.47,17.77,42.2,1582.81,21.44
 19,Swimming,21.38,86.92,30.08,42.67,39.65,95.98,12.18,163.98,6.37,15.59,18.32,44.22,67.94,67.2,1311.89,1.46,2750.58,14.64,10.31,82.99,7.74,21.71,42.79,1708.06,20.2
 20,Swimming,80.4,76.84,30.89,46.93,4.52,89.75,9.6,167.86,5.08,28.19,23.13,46.41,54.58,53.89,687.6,0.97,2562.07,18.51,12.44,76.86,3.89,17.96,71.83,2185.0,20.25
 21,Swimming,42.3,88.15,31.29,26.37,89.93,64.10,33,185.18,12.6,45.14,18.67,46.04,45.95,70.28,750.19,0.95,3354.32,8.59,5.98,36.37,7.13,23.3,72.25,2199.91,14.92
 22,Swimming,34.82,71.85,31.37,38.2,43.04,101.92,16.12,187.55,8.54,79.13,26.33,44.34,45.31,74.76,1929.18,1.06,3393.84,16.1,1.5,95.70,57.4,8.88,23.6,62.35,2095.4,4.28
 23,Swimming,46.27,92.46,31.05,36.13,28.02,103.49,12.91,185.93,12.79,62.45,21.33,56.86,42.18,50.68,1429.63,1.16,3089.86,14.82,4.41,79.52,7.95,16.87,62.98,1928.97,20.52
 24,Swimming,49.4,86.83,30.38,22.52,50.14,99.04,2.61,189.23,11.82,83.55,25.74,30.76,49.26,67.43,994.75,1.17,2810.73,8.67,14.96,58.12,3.76,16.36,43.85,1729.24,15.67
 25,Swimming,16.34,92.34,32.54,33.78,41.11,99.71,14.87,186.26,8.72,102.24,25.74,40.66,63.61,46.33,1697.66,1.42,1884.68,13.16,7.83,85.04,3.29,22.14,57.98,1349.3,12.91
 26,Swimming,8.94,65.43,37.49,24.83,10.7,9.99,42,22.69,179.16,11.0,29.79,26.31,47.43,62.67,67.68,1821.56,1.02,1980.84,18.97,13.95,51.18,8.32,18.39,43.33,1720.32,18.96
 27,Swimming,14.85,93.53,35.64,35.34,54.25,45,88,67,21.11,183.36,13.15,79.8,22.77,42.25,57.59,55.47,1637.59,15.56,10.8,55.04,2.36,20.07,64.37,1831.99,3.35
 28,Swimming,35.85,82.54,31.54,54.12,64.27,77,100.7,24.73,174.88,3.17,11.47,25.5,51.32,60.59,62.32,997.92,1.16,2213.8,5.94,5.9,69.73,4.59,20.65,47.16,1205.35,12.07
 29,Swimming,59.11,84.71,36.16,40.46,63.09,85.21,18,18,192.56,2.81,8.58,24.83,45.64,70.98,71.35,1894.41,1.09,2689.04,3.79,4.97,82.93,2.76,16.37,46.45,1571.55,18.46
 30,Swimming,59.01,51.27,35.36,40.01,62.11,100.17,26.21,166.01,13.18,87.65,26.43,50.71,51.13,45.12,765.18,1.14,3262.86,10.73,4.09,21.67,2.09,17.7,63.26,1074.39,2.86

Rysunek 5. 30 przypadków dla dyscypliny pływanie ze zbioru danych testowych

Źródło: Opracowanie własne.



Rysunek 6. Histogram przedstawiający liczbę przypadków dla każdej dyscypliny

Źródło: Histogram wygenerowany w Azure Machine Learning Studio na podstawie zbioru danych testowych.

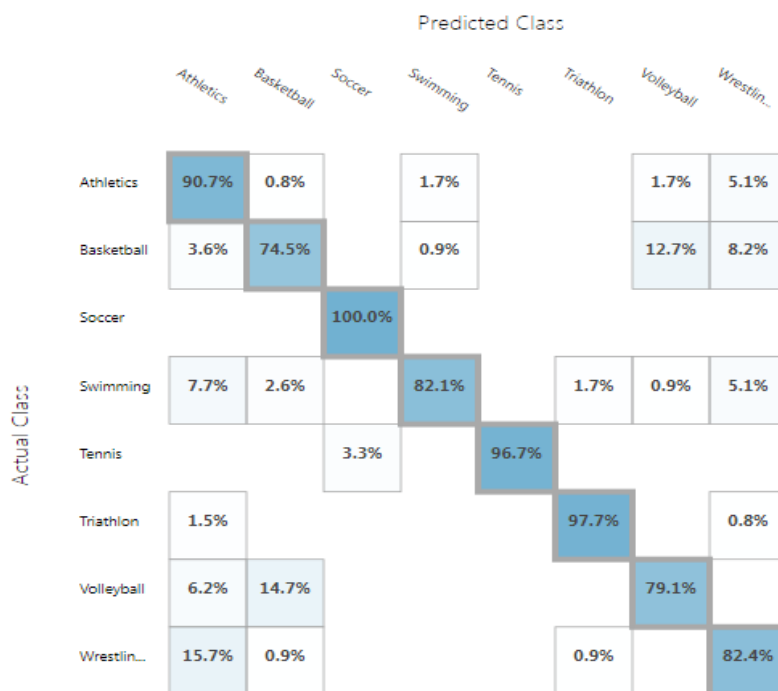
Metrics

| | |
|--------------------------|----------|
| Overall accuracy | 0.882292 |
| Average accuracy | 0.973843 |
| Micro-averaged precision | 0.882292 |
| Macro-averaged precision | NaN |
| Micro-averaged recall | 0.882292 |
| Macro-averaged recall | NaN |

Rysunek 7. Wskaźniki jakości modelu

Źródło: Wynik eksperymentu przeprowadzonego w Azure Machine Learning Studio.

Confusion Matrix



Rysunek 8. Macierz pomyłek

Źródło: Wynik eksperymentu przeprowadzonego w Azure Machine Learning Studio.

Przeprowadzono serię eksperymentów w trakcie, których zmieniano parametry algorytmu. Najlepszy uzyskany wynik dla algorytmu drzew decyzyjnych to ogólna dokładność na poziomie 0.88 (liczba prawidłowo zaklasyfikowanych przypadków / liczba wszystkich przypadków). Na macierzy pomyłek widać, że model najlepiej poradził sobie z klasyfikacją przypadków dla dyscypliny piłka nożna (100% dobrze zaklasyfikowanych), a najgorzej dla koszykówki (74,5%), gdzie 12.7% przypadków zostało zaklasyfikowanych, jako siatkówka.

8. Podsumowanie

Obydwie metody przedstawione w niniejszym opracowaniu dają zachęcające rezultaty do dalszej analizy. Przedstawiony schemat procedury może zostać wykorzystany do implementacji komputerowego systemu wspomaganie identyfikacji talentów sportowych (KSWITS). Implementacja ta połączona z ogólnokrajową bazą wzorców parametrów dla każdej dyscypliny mogłaby znaleźć zastosowanie w szkołach, sekcjach i klubach sportowych. Budowa ogólnokrajowego

KSWITS może stać się kontynuacją działań Instytutu Sportu i Ministerstwa Sportu i Turystyki, które udostępniło innowacyjne elektroniczne narzędzie „Narodową Bazę Talentów”. Za pomocą tego narzędzia gromadzone są dane dotyczące sprawności fizycznej dzieci i młodzieży.

Literatura

- [1] Vučetić, V., Babić V., Šentija D., Nekić B. (2015), “*Anthropometric and morphological characteristics of runners.*”, University of Zagreb, Croatia.
- [2] Santos D. A., Dawson J. A., Matias C. N., Rocha P. M., Minderico C. S., Allison D. B., Sardinha L. B., Silva A. M. (2014), “*Reference Values for Body Composition and Anthropometric Measurements in Athletes.*”, PLOS ONE, Retrieved from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0097846>.
- [3] Hugo K. (2004), “*A model for talent identification and development for team sports in South Africa.*”, Dissertation presented for the degree of Doctor in Sport Science at the University of Stellenbosch.
- [4] Ameljańczyk A. (2009), “*Matematyczne aspekty modelowania pajęczynowego obiektów.*”, Biuletyn Instytutu Systemów Informatycznych, WAT, Warszawa.
- [5] Ameljańczyk A. (2009), “*Wielokryterialne mechanizmy wspomagania podejmowania decyzji medycznych w modelu repozytorium w oparciu o wzorce.*”, Biuletyn Instytutu Systemów Informatycznych, WAT, Warszawa.
- [6] Ameljańczyk A. (2010), “*Model formalny informatycznego komponentu wspomagania decyzyjnego ustalania wstępnej diagnozy medycznej.*”, Biuletyn Instytutu Systemów Informatycznych, WAT, Warszawa.
- [7] Kusy K., Zieliński J. (2017), “*Nowoczesne metody diagnostyczne w sporcie. Przewodnik Trenera.*”, AWF, Poznań.
- [8] Wońska M., Szmidt C., (2005), “*Metody i techniki wylaniania kadry o wysokim potencjale (wnioski z praktyki)*”, [w:] S. Borkowska (red.), Zarządzanie talentami, IPISS, Warszawa.: 71.
- [9] Poczowski A. (2008) “*Zarządzanie talentami w organizacji*”, Wyd. Wolters Kluwer, Warszawa. 38-40.
- [10] Panfil Ł. (2012), “*Model wpierania rozwoju talentów sportowych w procesie zarządzania nimi – badania pilotażowe*”, prace naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Sukces w zarządzaniu kadrami. Elastyczność w zarządzaniu kapitałem ludzkim Problemy zarządczo-ekonomiczne. ISSN 1899-3192, str. 327- 336.
- [11] Morton L. (2004) “*Integrated and Integrative Talent Management: A Strategic HR Framework*”, Conference Board.
- [12] Listwan T. (2005) , “*Zarządzanie talentami – wyzwanie współczesnych organizacji*”.

Strony internetowe

- [13] <https://pl.wikipedia.org>
- [14] <http://mlodziejowyfutbol.pl/poradnik-sukcesu-pilkarza/dla-rodzica/poradnik-mf-talent-pilce-noznej-identyfikowanie-rozwoj-cz>