

# Analiza wydolności sportowców na podstawie krzywych kwasu mlekowego

Jakub Nieśmiał<sup>a,\*</sup>, Hubert Król<sup>b</sup> and Krzysztof Malisak<sup>c</sup>

<sup>a</sup>Wydział Elektrotechniki, Automatyki i Informatyki, Politechnika Opolska, Opole, Polska (inżynier)

<sup>b</sup>Wydział Elektrotechniki, Automatyki i Informatyki, Politechnika Opolska, Opole, Polska (inżynier)

<sup>c</sup>Wydział Elektrotechniki, Automatyki i Informatyki, Politechnika Opolska, Opole, Polska (inżynier)

## ARTICLE INFO

### Keywords:

wydolność  
sportowcy  
krzywe kwasu mlekowego  
uczenie maszynowe  
Politechnika Opolska

## ABSTRACT

W pracy przedstawiono eksperyment badawczy z wieloma testami wydolności sportowców. Dane oraz modele predykcyjne oparto na zestawie cech fizjologicznych i krzywych kwasu mlekowego, co umożliwiło ocenę i personalizację obciążeń treningowych. Zbiór danych poddano analizie pod kątem wykrywania anomalii, wykorzystując algorytm *Isolation Forest* (1). Otrzymane wyniki ukazują potencjał metody w wykrywaniu nietypowych przypadków, mogących wskazywać na błędy pomiarowe, skrajne wyniki bądź inne nieprawidłowości.

## 1. Wstęp

Podczas przeprowadzonego eksperymentu badawczego zrealizowano szereg testów wydolnościowych dla różnych sportowców. Po wstępnym przetworzeniu (usunięciu rekordów z brakami danych i wartości odstających) uzyskano zbiór 183 kompletnych raportów pomiarowych. Dodatkowo, przed każdym testem przeprowadzano wywiad z zawodnikiem, aby uzyskać szczegółowe informacje dodatkowe na temat jego kondycji, aktualnego stanu zdrowia i specyfiki dyscypliny sportowej.

Na podstawie zebranych i zmierzonych danych utworzono zestaw, w którym wyróżniono następujące cechy predykcyjne: wiek, wzrost, waga,  $VO_{2max}$ , VE, R, HRmax, RF,  $VO_{2max\_1\_m}$ . Zbiór danych rozszerzono także o cechy odzwierciedlające poziom kwasu mlekowego oraz tętno przy różnych obciążeniach (6, 8 i 10 km/h), a także charakterystyczne punkty występujące na krzywej kwasu mlekowego:

- $la_{min}$  – najniższy zarejestrowany poziom kwasu mlekowego,
- $hr_{min\_la}$  – wartość tętna odpowiadająca najniższemu poziomowi kwasu mlekowego,
- $la_{max}$  – najwyższa zarejestrowana wartość poziomu kwasu mlekowego,
- $hr_{max\_la}$  – wartość tętna odpowiadająca najwyższej wartości kwasu mlekowego,
- $la_{max\_speed}$  – obciążenie (prędkość), przy którym pojawiła się najwyższa wartość kwasu mlekowego.

W związku z faktem, że poszczególni sportowcy osiągalni różnicowane maksymalne obciążenia w testach (6–22 km/h) i nie zawsze było możliwe uzyskanie pełnego zakresu pomiarów, dodanie powyższych cech okazało się istotne dla zachowania użyteczności danych. W dalszych etapach

przetwarzania i analizy szczególną uwagę zwracano na kontekst poszczególnych zawodników (m.in. ich stan nawodnienia, porę dnia testu i inne indywidualne uwarunkowania).

Zgodnie z przeglądem literatury w obszarze detekcji anomalii (2; 5), poprawne zdiagnozowanie punktów odstających wymaga dostosowania narzędzi do charakteru analizowanego problemu. Przy zróżnicowanym zbiorze danych – zarówno pod kątem dyscyplin sportowych, jak i poziomu wytrenowania – uzasadnione jest zastosowanie metod odpornych na różne typy wariancji i nieciągłości w danych.

## 2. Cel i zakres

Głównym celem niniejszej pracy było zastosowanie algorytmu **Isolation Forest** do analizy danych w celu identyfikacji anomalii, mogących wskazywać na nietypowe lub niepożądane zachowania w zbiorze. Istotnym zadaniem była ocena, czy wychwycone obserwacje mogą sygnalizować błędy pomiarowe lub nietypową reakcję organizmu (3; 4).

### 2.1. Identyfikacja anomalii

Pierwszym etapem było zidentyfikowanie odstających obserwacji z wykorzystaniem algorytmu *Isolation Forest*. Poprzez analizę wskaźników takich jak liczba wykrytych anomalii i ich rozkład w przestrzeni cech, można było określić, które dane odbiegały od dominującego trendu.

### 2.2. Analiza efektywności modelu

Drugim ważnym celem była ocena skuteczności modelu **Isolation Forest** w wykrywaniu anomalii. Ze względu na brak etykiet prawdziwych (ground truth), skupiono się na analizie proporcji wykrytych anomalii, ich rozkładu w przestrzeni wybranych cech oraz na wizualizacji wyników.

### 2.3. Zastosowania w praktyce

Wyniki analizy anomalii mogą znaleźć zastosowanie w monitorowaniu zdrowia i wydolności sportowców. Wykrycie potencjalnych nieprawidłowości na wczesnym etapie może

\*Autor do korespondencji: j.niesmiala@po.edu.pl  
ORCID(s):

pomóc w zapobieganiu kontuzjom, dostosowywaniu intensywności treningów czy też ulepszaniu procedur pomiarowych.

### 3. Zbiór danych (Dataset)

Analizowany zbiór danych obejmował łącznie 183 próbki (pomiarów) uzyskane od sportowców trenujących różne dyscypliny. Każdy rekord składał się z następujących elementów:

- **Dane osobowe:**
  - Płeć (male, female),
  - Wiek (wartość całkowita),
  - Wzrost (w cm),
  - Waga (w kg).
- **Dyscyplina sportowa:**
  - Rodzaj uprawianej dyscypliny (np. biegacz, triathlonista, koszykarz).
- **Wyniki testów wydolnościowych:**
  - AeT (Aerobic Threshold) – próg tlenowy.
  - AnT (Anaerobic Threshold) – próg beztlenowy.
  - VO2max – maksymalna pojemność tlenowa.
  - VO2 at AnT – VO2 przy progu beztlenowym.
  - ve – objętość oddechu w ciągu minuty.
  - hrmax – maksymalne tętno.
- **Pomiar reakcji organizmu w trakcie wysiłku:**
  - Tętno (hr) przy różnych intensywnościach (np. 6, 8, 10, 12 minut testu).
  - Poziom kwasu mlekowego (la) przy różnej intensywności wysiłku.

Tak zorganizowany zbiór danych okazał się bogatym źródłem informacji, pozwalającym na analizę efektywności treningowej, ocenę wydolności w zależności od dyscypliny oraz planowanie optymalnych obciążeń treningowych.

### 4. Algorytm Isolation Forest

Metoda **Isolation Forest**, opisana szczegółowo w pracy (1), opiera się na idei izolowania nietypowych punktów poprzez losowe partycjonowanie przestrzeni cech z użyciem drzew decyzyjnych. Każdy punkt danych jest szybko „odcinany” od pozostałych, a miarą anomalii jest łatwość (lub szybkość) izolacji danej obserwacji.

#### 4.1. Parametry modelu

- **Liczba estymatorów ( $n\_estimators$ ):** 500 drzew decyzyjnych, co sprzyja stabilności wyników.
- **Contamination:** 10,3% – określa przybliżony odsetek anomalii w zbiorze.
- **random\_state:** Ustawiony na stałą wartość (2137) dla powtarzalności wyników.

Po wytrenowaniu modelu, każdej obserwacji przypisywany jest wskaźnik anomalii, który umożliwia określenie, czy dana próbka jest „odstająca” w stosunku do pozostałych.

### 5. Zastosowane metryki oceny

Ze względu na brak dostępnych etykiet prawdziwych (ground truth), ocena modelu Isolation Forest została przeprowadzona w sposób przybliżony, z wykorzystaniem następujących metod:

- **Analiza proporcji anomalii:** Porównano liczbę wykrytych anomalii z ogólną liczebnością zbioru, uzyskując odsetek odstających próbek.
- **Wizualizacja rozkładu:** Zweryfikowano rozkład anomalii i obserwacji normalnych w przestrzeni wybranych cech (np. VO2max oraz HRmax), aby sprawdzić, czy anomalie rzeczywiście odbiegają od trendu.

Takie podejście pozwoliło na pogładową ocenę skuteczności algorytmu w kontekście dostępnych danych, choć bez precyzyjnej walidacji względem obiektywnych etykiet (2).

### 6. Zastosowanie i implementacja algorytmu

**Isolation Forest** zaimplementowano w języku Python z użyciem biblioteki *scikit-learn*. Proces przygotowania danych obejmował:

- Usunięcie braków danych i anomalii ewidentnych (np. błędnych pomiarów).
- Standaryzację (normalizację) poszczególnych cech.
- Uruchomienie modelu z zadanymi parametrami ( $n\_estimators=500$ ,  $contamination=0.103$ ,  $random\_state=2137$ ).

Po zastosowaniu modelu, próbki sklasyfikowane jako odstające mogą stanowić przypadki szczególne, wymagające dalszej analizy, np. w celu wyjaśnienia, czy wynikają z nietypowej formy zawodnika, błędu pomiarowego, czy też wyjątkowych warunków zewnętrznych.

### 7. Wyniki

Po przetworzeniu 183 próbek metodą **Isolation Forest** otrzymano następujące obserwacje:

- **Liczba wykrytych anomalii:** 19 próbek (ok. 10,38%).

- **Liczba próbek uznanych za normalne:** 164 próbek (ok. 89,62%).

Otrzymane wyniki wskazują, że **Isolation Forest** wyłonił niewielką, lecz znaczącą populację obserwacji odstających. Niski procent anomalii potwierdza, że większość danych cechuje się spójnością, a ewentualne odchylenia mogą dotyczyć pojedynczych zawodników lub specyficznych warunków pomiarowych.

## 8. Podsumowanie i wnioski

W przeprowadzonej analizie z wykorzystaniem metody **Isolation Forest** udało się wyodrębnić około 10% anomalii wśród 183 przebadanych próbek. Choć nie posiadano etykiet prawdziwych (ground truth), procentowy udział anomalii wraz z ich rozkładem w przestrzeni wybranych cech sugeruje, że algorytm skutecznie identyfikuje potencjalnie nietypowe wyniki.

### Najważniejsze spostrzeżenia:

- **Możliwość wykrycia błędów pomiarowych:** Niewielka część obserwacji mogła odbiegać od normy z powodu problemów z aparaturą lub nieprawidłowego zapisu wyników (3).
- **Identyfikacja sportowców „odstających”:** U części zawodników nietypowe wartości VO<sub>2</sub>max, tętna maksymalnego czy poziomu kwasu mlekowego mogą wynikać z fenotypowych predyspozycji bądź specyfiki treningu.
- **Wartość w kontekście zdrowia i prewencji kontuzji:** Wczesne wyłapanie nietypowych wskaźników może być pomocne w szybkiej reakcji trenerskiej lub medycznej.

W przyszłości warto rozważyć:

- Rozszerzenie zbioru o dodatkowe pomiary i zmienne kontekstowe (m.in. informacje o nawodnieniu, porze dnia, diecie).
- Porównanie skuteczności wykrywania anomalii przy użyciu innych metod (np. *Local Outlier Factor*, *One-Class SVM*), co zostało szerzej opisane w literaturze (2; 5).
- Głębszą analizę biologiczną i zdrowotną wykrytych anomalii, aby trafniej interpretować ich przyczynę.

## 9. Bibliografia

### References

- [1] Liu, F. T., Ting, K. M., & Zhou, Z. H. (2012). Isolation Forest. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2012)* (pp. 413–422). IEEE. <https://doi.org/10.1109/ICDM.2012.60>
- [2] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1–58. <https://doi.org/10.1145/1541880.1541882>
- [3] Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers*. SAGE Publications.
- [4] Aggarwal, C. C. (2017). *Outlier Analysis*. Springer. <https://doi.org/10.1007/978-3-319-51656-7>
- [5] Boudahri, A., & Dahmani, H. (2021). A Comprehensive Survey on Anomaly Detection Methods in Data Mining. *Procedia Computer Science*, 179, 361–368. <https://doi.org/10.1016/j.procs.2021.01.051>