



Dokumentacja projektowa

INŻYNIERIA BIOMEDYCZNA

Jan Dylik

Informatyka

Wydział Matematyki Stosowanej

Rok akademicki 2021/2022

Tytuł:

Analiza danych antropometrycznych w celu określenia wpływu otyłości lub nadwagi na
mięśnie kończyn dolnych

Wstęp

Procent populacji ludzkiej znajdujący się w stanie nadwagi lub otyłości we współczesnych nam czasach na skutek najłatwiejszego w historii ludzkości dostępu do wysokokalorycznych produktów i siedzącego stylu życia zwiększył się do rekordowych rozmiarów. Kiedyś widok otyłej osoby na ulicy przyciągał wzrok, teraz jest to właściwie codzienność. Można z tego wywnioskować, że takie osoby, pomimo licznych problemów koordynacyjnych, starają się normalnie żyć. A więc chodzić do pracy, szkoły, uczestniczyć w życiu społecznym... Część z tych czynności osoba z nadmiarowymi kilogramami jest w stanie sobie uprościć (dla przykładu udając się do pracy samochodem, a nie pieszo), niemniej większość z nich będzie zmuszona wykonywać w niezmienny sposób. W ten sposób angażuje swoje ciało, a konkretnie mięśnie w wysiłek. Część z nich będzie pracować mocniej, część słabiej. Bezsprzecznie jednymi z regularnie obciążanych mięśni są u człowieka te położone w nogach – bez nich trudno byłoby się przemieszczać. U osoby otyłej takie mięśnie muszą wykonać nieproporcjonalnie większą pracę niż u osoby z wagą w normie – udźwignięcie często gigantycznego ciężaru całej górnej części ciała to nie lada wyzwanie. Jednocześnie jest to potężna stymulacja dla rozwoju tych mięśni – przecież ich odpowiednia moc i siła to być albo nie być dla możliwości ruchu takiej osoby. W mojej analizie spróbuję dowieść, że istotnie zwiększona waga i rozmiary człowieka to bodźce powodujące zwiększenie masy mięśniowej w stosunku do tej, którą posiada osoba o wadze w normie.

Analiza zagadnienia

Aby poprawnie zanalizować problem, trzeba ostrożnie wyselekcjonować niezbędne informacje. Wzięcie pod uwagi całościowej wagi nogi nie ma sensu, ponieważ jest to jedna z części ciała, w której najczęściej odkłada się tłuszcz [1][2]. W związku z tym użyję miary, która ignoruje całą masę tłuszczową w danym miejscu, a skupia się na pozostałych elementach – **Fat-Free Mass (FFM)**. Bierze ona pod uwagę takie aspekty, jak właśnie masę mięśniową, wodę w organizmie czy choćby tkankę kostną. Główną przeszkodą wydawać by się mogły kości, jednak jak się okazuje, nie zwiększają one swojej masy u osób otyłych [3]. Z tego powodu można wysunąć tezę, że gdy ta miara będzie wynosić więcej niż powinna, będzie to oznaczać większy poziom masy mięśniowej.

Wspomniany wskaźnik FFM będę rozważał dla obu kończyn osobno aby upewnić się, że rezultaty nie różnią się znacząco.

Rozwiązanie polega na zastosowaniu algorytmów klasyfikacyjnych uczenia maszynowego, które operując na dostępnych danych antropometrycznych będą w stanie przewidzieć zwiększony poziom wskaźników FFM, a co za tym idzie wskazać osobnika otyłego. Uprzednio należałoby jednak udowodnić, że podwyższone wskaźniki FFM korelują z otyłością, czyli wyższymi parametrami antropometrycznymi.

Dane antropometryczne, na których będę polegał, to zbiór różnorodnych pomiarów poszczególnych części ludzkiego ciała. Z pozoru większość wydaje się być przydatna w kontekście określenia otyłości, jednak jak się okazuje, niektóre z nich są bardziej, a drugie mniej pomocne.

Dlatego postanowiłem przeprowadzić oddzielne serie obliczeń dla dwóch zbiorów danych: pierwszy to komplet wszystkich danych, drugi natomiast to pewne cechy mające w moim mniemaniu kluczowy wpływ na odpowiedź na pytanie, czy dana osoba jest otyła czy też nie. Zaliczają się do nich:

- obwód pasa,
- obwód uda,
- obwód bioder,
- fałd na brzuchu,
- ciśnienie krwi skurczowe,
- ciśnienie krwi rozkurczowe,
- tętno krwi.

Jak wiadomo, otyłe osoby są narażone na nadciśnienie, jak i na wyższe tętno krwi [4]. Rejon brzucha to miejsce, gdzie tłuszcz gromadzi się najmocniej w organizmie. Natomiast zmianie przy otyłości z pewnością podlegają powyższe trzy obwody, szczególnie pasa czy bioder. Dowodem może być specjalnie stworzony wskaźnik **Waist-to-Hip Ratio (WHR)**[5]. Jest on obliczany, jak sama nazwa wskazuje, przy pomocy obwodów pasa i bioder, a otrzymany stosunek, gdy nie mieści się w dopuszczalnych ramach, wskazuje na otyłość. Bierze się to stąd, że u normalnie zbudowanego człowieka biodra są wyraźnie szersze niż pas, natomiast u otyłej osoby ta różnica znacznie się zmniejsza.

Cechy, które można odrzucić jako wpływające w mniejszym stopniu lub wcale to np. wzrost badanego czy obwód klatki piersiowej – nie wzrasta on aż tak wyraźnie u człowieka z nadwagą.

Tak przygotowane dane wykorzystam do klasyfikacji czterema algorytmami:

- K najbliższych sąsiadów,
- drzewem decyzyjnym,
- maszyną wektorów nośnych (SVM) o jądrze liniowym,
- maszyną wektorów nośnych (SVM) o jądrze wielomianowym.

Otrzymane wyniki porównam zwracając uwagę na dwa wskaźniki: dokładność i czułość. Czułość jest tu dosyć istotna, ponieważ zakładając, że pracuję na danych medycznych, a sama analiza ma również taki charakter i może służyć potencjalnie do celów diagnostycznych osób z otyłością, ostatnim czego chcę, to pominiecie jakiejś osoby z otyłością przez klasyfikator. W tym celu można poświęcić nawet błędne zdiagnozowanie jakiejś osoby jako otyłej, ponieważ skutki mogą być dalece mniej poważne. Z tego powodu jeżeli osoba o wyższym poziomie mięśni w udach zostanie oznaczona jako pozytywna, to zmaksymalizowanie czułości spowoduje zmaksymalizowanie wykrywalności właśnie takich przypadków, ponieważ czułość to stosunek osób określonych jako pozytywne przez model, które faktycznie pozytywne są do wszystkich osób faktycznie pozytywnych.

Specyfikacja wewnętrzna

Całość projektu znajduje się w jednym pliku R („Projekt.R”), a dane pobierane są z dwóch plików Excel („dane_antropo_13”, „dane_inbo”).

Początkowym etapem jest czyszczenie danych, które składa się głównie z zawężenia danych do kilku potrzebnych kolumn.

Kolumna DataBadania wymagała usystematyzowania, jako że istniało kilka formatów zapisu daty. W pliku z danymi antropometrycznymi było to różne ustawienie dnia, miesiąca i roku badania, a także odmienne separatory. W pliku „dane_inbo” niepotrzebnie znajdują się informacje o godzinie badania, dlatego zostały usunięte.

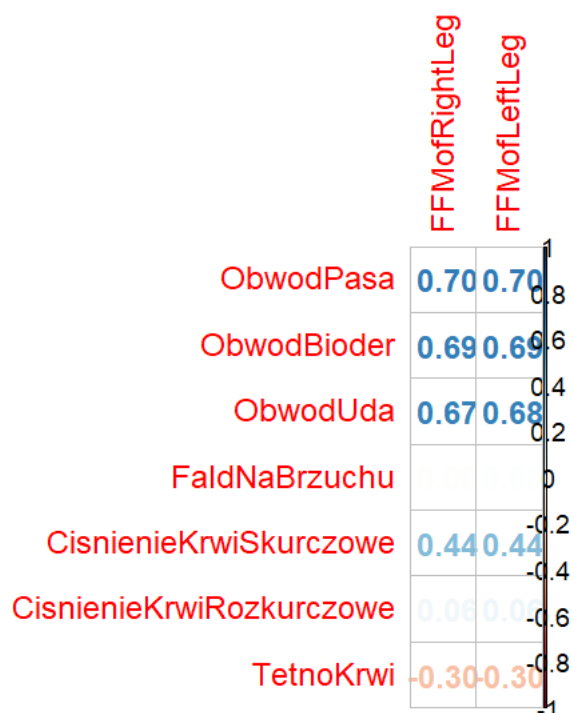
```
antropo$DataBadania = gsub("-", ".", antropo$DataBadania)
entire_antropo$DataBadania = gsub("-", ".", entire_antropo$DataBadania)
for(j in 1:nrow(antropo))
{
  splited = unlist(strsplit(antropo[j,2], split=".[.]"))
  ent_splited = unlist(strsplit(entire_antropo[j,2], split=".[.]"))
  if (nchar(splited[1]) != 4)
  {
    day = splited[1]
    month = splited[2]
    year = splited[3]
    date = paste(paste(year, month, sep="."), day, sep=".")
    antropo[j,2] = date
    entire_antropo[j,2] = date
  }
}
inbo$DataBadania = substr(inbo$DataBadania, 1, 10)
```

W celu otagowania danych tworzę kolumny opisujące stan umięśnienia nóg pacjentów – bazuję przy tym na kolumnach opisujących Fat-Free Mass dla każdej z nóg i wartości, pomiędzy którymi dana pacjenta powinna się znajdować. Jeżeli Fat-Free Mass osoby jest wyższa niż górna granica pożądanego przedziału, wartość kolumny to 1 – mięśnie są mocniej rozwinięte. Jeżeli są w normie lub wartość jest mniejsza niż dolna granica przedziału – przypisane zostanie 0.

```
for(j in 1:nrow(final))
{
  if (final[j,15] > final[j,17])
  {
    final[j, 18] = 1
  }
  else
  {
    final[j, 18] = 0
  }
}
colnames(final)[18] = "Miesnielewejnogi"
final[,18]=factor(final[,18])
```

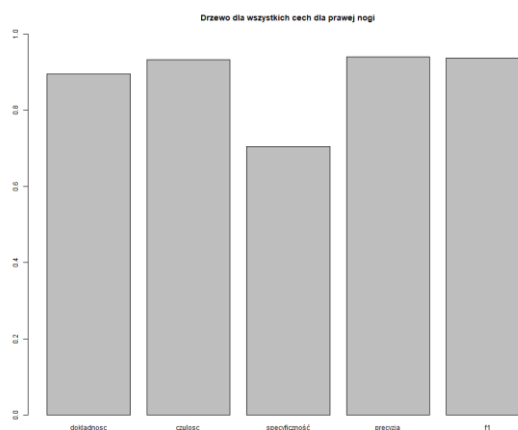
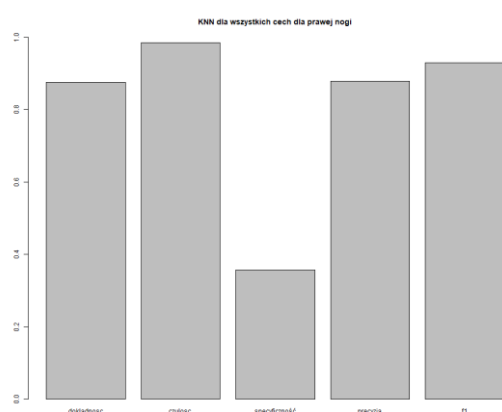
Wyniki

Pierwszą kluczową dla projektu kwestią jest możliwość skorelowania danych antropometrycznych, na których operują modele, z danymi inbody, z których powstaje później otagowanie. Jak się okazuje po zbadaniu współczynników korelacji Pearsona, między danymi istnieją liniowe zależności, szczególnie pomiędzy obwodami pasa, bioder i uda a Fat-Free Mass obu nóg.

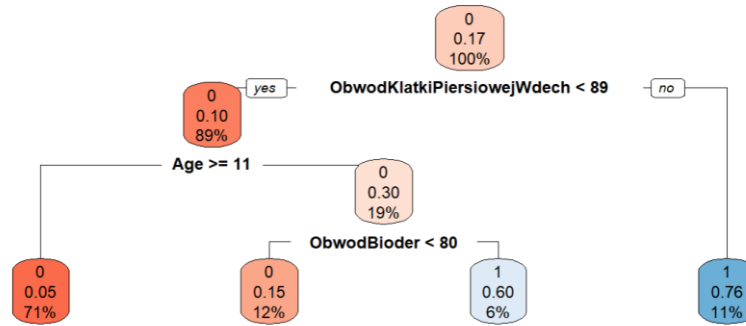


Wyniki modeli:

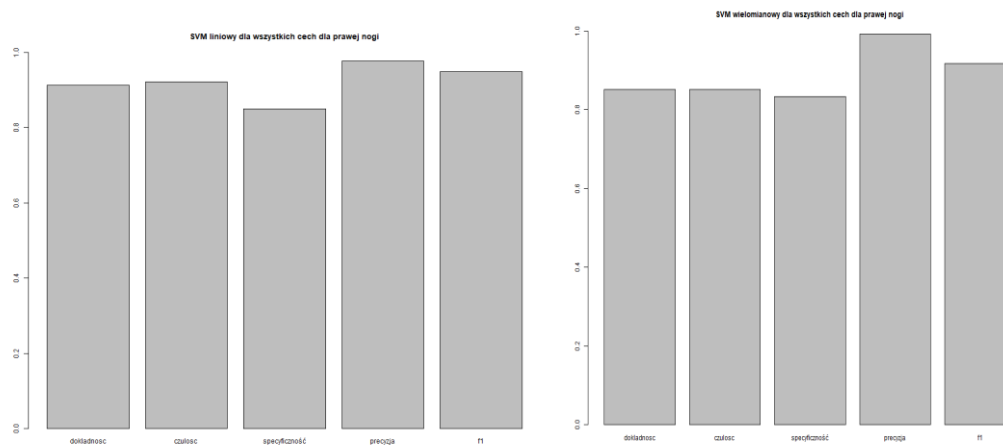
- dla wszystkich cech, FFM dla prawej nogi:



Powyżej wyniki KNN i drzewa decyzyjnego



Powyżej wykreowane drzewo

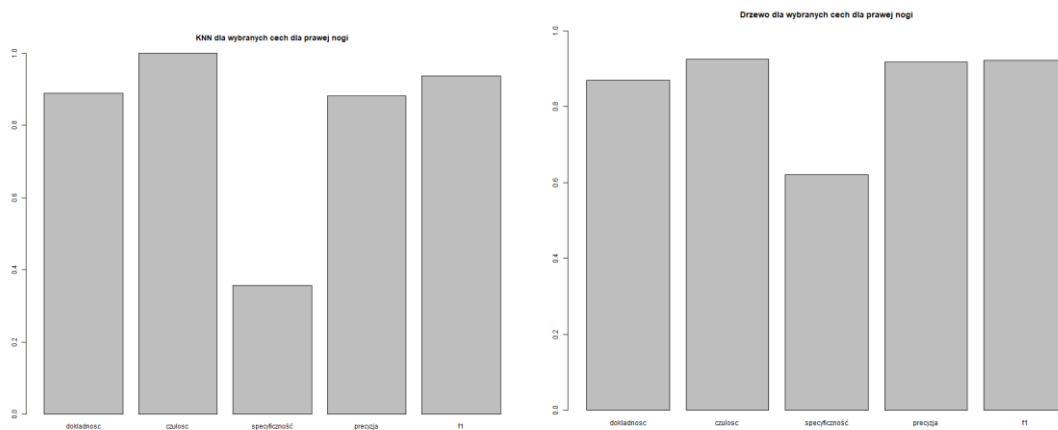


Powyżej wyniki SVM

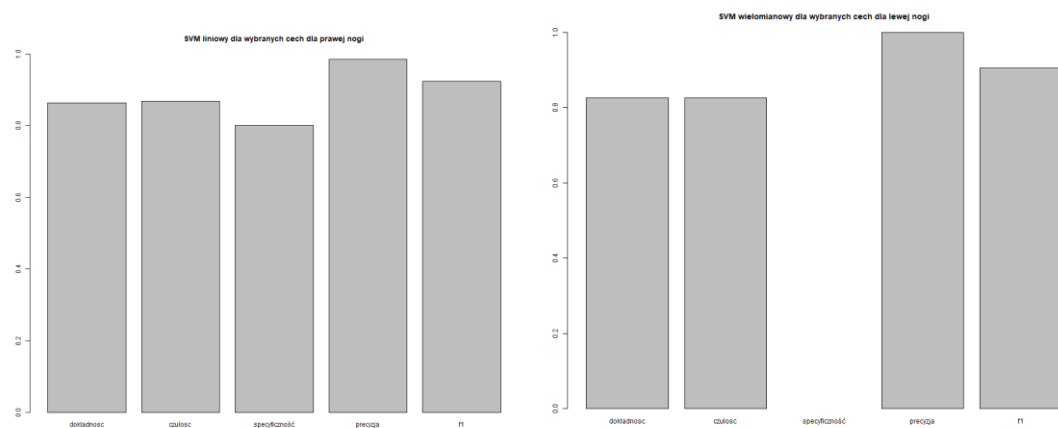
	KNN	Drzewo	SVM lin.	SVM poly.
accuracy	88%	89%	91%	88%
sensitivity	98%	93%	92%	88%

Zbiorne wyniki

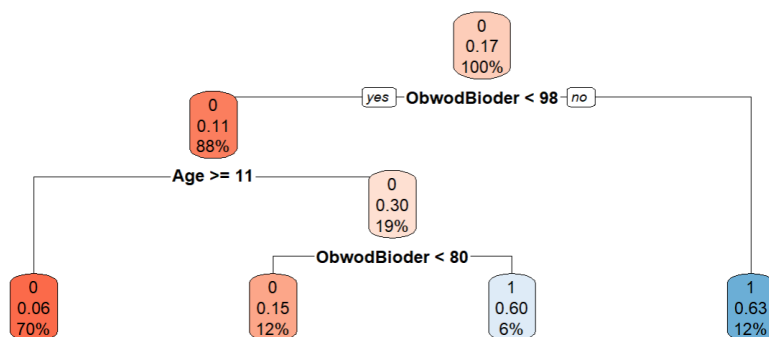
- dla wybranych cech, FFM dla prawej nogi:



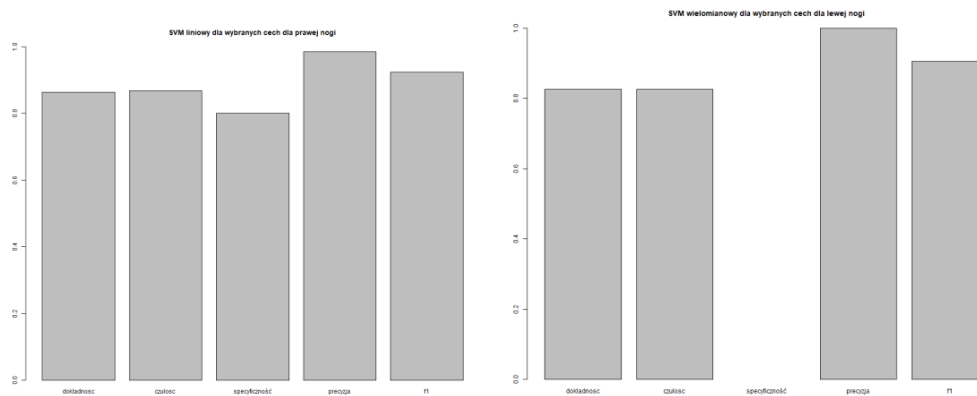
Powyżej wyniki KNN i drzewa decyzyjnego



Powyżej wyniki SVM



Powyżej wykreowane drzewo

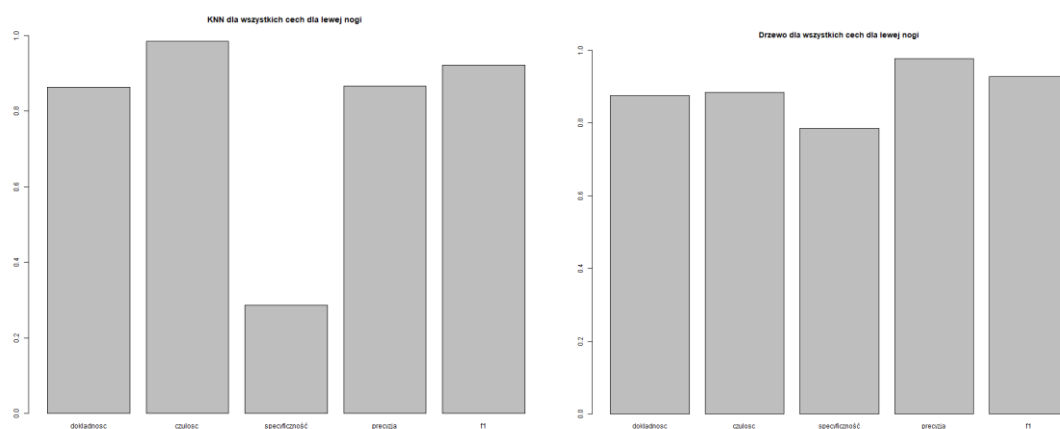


Powyżej wyniki SVM

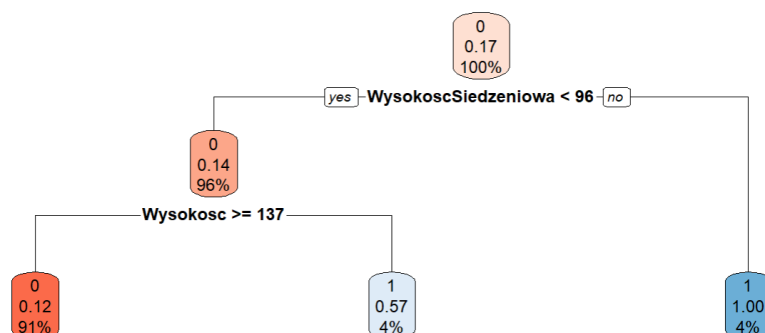
	KNN	Drzewo	SVM lin.	SVM poly.
accuracy	88%	87%	86%	85%
sensitivity	100%	92%	87%	85%

Zbiorcze wyniki

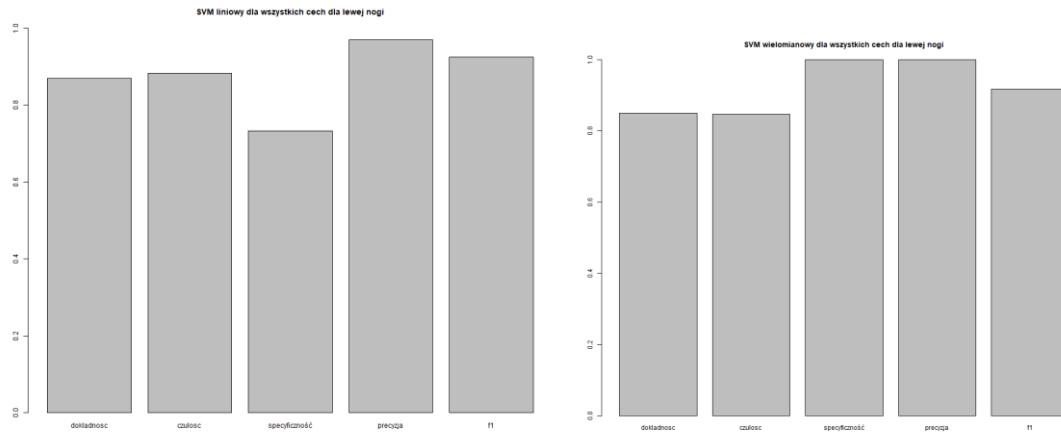
- dla wszystkich cech, FFM dla lewej nogi:



Powyżej wyniki KNN i drzewa



Powyżej drzewo decyzyjne

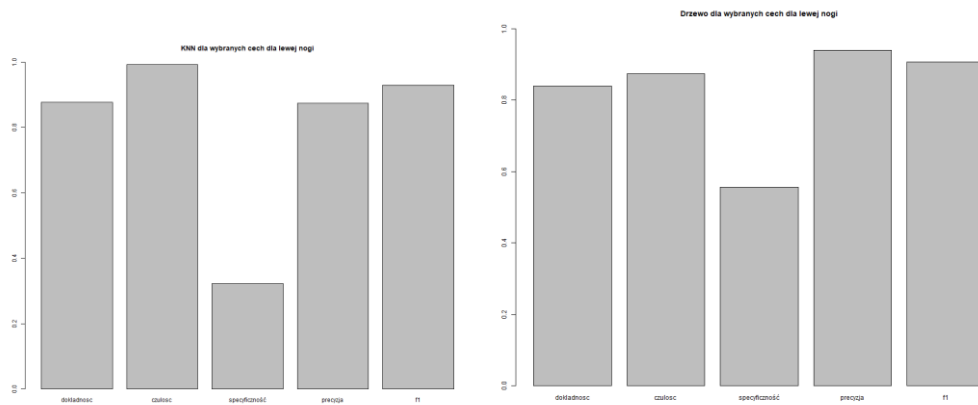


Powyżej SVM

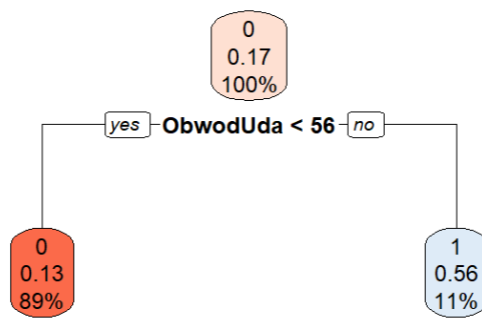
	KNN	Drzewo	SVM lin.	SVM poly.
accuracy	86%	88%	87%	85%
sensitivity	98%	88%	88%	84%

Zbiorcze wyniki

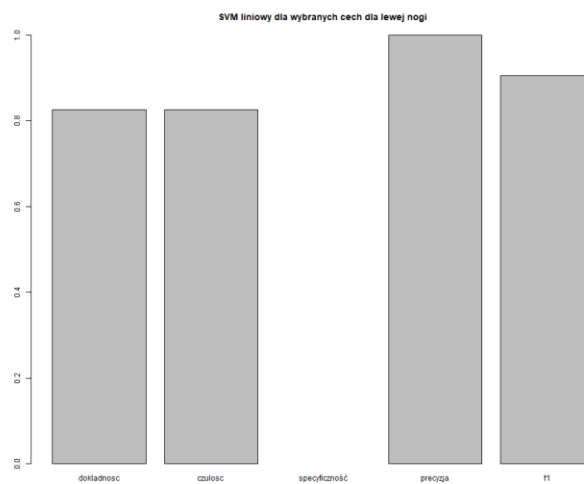
- dla wybranych cech, FFM dla lewej nogi:



Wyniki dla KNN i drzewa



Wygenerowane drzewo decyzyjne



Wyniki SVM

	KNN	Drzewo	SVM lin.	SVM poly.
accuracy	88%	84%	82%	82%
sensitivity	99%	97%	82%	82%

Zbiorcze wyniki

Wnioski

Wszystkie wyniki uzyskane przez modele w kategoriach dokładności i czułości są bardzo wysokie – żaden wynik nie spadł poniżej 80%. To nie dziwi, patrząc na współczynniki Pearsona – z badania korelacji wynika, że jak najbardziej pomiędzy niektórymi, uwzględnionymi danymi antropometrycznymi a Fat-Free Mass obu nóg zachodzą liniowe zależności (a więc wzrost masy tłuszczowej w dolnych rejonach ciała pociąga za sobą wzrost masy mięśniowej w nogach), co też prowadzi do konkluzji, że:

- 1) otyłość powoduje wzrost masy beztłuszczowej, czyli przede wszystkim mięśniowej kończyn dolnych [6],
- 2) jeżeli podczas badań zostanie wykryta wyższa niż normalnie masa mięśniowa nóg, przyczyną może być otyłość.

Ciekawą obserwacją jest fakt, że pomimo wykrycia korelacji między zastosowanymi zmiennymi w drugim, mniejszym zestawie danych wyniki wcale nie są lepsze od zestawu pełnych danych antropometrycznych. Można by nawet postawić tezę, że są one gorsze, i to zarówno pod kątem dokładności, jak i czułości.

Interesujące są również wnioski płynące z porównania obu kończyn – prawa noga ma wyższe rezultaty niż lewa. Nie są to ogromne różnice, jednak jak wiemy, pomimo względnej symetryczności tych kończyn, nie są one w całości identyczne. Prawa noga jest zazwyczaj dominująca – sprawia to, że lewa rozwija mięśnie na delikatnie wyższym poziomie. Dlaczego więc to prawa noga ma wyższe wyniki? Na to pytanie nie jestem w stanie odpowiedzieć.

Całościowe wyniki wskazują na SVM o liniowym jądrze i drzewo decyzyjne jako najoptymalniejsze modele do predykcji w moim problemie. Warto tu jednak zwrócić uwagę też na KNN, który o ile osiąga niekiedy średnie (w porównaniu do reszty) dokładności, o tyle nadrabia czułością, która jak już wspomniałem może być dosyć istotna – zawsze są to wyniki niedaleko perfekcji.

Podsumowanie

Dzięki zbadaniu korelacji między różnymi parametrami fizycznymi, a także predykcji masy beztłuszczowej nóg udało się ustalić i potwierdzić, że w wyniku przybierania człowieka na masie mięśnie w kończynach dolnych stają się większe. W pewien sposób działa to również w drugą stronę, ponieważ dzięki takiej zależności po odkryciu ponadprzeciętnych mięśni w nogach można zdiagnozować nadwagę. Projekt zasugerował również możliwość pochylenia się nad różnicami w umięśnieniu pomiędzy dwoma pozornie symetrycznymi kończynami – być może jest to po prostu przypadek, być może nie, niemniej dobrze byłoby w przyszłości pochylić się nad tym zagadnieniem i dogłębniej je zbadać.

Literatura

1. **Goodpaster BH, Thaete FL, Kelley DE** 2000 Thigh adipose tissue distribution is associated with insulin resistance in obesity and in type 2 diabetes mellitus.
2. **Michael D. Jensen** Role of Body Fat Distribution and the Metabolic Complications of Obesity, 2008
3. **Urszula T. Iwaniec, Russell T. Turner** Influence of Body Weight on Bone Mass, Architecture and Turnover, 2017
4. **C. P. Lucas, J. A. Estigarribia, L. L. Darga, G. M. Reaven** Insulin and blood pressure in obesity, 1985
5. **Waist Circumference and Waist–Hip Ratio: Report of a WHO Expert Consultation** Geneva, 8–11 December 2008

6. D. J. Tomlison, R. M. Erskine, C. I. Morse, K. Winwood, Gladys Onambele-Pearson
The impact of obesity on skeletal muscle strength and structure through
adolescence to old age, 2015