



Dokumentacja projektowa

INŻYNIERIA BIOMEDYCZNA W PRAKTYCE

Jakub Nocoń

Sekcja 3

Kierunek: Informatyka

Wydział Matematyki Stosowanej [RMS]

Rok akademicki 2021/2022

Tytuł: *Body Fat Mass a fałdy skórno-tłuszczowe*

Spis treści

1	Wstęp	2
2	Analiza zagadnienia	2
2.1	Tkanka tłuszczowa	2
2.2	Fałdy skórno-tłuszczowe	3
2.3	Klasyfikatory	4
3	Specyfikacja wewnętrzna	4
4	Wyniki	8
4.1	Obliczanie BFM	8
4.2	Analiza wykresów	8
4.3	Wyniki klasyfikatorów	11
5	Wnioski	12
5.1	Obliczone BFM	12
5.2	Wykresy - wnioski	12
5.3	Klasyfikatory - wnioski	12
6	Podsumowanie	13

1 Wstęp

Tematem mojego projektu jest zbadanie zależności parametru Body Fat Mass do grubości fałdów skórno-tłuszczowych na ramieniu, brzuchu oraz pod łopatką na podstawie arkuszy z wynikami pomiarów antropometrycznych i z *InBody*. Zdecydowałem się na taki temat, ponieważ chciałem przedstawić przedstawienia korelacji między różnymi parametrami związanymi z budową ludzkiego ciała, ale mającymi ze sobą coś wspólnego. Oba badane parametry są powiązane z tkanką tłuszczową, więc wydało mi się w miarę logiczne sprawdzenie zależności między nimi.

Tkanka tłuszczowa w obiegowej opinii uznawana jest za niekorzystną składową naszego ciała. Jednak jej niedobór również może mieć negatywne konsekwencje dla zdrowia - kluczową kwestią jest jej poziom w naszym organizmie. Jest to ważny składnik ludzkiego organizmu, który pełni istotne funkcje. Ciało potrzebuje pewnej podstawowej ilości tkanki tłuszczowej by funkcjonować, jednak musi być ona utrzymywana w ustalonym przedziale dla zachowania optymalnego zdrowia.

Antropometryczny pomiar grubości fałdów skórno-tłuszczowych pomaga w ocenie wielkości skumulowanej pod skórą warstwy tłuszczu, a Body Fat Mass (BFM) określa ilość masy tkanki tłuszczowej zawartej w organizmie.

W swoim projekcie użyłem następujących klasyfikatorów:

- k najbliższych sąsiadów,
- proste drzewo decyzyjne,
- random forest.

2 Analiza zagadnienia

2.1 Tkanka tłuszczowa

Całkowitą wagę ludzkiego organizmu można rozbić na tkankę tłuszczową i masę beztłuszczową. Do masy beztłuszczowej zalicza się wszystko co nie jest tkanką tłuszczową, czyli mięśnie, białko, woda, kości, minerały, narządy.

Tkanka tłuszczowa bierze udział w procesach metabolicznych, wydziela biologicznie czynne substancje, które stymulują pracę układu odpornościowego. Ponadto ma właściwości termoizolacyjne, dzięki czemu chroni przed utratą ciepła w niskich temperaturach. Najczęściej za optymalną u ludzi dorosłych uznaje się

20–25% masy ciała u kobiet i 15–20% masy ciała u mężczyzn. Niższe wartości klasyfikowane są jako niedowaga natomiast wyższe – otyłość. Tłuszczowa masa ciała, może być podzielona na mniejsze składowe, aby mieć lepsze pojęcie tkance tłuszczowej należy je rozróżniać i zrozumieć jakie funkcje pełnią w organizmie. Podział tkanki tłuszczowej:

- ze względu na rozmieszczenie:
 - tkanka tłuszczowa powierzchniowa (podskórna) - znajduje się bezpośrednio pod skórą, w okolicach brzusznych, podłopatkowych, pośladkowych i udowych. Nie komunikuje się ona z narządami wewnętrznymi. Stanowi 80-90% tłuszczu w organizmie.
 - tkanka tłuszczowa wewnętrzna (trzewna) - znajduje się głównie wokół narządów jamy brzusznej, w tym wątroby i jelit. Odprowadza ona wolne kwasy tłuszczowe i adipokiny do krążenia wrotnego, gdzie mogą wywierać wpływ na metabolizm. Stanowi 6-20% tłuszczu w organizmie, a wyższe stężenie tłuszczu trzewnego w stosunku do podskórnego powoduje zwiększenia ryzyka chorób metabolicznych.
- ze względu na funkcję:
 - tkanka tłuszczowa „aktywna” - rodzaj tkanki tłuszczowej, który jest obecny w organach, szpiku kostnym, komórkach nerwowych czy w samym mózgu i jest odpowiedzialny za: utrzymanie odpowiedniej rezerwy energii („paliwo metaboliczne”); zachowanie ciepłoty ciała („izolator”); ochronę narządów wewnętrznych i stawów („amortyzator”);
 - tkanka tłuszczowa „zapasowa” - rezerwa energetyczna. Zawartość w organizmie tego rodzaju tkanki tłuszczowej oddziałuje na zewnętrzny wygląd ciała człowieka.

2.2 Fałdy skórno-tłuszczowe

Aby zmierzyć poziom tkanki tłuszczowej można zastosować pomiar fałdu skórno-tłuszczowego, który może być najdokładniejszy, jednak zależy od rozłożenia tkanki w ciele i dokładności pomiaru. Grubość fałdów skórno-tłuszczowych mierzy się po lewej stronie ciała. Pomiaru obejmują 3–9 miejsc anatomicznych organizmu. Fałd zostaje ujęty w odpowiednim miejscu, aby podnieść warstwę skóry i tkankę tłuszczową, ale nie mięsień. Następnie nakładany jest fałdomierz pod kątem prostym do wyznaczonego punktu, a wynik podawany w milimetrach. Należy wziąć

pod uwagę średnio 2 pomiary. Jeśli są one bardzo zróżnicowane, należy wykonać trzeci pomiar i obliczyć medianę. Przykładowe miejsca pomiarowe znajdują się:

- **nad mięśniem trójgłowym ramienia** - fałd podłużny, mierzony w połowie odległości między wyrostkiem barkowym łopatki a wyrostkiem łokciowym kości łokciowej, w tzw. punkcie środkowym ramienia przy rozluźnionej kończynie;
- **pod dolnym kątem łopatki** - fałd biegnący skośnie w dół i w bok od dolnego kąta łopatki;
- **na brzuchu** - fałd poprzeczny, mierzony nieco poniżej (ok. 1cm) i nieco w bok (ok. 5cm) od pępka.

2.3 Klasyfikatory

Algorytm **k najbliższych sąsiadów** (algorytm **k-nn** z ang. k nearest neighbours) – jeden z algorytmów regresji nieparametrycznej używanych w statystyce do prognozowania wartości pewnej zmiennej losowej jak i do klasyfikacji. k-nn to rodzaj klasyfikacji, w której funkcja jest tylko aproksymowana lokalnie, a wszystkie obliczenia są odraczane do czasu oceny funkcji.

Drzewo decyzyjne – graficzna metoda wspomagania procesu decyzyjnego. Algorytm drzew decyzyjnych jest również stosowany w uczeniu maszynowym do pozyskiwania wiedzy na podstawie przykładów. Technika drzew decyzyjnych (klasyfikacyjnych) pozwala na: wyznaczenie zasad decyzyjnych opisujących reguły przypisywania obiektów do wyróżnionych klas (zasady odwołują się do wartości atrybutów opisujących obiekty), analizowanie zbioru obiektów opisywanych przez przyjęty zestaw atrybutów.

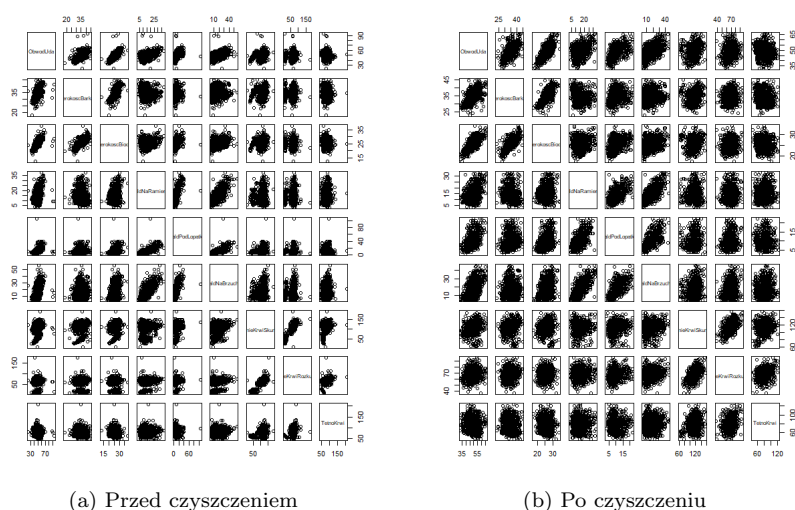
Random forest (losowy las decyzyjny) – metoda zespołowa uczenia maszynowego dla klasyfikacji, regresji i innych zadań, która polega na konstruowaniu wielu drzew decyzyjnych w czasie uczenia i generowaniu klasy, która jest dominantą klas (klasyfikacja) lub przewidywaną średnią (regresja) poszczególnych drzew. Losowe lasy decyzyjne poprawiają tendencję drzew decyzyjnych do nadmiernego dopasowywania się do zestawu treningowego.

3 Specyfikacja wewnętrzna

Sekcja ta poświęcona jest omówieniu napisanego kodu i generalnemu przedstawieniu poszczególnych fragmentów kodu o bardziej kluczowym znaczeniu.

Pierwszym krokiem było pobranie odpowiednich bibliotek (stringr, dummies, caTools, rpart, rpart.plot, randomForest, class, corrplot, PerformanceAnalytics, scatterplot3d) i podpięcie arkuszy z pomiarami. Następną część stanowi ogólne czyszczenie wstępne obu arkuszy z błędnych pozycji.

W arkuszu z danymi antropometrycznymi na początek poprawiono nierealne wyniki z kolumn z wysokością i wysokością siedzeniową. Następnie usunięto najbardziej odstające wyniki z pozostałych kolumn. W miejsce pustych komórek i tych z 0 wpisane zostało *NA*.



Rysunek 1: Wykresy zgodności

W arkuszu *InBody* na początek usunięto ponad 100 kolumn z niepotrzebnymi danymi, np. tych z impedancją. Poprawiono też nieprawidłowe wartości wieku. W miejsce pustych komórek i tych z 0 wpisane zostało *NA* - wyjątkiem były kolumny: WeightControl, BFMCControl i FFMControl, gdzie pozostawiono wartości zerowe.

Następny krok stanowiło posortowanie danych z obu arkuszy i stworzenie na ich podstawie 2 macierzy. Obie macierze scalono na podstawie numeru ID oraz daty badania. Otrzymana scalona macierz została zapisana jako ramka danych i ze-stack-owana.

W celu łatwiejszej pracy z projektem wyizolowano kolumny związane z: ID, wagą, ilością wody, protein, minerałów w organizmie i masą tkanki tłuszczowej. Najpierw wyznaczono tylko te wiersze, w których wartość minerałów nie przekracza 10 i przepisano je do nowej kolumny.

```

# mineraly<10
mn=11
mnrls=data.frame()
for(u in c(1:nrow(needed_1)))
{
  if(!is.na(needed_1[u, mn])) # kolumny niepuste
  {
    if(needed_1[u, mn] < 10) # wyszukanie wartosci <10
    {
      mnrls=rbind(mnrls, needed_1[u, mn]) # przepisanie wartosci
    }
    else
    {
      mnrls=rbind(mnrls, NA) # zamiana >10 na NA
    }
  }
  else
  {
    mnrls = rbind(mnrls, NA)
  }
}
needed_1=cbind(needed_1,mnrls) # dodanie kolumny z przepisanyimi wynikami
names(needed_1)[17]="Minerals2" # zmiana nazwy kolumny

```

Rysunek 2: Pętla dla minerałów

Następnie w kolumnach zawierających dolną i górną granicę prawidłowej wartości parametru uporządkowano liczby w wierszach tak, aby w kolumnie z granicą dolną znajdowała się wartość mniejsza od tej w kolumnie z granicą górną. Proces ten przeprowadzono dla zakresów skrajnych wagi, protein, wody (TBW), minerałów i tkanki tłuszczowej (BFM).

```

# lower/upper range dla BFM
for(w in c(1:nrow(needed_1)))
{
  if(!is.na(needed_1[w, 15]) & !is.na(needed_1[w, 16]))
  {
    if(needed_1[w, 15] != "" & needed_1[w, 16] != "")
    {
      if(needed_1[w, 15] > needed_1[w,16]) # sprawdzenie, czy wartosc lower > upper
      {
        flu = needed_1[w, 15] #
        needed_1[w, 15] = needed_1[w, 16] # zamiana miejscami, aby wartosc wieksza byla
        needed_1[w, 16] = flu # w kolumnie upper range
      }
    }
  }
}

```

Rysunek 3: Pętla segregująca wartości skrajne

Potem na nowo policzono za pomocą pętli *for* wartość Body Fat Mass (BFM) ze wzoru:

$$BFM = waga - (TBW + protein + minerals).$$

Wynik zapisano do nowej kolumny w tej samej ramce danych co pierwotna war-

tość parametru.

```
# BFM=waga-(tbw+proteiny+mineraly)
nowe_BFM = data.frame()
wg=2
wd=5
pr=8
mnr=17
for(u in c(1:nrow(needed_1)))
{
  # sprawdzenie, czy potrzebne komórki sa niepuste
  if(!is.na(needed_1[u, wg]) & !is.na(needed_1[u, wd]) & !is.na(needed_1[u, pr]) & !is.na(needed_1[u, mnr]))
  {
    if(needed_1[u, wg] != "" & needed_1[u, wd] != "" & needed_1[u, pr] != "" & needed_1[u, mnr] != "")
    {
      # obliczenie nowego BFM ze wzoru, jesli wszystkie potrzebne skladniki wystepuja
      nowe_BFM = rbind(nowe_BFM, needed_1[u, wg]-(needed_1[u, wd]+needed_1[u, pr]+needed_1[u, mnr]))
    }
    else
    {
      # wpisanie NA, jesli ktorys z potrzebnych skladnikow nie wystepuje
      nowe_BFM = rbind(nowe_BFM, NA)
    }
  }
  else
  {
    nowe_BFM = rbind(nowe_BFM, NA)
  }
}
needed_1=cbind(needed_1,nowe_BFM) # dopisanie kolumny z obliczonym BFM
names(needed_1)[18]="New_BFM" # zmiana nazwy kolumny
```

Rysunek 4: Pętla obliczająca nowe BFM

Następnie wyrysowano kilka wykresów, które przedstawiono dokładniej w sekcji „Wyniki”. Potem stworzono 2 ramki danych, które w późniejszej fazie wykorzystano w pracy z klasyfikatorami. Pierwsza ramka: płeć (1 - mężczyzna, 0 - kobieta), wysokość, fałd na ramieniu, fałd na brzuchu, fałd pod łopatką, wiek, waga i samodzielnie policzone BFM. Druga ramka: płeć (1 - mężczyzna, 0 - kobieta), wysokość, fałd na ramieniu, fałd na brzuchu, fałd pod łopatką, wiek, waga i samodzielnie policzone średnie BFM. Owe średnie BFM - *sr_BFM* - powstało przez porównanie obliczonego BFM do skrajnych zakresów. Jeśli BFM znajduje się między granicami to wartość *sr_BFM* wynosi 1 - analogicznie w przeciwnym wypadku wynosi 0.

Dalej rozpisana została funkcja oceniająca, którą przedstawiono w czasie zajęć projektowych. Później wydzielono zestawy testowe i treningowe. W końcowej części skryptu znajdują się klasyfikatory KNN, proste drzewo decyzyjne i random forest, które wykorzystują gotowe komendy z bibliotek specjalnie przeznaczonych dla tych metod.

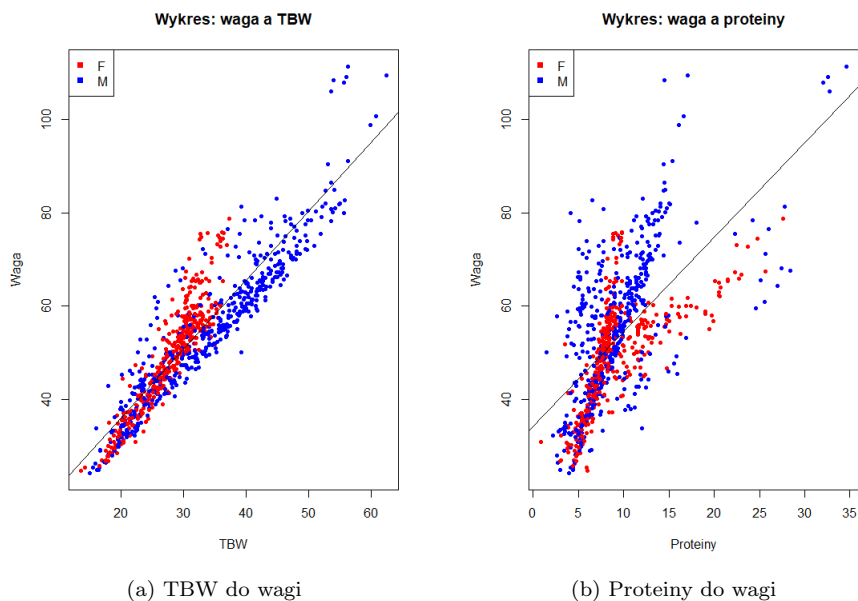
4 Wyniki

4.1 Obliczanie BFM

Analizę wyników należy zacząć porównania rezultatów samodzielnie obliczonego Body Fat Mass z tym pierwotnie znajdującym się w arkuszu *InBody*. Są one praktycznie identyczne, ponieważ pierwotne dane z arkusza zapisano z dokładnością do 1 cyfry po przecinku, a obliczone przeze mnie nowe BFM uwzględniła 2 cyfry po przecinku.

4.2 Analiza wykresów

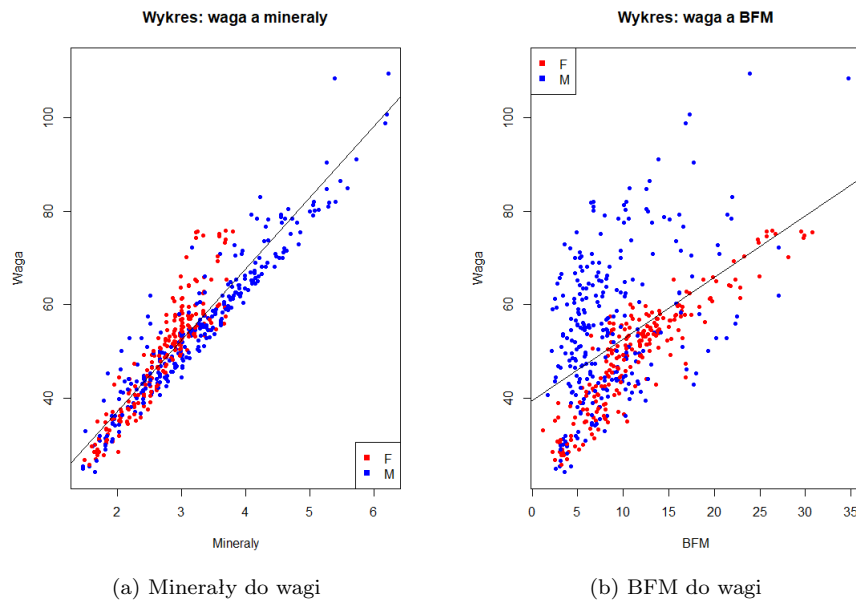
Na początek przygotowano wykresy zależności wagi od każdej z jej składowych, czyli Total Body Water (TBW), protein, minerałów i BFM. Na wszystkich z tych wykresów otrzymano, więc zależność taką, że kiedy jeden z powyższych parametrów rośnie to waga też.



Rysunek 5: Waga a jej składowe (1)

Na powyższym wykresie (a) można zauważyć, że u kobiet waga nie przekracza wartości 80kg, a zawartość wody w organizmie nie przekracza 40l. Na powyższym wykresie (b) można zauważyć, że znaczna większość wyników zawartości

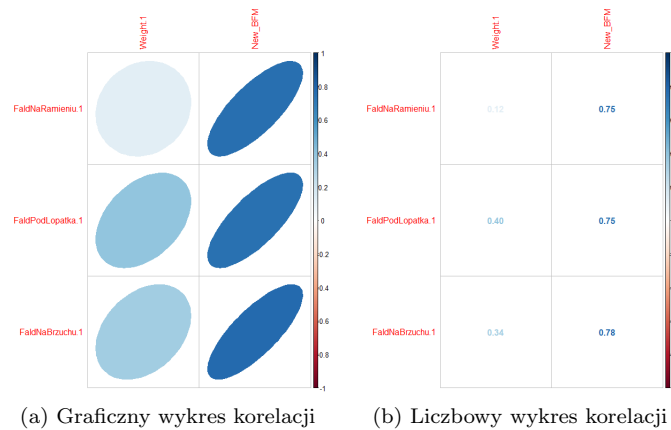
białka u kobiet i mężczyzn zawiera się w przedziale 2-15kg. Wyniki spoza tego przedziału należą głównie do kobiet.



Rysunek 6: Waga a jej składowe (2)

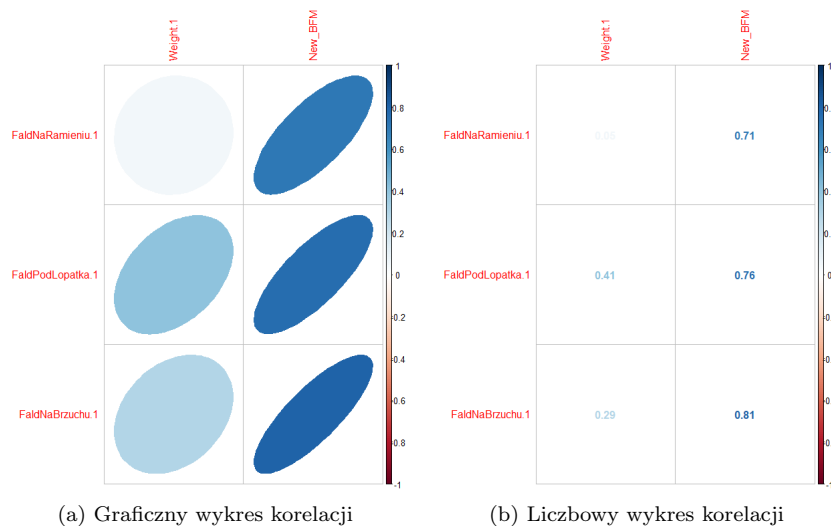
Na powyższym wykresie (a) można zauważyć, że wartość minerałów u kobiet nie przekracza 4kg. Kobiety posiadające wagę z przedziału od ok. 70 do 80kg mają mniej minerałów od mężczyzn o tej samej wadze. Na podstawie powyższego wykresu (b) można stwierdzić, iż kobiety posiadają większą masę tkanki tłuszczowej w stosunku do mężczyzn o tej samej wadze.

Kolejne wykresy przedstawiają korelację między poszczególnymi fałdami skórno-tłuszczowymi a wagą i BFM. Dodatnie korelacje są wyświetlane w kolorze niebieskim, a ujemne w kolorze czerwonym. Intensywność koloru i spłaszczenie elipsy są proporcjonalne do współczynników korelacji - im ciemniejszy kolor i bardziej spłaszczona elipsa, tym korelacja jest silniejsza.



Rysunek 7: Wykresy korelacji - metoda Pearsona

Na podstawie wykresów korelacji, w których zastosowana została metoda Pearsona, można zauważyć w miarę silną korelację fałdami skórno-tłuszczowymi a BFM. Współczynnik korelacji między każdym z nich a BMF wynosi co najmniej 0,75. Natomiast korelacja między fałdami a wagą nie jest aż tak silna - współczynnik waha się od 0,12 do 0,4 i największy jest dla fałdu pod łopatką.



Rysunek 8: Wykresy korelacji - metoda Spearmana

Na podstawie wykresów korelacji, w których zastosowana została metoda Spe-

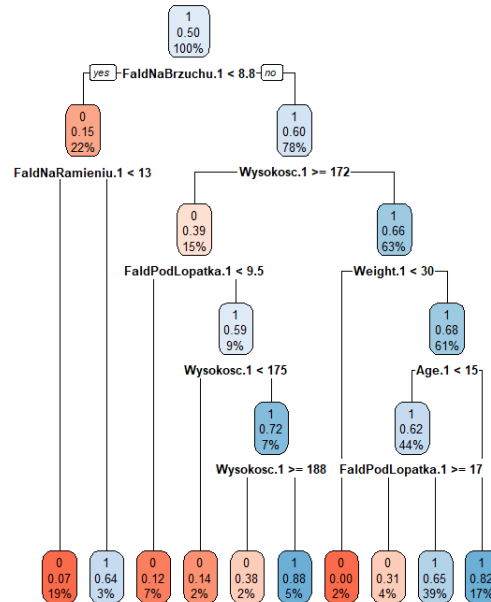
armana, można zauważyć, że współczynniki korelacji są bardzo zbliżone do tych uzyskanych metodą Pearsona, lecz w przypadku BFM różnica między najmniejszą, a największą wartością współczynnika jest większa. Taka sama sytuacja występuje w przypadku współczynników dla wagi.

4.3 Wyniki klasyfikatorów

klasyfikator	dokładność	czułość	specyficzność	precyzja	f1
KNN	0,731	0,696	0,766	0,744	0,719
drzewo decyzyjne	0,688	0,73	0,661	0,587	0,651
las losowy	0,71	0,694	0,727	0,739	0,716

Tablica 1: Oceny klasyfikatorów

KNN uzyskał najwyższe wyniki w dokładności, specyficzności, precyzji i f1. Największą wartość czułości otrzymało natomiast drzewo decyzyjne, ale w pozostałych kategoriach uzyskało najniższe wyniki. Las losowy uzyskał wyniki bardzo zbliżone do KNN-a, ale trochę niższe.



Rysunek 9: Graf drzewa decyzyjnego

5 Wnioski

5.1 Obliczone BFM

Otrzymane przeze mnie BFM nie różni się od tego pierwotnie znajdującego się w arkuszu *InBody*. W ramach podsumowania można stwierdzić, że pierwotna masa tkanka tłuszczowej jest zaokrągleniem tej samodzielnie obliczonej. W przypadku dokładniejszego czyszczenia arkusza otrzymane wyniki mogłyby się różnić w większym stopniu.

5.2 Wykresy - wnioski

W porównaniu do kobiet o tej samej wadze, mężczyźni posiadają większą ilość wody w organizmie - spowodowane może być to tym, że tkanka mięśniowa zawiera zdecydowanie więcej wody (75-80%) niż tkanka tłuszczowa (10-20%).

Kobiety posiadają większą masę tkanki tłuszczowej w porównaniu do mężczyzn. Kobiety w składzie ciała mają więcej tłuszczu (od 5% do 8%) niż mężczyźni. Jest to między innymi wynik ewolucji. Pełni on rolę magazynu energii na czas porodu, ciąży czy karmienia piersią. Jest też niezbędny dla zachowania równowagi hormonalnej.

W przypadku wykresów korelacji bardziej rzetelna wydaje się metoda Pearsona, gdyż w czasie czyszczenia zredukowano ilość odstających przypadków.

5.3 Klasyfikatory - wnioski

Najbardziej efektywnym z zastosowanych klasyfikatorów dla przyjętej próbki danych wydaje się KNN, ponieważ otrzymał on najlepsze rezultaty dla prawie wszystkich kategorii, z wyjątkiem czułości, gdzie górowało proste drzewo decyzyjne. Random forest osiągnął wyniki zbliżone do KNN-a. Mimo to KNN i tak wydaje się najbardziej opłacalny dla posiadanej próbki, ponieważ las losowy osiągał mniejsze wartości w poszczególnych kategoriach.

6 Podsumowanie

Celem mojego projektu było przedstawienie zależności pomiędzy fałdami skórno-tłuszczowymi człowieka a masą tkanki tłuszczowej organizmu. Wydaje mi się, że cel ten udało mi się osiągnąć w zadowalającym stopniu ze zwróceniem uwagi na powiązanie z wagą i innymi jej składowymi jak Total Body Water, czy minerały z podziałem na płeć.

W kwestii rozwoju projektu można by zacząć od dokładniejszego czyszczenia danych znajdujących się w arkuszach. Dobrym pomysłem mogłoby być również włączyć do projektu wskaźniki bazujące na wadze jak BMI, czy PBF. Można by też przeprowadzić podobne obliczenia dla innych składowych wagi jak Fat Free Mass, Soft Lean Mass lub Skeleton Muscle Mass i porównanie rezultatów z wyróżnieniem występujących zależności.

Literatura

- [1] Drzewo decyzyjne. https://mfiles.pl/pl/index.php/Drzewo_decyzyjne.
- [2] k-nearest neighbors algorithm. https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm.
- [3] Las losowy. https://pl.wikipedia.org/wiki/Las_losowy.
- [4] Result sheet interpretation. <https://uk.inbody.com/about-inbody/result-sheet-interpretation/>.
- [5] A. BROŃCZYK-PUZOŃ, A. KOSZOWSKA, J. B. Podstawowe pomiary antropometryczne i pochodne wskaźniki w poradnictwie dietetycznym – część pierwsza. *Pielęgniarstwo i Zdrowie Publiczne* (2018).
- [6] OŁDAK, K. Zawartość tkanki tłuszczowej w organizmie – jak ją obliczyć i zmierzyć? <https://dietetycy.org.pl/zawartosc-tluszczu-wzor/>, 2021.
- [7] TOMBOC, K. A guide to achieving a healthy body fat percentage. <https://inbodyusa.com/blogs/inbodyblog/your-body-and-you-a-guide-to-body-fat/>, 2020.