



Dokumentacja projektowa

Inżynieria Biomedyczna w Praktyce (Informatyka st. I sem. 4)

**Artur Łukaszek**

Sekcja 2

Kierunek: Informatyka

Wydział Matematyki Stosowanej

Rok akademicki 2021/2022

**Tytuł:** *Przedstawienie BMI na podstawie obwodów części ciała.*

# 1 Wstęp

Celem mojego projektu jest przedstawienie BMI w skali: niedowaga, waga prawidłowa, nadwaga oraz otyłość. Do określenia BMI wykorzystałem obwody następujących części ciała: szyi, klatki piersiowej, pasa, bioder, uda i ramienia. W projekcie wykorzystałem 3 klasyfikatory do przewidywania BMI: K-najbliższych sąsiadów, Random Forest oraz Drzewo Decyzyjne.

Dokonując pomiarów obwodów, należy zwrócić uwagę na to, aby centymetr był prowadzony poprzecznie w stosunku do danego mięśnia i nie był przyłożony ani zbyt luźno, ani zbyt ciasno. Pomiary dokonuje się przy rozluźnionych mięśniach z dokładnością do 0,5cm. Jeżeli prawy lub lewy wymiar tego samego mięśnia są różne, można zapisać średnią lub mniejszy wynik.

BMI jest wskaźnikiem, który jest obliczany przez porównanie wzrostu z masy ciała. Jego wartość jest pomocna w ocenie ryzyka wystąpienia chorób związanych z nadwagą takich jak miażdżyca lub choroba niedokrwienna serca. Im mniejsza wartość BMI, tym ryzyko wystąpienia chorób jest mniejsze. Wartość BMI oblicza się jako współczynnik powstały przez podzielenie masy ciała podanej w kilogramach przez kwadrat wysokości podanej w metrach.

## 2 Analiza zagadnienia

Aby dobrze wykonać pomiary należy to robić w następujący sposób:

- udo – stoimy w lekkim rozkroku, ciężar ciała rozłożony równomiernie, mięśnie ud rozluźnione, należy mierzyć tuż poniżej fałdy pośladkowej, centymetr prowadzony równolegle do podłoża;
- biodra – stoimy w lekkim rozkroku, ciężar ciała rozłożony równomiernie, mięśnie pośladków rozluźnione, mierzymy przez środek pośladków, centymetr prowadzimy równolegle do podłoża;
- klatka piersiowa - mierzymy 1 cm powyżej linii sutków, mięśnie piersiowe i pleców rozluźnione, plecy wyprostowane, ręce opuszczone, tuż po swobodnym wydechu, centymetr prowadzimy równolegle do podłoża;
- szyja - mierzymy w największym miejscu, mięśnie karku rozluźnione, postawa wyprostowana, centymetr prowadzimy równolegle do podłoża;
- ramię - mierz zawsze u ręki niedominującej. Obwód ramienia wykonuj stojąc z opuszczonym wzdłuż ciała ramieniem w połowie odległości pomiędzy barkiem a łokciem;
- pas - miarę prowadzimy w najszerszym miejscu (zazwyczaj ok. 1 cm poniżej pępka), oczywiście na całym obwodzie równolegle do podłoża.

Wskaźnik masy ciała, czyli Body Mass Index (BMI) to sposób oceny aktualnej masy ciała i stopnia odżywienia organizmu. Wartość BMI nie jest adekwatna dla oceny zawartości tłuszczu u kobiet w ciąży i karmiących piersią. Równie ważne są inne wskaźniki mówiące o stanie zdrowia, takie jak obwód talii, zawartość tłuszczu w organizmie, wartość

ciśnienia tętniczego, stężenie cholesterolu, poziom aktywności fizycznej, używki oraz sposób żywienia. Wskaźnik BMI jest niezbędny w przypadku oceny zagrożenia chorobami związanymi z nadwagą i otyłością m.in. miażdżycą, cukrzyca. BMI kontrolować należy w przypadku częstych wahań masy ciała i podczas leczenia dietetycznego. Zaleca się, aby masę ciała kontrolować nie częściej niż raz w tygodniu, na czczo o tej samej porze dnia.

Szczegółowa klasyfikacja BMI:

< 16,0 wygłodzenie

16,0–16,9 wychudzenie

17,0–18,4 niedowagę

18,5–24,9 wartość prawidłową

25,0–29,9 nadwagę

30,0–34,99 I stopień otyłości

35,0–39,99 II stopień otyłości

≥ 40,0 III stopień otyłości

Tak wygląda tabela która pomaga określić BMI

wzrost cm	waga kg																							
	45.5	47.7	50.0	52.3	54.5	56.8	59.1	61.4	63.6	65.9	68.2	70.5	72.7	75.0	77.3	79.5	81.8	84.1	86.4	88.6	90.9	93.2	95.5	97.7
	Niedowaga								Norma								Nadwaga							
152.4	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
154.9	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	36	37	38	39	40
157.4	18	19	20	21	22	22	23	24	25	26	27	28	29	30	31	32	33	33	34	35	36	37	38	39
160.0	17	18	19	20	21	22	23	24	24	25	26	27	28	29	30	31	32	32	33	34	35	36	37	38
162.5	17	18	18	19	20	21	22	23	24	24	25	26	27	28	29	30	31	31	32	33	34	35	36	37
165.1	16	17	18	19	20	20	21	22	23	24	25	25	26	27	28	29	30	30	31	32	33	34	35	36
167.6	16	17	17	18	19	20	21	21	22	23	24	25	25	26	27	28	29	29	30	31	32	33	34	34
170.1	15	16	17	18	18	19	20	21	22	22	23	24	25	25	26	27	28	29	29	30	31	32	33	33
172.7	15	16	16	17	18	19	19	20	21	22	22	23	24	25	25	26	27	28	28	29	30	31	32	32
175.2	14	15	16	17	17	18	19	20	20	21	22	22	23	24	25	25	26	27	28	28	29	30	31	31
177.8	14	15	15	16	17	18	18	19	20	20	21	22	23	23	24	25	25	26	27	28	28	29	30	30
180.3	14	14	15	16	16	17	18	18	19	20	21	21	22	23	23	24	25	25	26	27	28	28	29	30
182.8	13	14	14	15	16	17	17	18	19	19	20	21	21	22	23	23	24	25	25	26	27	27	28	29
185.4	13	13	14	15	15	16	17	17	18	19	19	20	21	21	22	23	23	24	25	25	26	27	27	28
187.9	12	13	14	14	15	16	16	17	18	18	19	19	20	21	21	22	23	23	24	25	25	26	27	27
190.5	12	13	13	14	15	15	16	16	17	18	18	19	20	20	21	21	22	23	23	24	25	25	26	26
193.0	12	12	13	14	14	15	15	16	17	17	18	18	19	20	20	21	22	22	23	23	24	25	25	26

W projekcie zostały wykorzystane 3 klasyfikatory, czyli K-najbliższych sąsiadów, Random Forest oraz Drzewo Decyzyjne.

- KNN - Algorytm K Najbliższych Sąsiadów (ang. KNN) jest najprostszym i najwolniejszym algorytmem klasyfikacyjnym. Problematyczny staje się przy dużych zestawach danych. Polega na znalezieniu **k** elementów już sklasyfikowanych (sąsiadów) najbliższych nowemu elementowi i przydzieleniu tego elementu do grupy, do której należy większość jego sąsiadów. Używany jest w statystyce do prognozowania wartości pewnej zmiennej losowej. Może również być używany do klasyfikacji.

- Random Forest - to metoda klasyfikacji (oraz regresji) polegająca na tworzeniu wielu drzew decyzyjnych na podstawie losowego zestawu danych. Idea tego algorytmu polega na zbudowaniu konsylium ekspertów z losowych drzew decyzyjnych, gdzie w odróżnieniu do klasycznych drzew decyzyjnych, losowe drzewa budowane są na zasadzie, iż podzbiór analizowanych cech w węźle dobierany jest losowo. Ponadto, poszczególne drzewa z losowych lasów drzew budowane są zgodnie z koncepcją Bugging. Najlepsze cechy tego algorytmu to: skuteczne działanie na dużych bazach, jest najlepszy jeśli chodzi o dokładność wśród pozostałych algorytmów, utrzymuje dokładność w przypadku braku danych, daje oszacowanie, które zmienne są istotne w klasyfikacji.
- Drzewo Decyzyjne - to nieparametryczny algorytm uczenia nadzorowanego, który jest wykorzystywany zarówno do zadań klasyfikacji, jak i regresji. Ma hierarchiczną, drzewiastą strukturę, która składa się z węzła głównego, gałęzi, węzłów wewnętrznych i węzłów liści. Konstrukcja drzewa decyzyjnego: Drzewa można się „nauczyć”, dzieląc zbiór źródłowy na podzbiory na podstawie testu wartości atrybutów. Ten proces jest powtarzany na każdym wyprowadzonym podzbiorku w sposób rekurencyjny zwany partycjonowaniem rekurencyjnym. Rekursja jest zakończona, gdy podzbiór w węźle ma taką samą wartość zmiennej docelowej lub gdy dzielenie nie dodaje już wartości do prognoz. Konstrukcja klasyfikatora drzewa decyzyjnego nie wymaga żadnej wiedzy o domenie ani ustawienia parametrów, dlatego jest odpowiednia do eksploracyjnego odkrywania wiedzy. Drzewa decyzyjne mogą obsługiwać dane wielowymiarowe. Ogólnie klasyfikator drzewa decyzyjnego ma dobrą dokładność. Indukcja drzewa decyzyjnego jest typowym indukcyjnym podejściem do uczenia się wiedzy na temat klasyfikacji.

### 3 Specyfikacja wewnętrzna

W projekcie wykorzystałem dane, które były używane na zajęciach i były wstępnie oczyszczone, lecz nadal musiałem w pewnym stopniu te dane uporządkować, aby pozostały tylko te, które będą potrzebne. Wyizolowałem więc te kolumny które będą mi głównie potrzebne. Następnie pozbyłem się wszelkich NA występujących w danych oraz dodałem kolumnę z obliczonymi wartościami BMI.

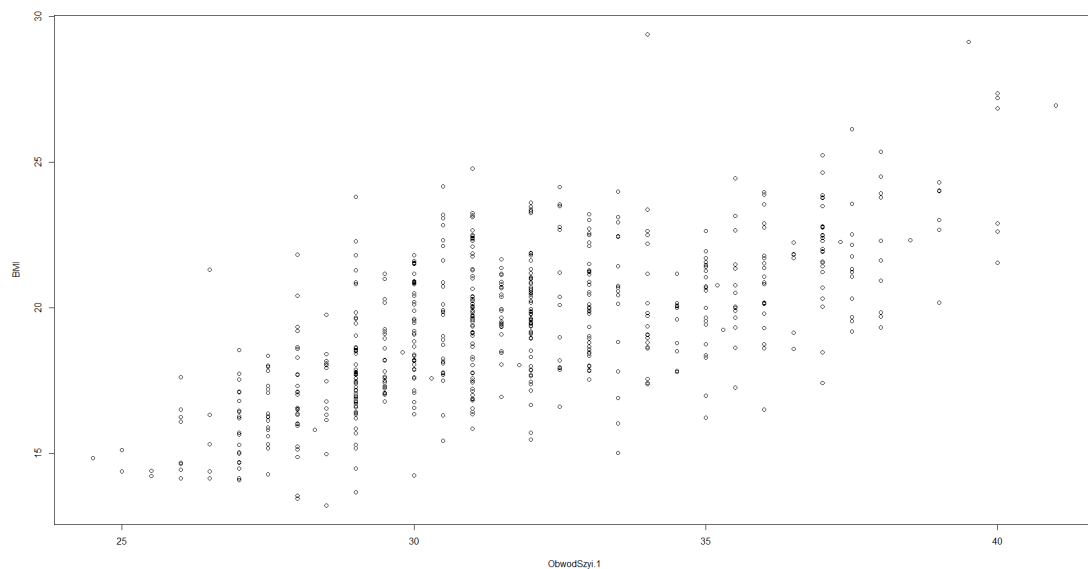
```
#stworzenie kolumny BMI
wysokosc = c(dane$wysokosc.1/100)
waga = c(dane$weight.1)
BMI = waga/(wysokosc*wysokosc)
dane["BMI"] = BMI
```

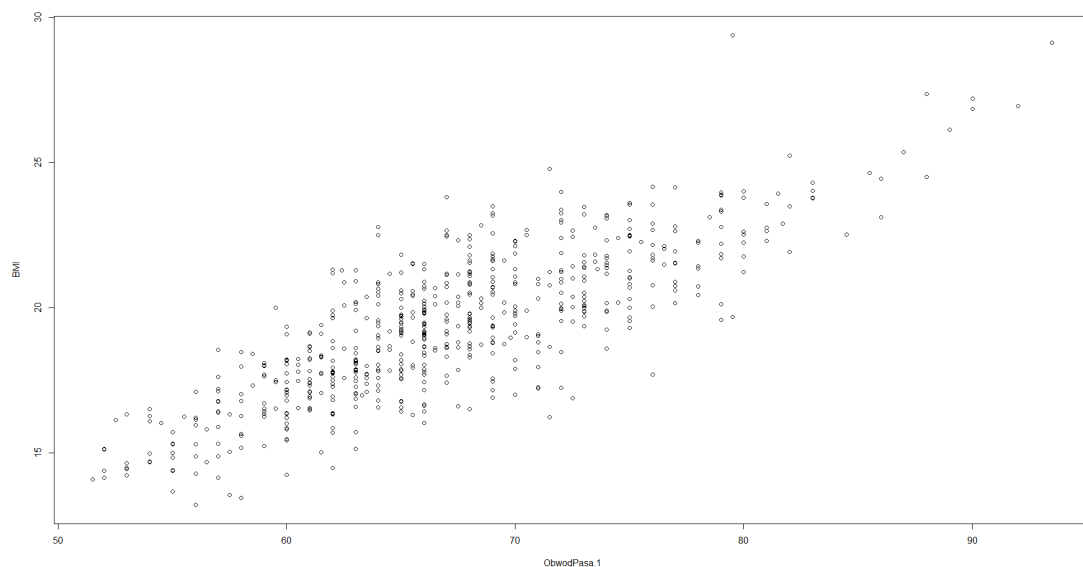
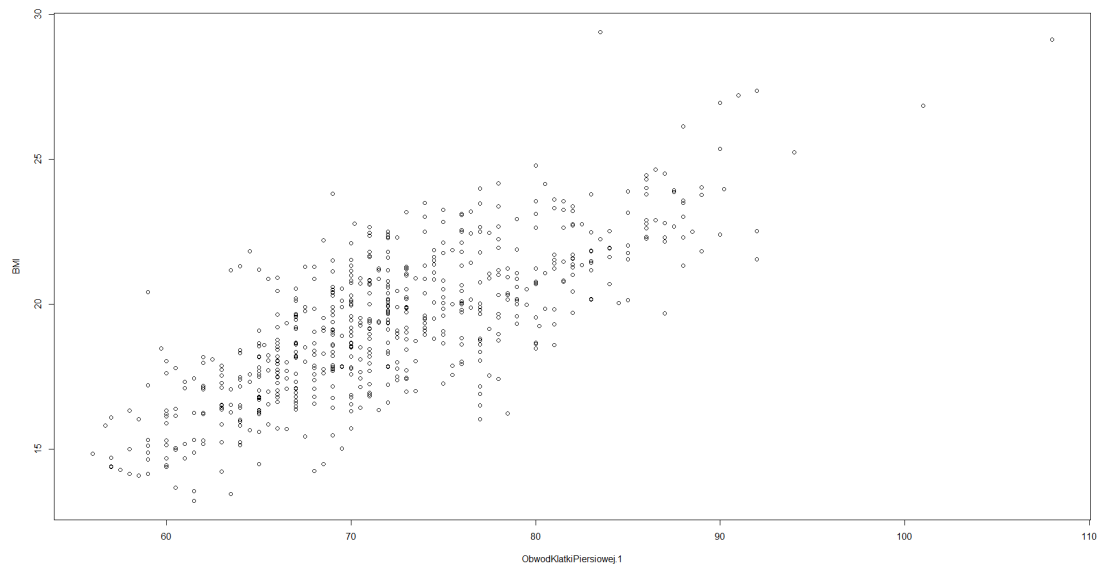
W kolejnym kroku dodałem kolumnę BMIlevel, w której na podstawie zwykłego BMI określałem czy dana osoba ma niedowagę, wagę prawidłową, nadwagę, czy też otyłość.

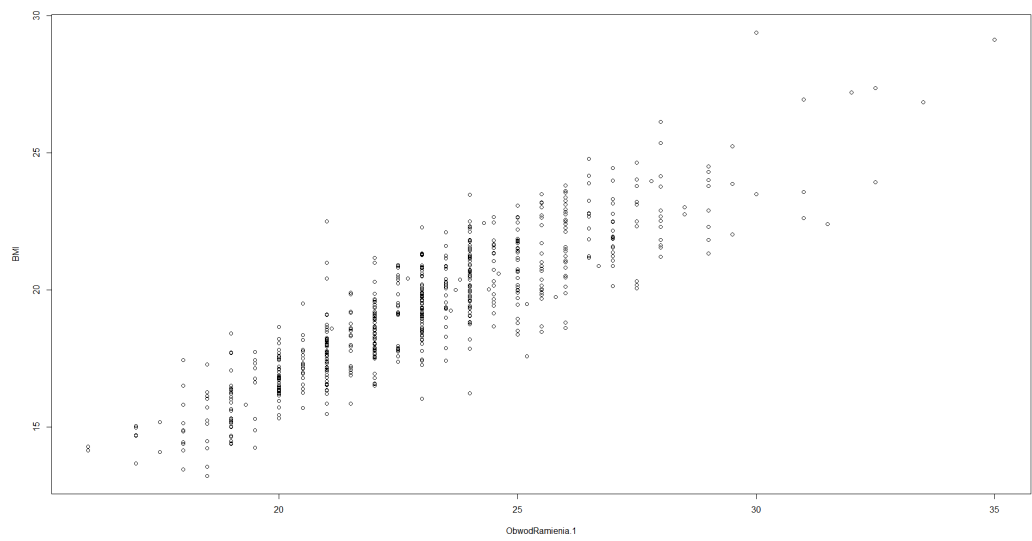
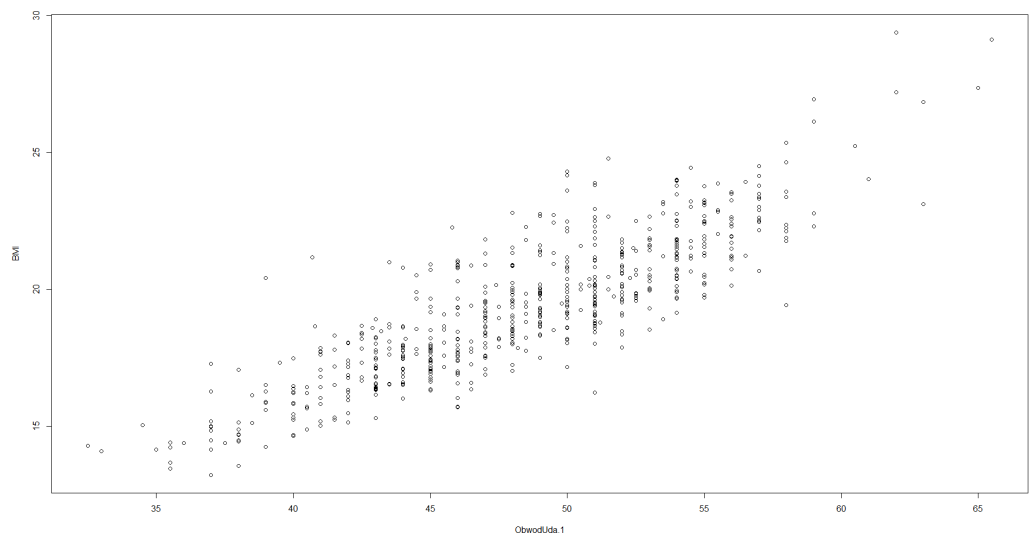
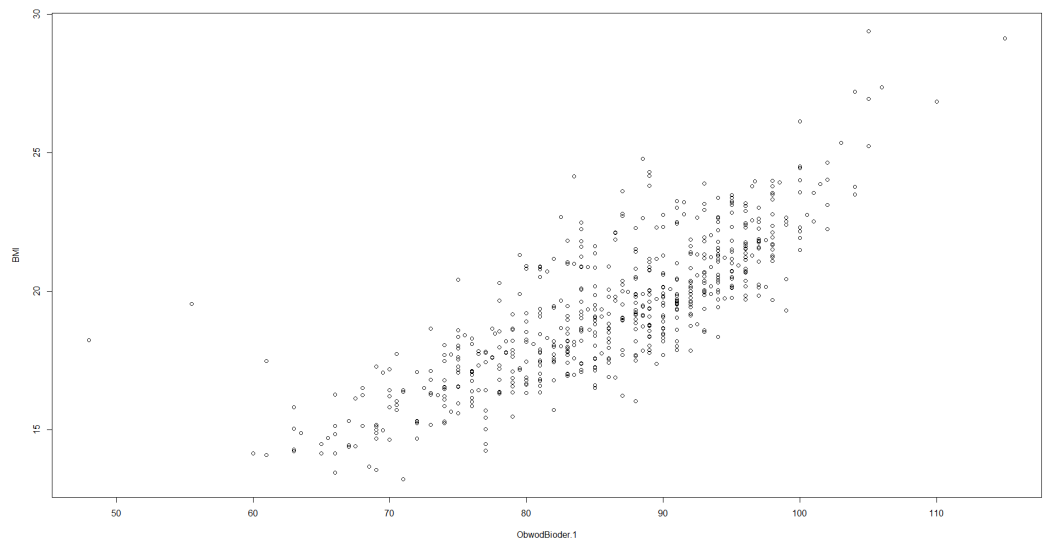
```
#przypisanie BMI do 4 wartości
#-1 = niedowaga
#0 = waga prawidłowa
#1 = nadwaga
#2 = otyłość
columnBMI=13
nrow(dane3)
BMIlevel = data.frame()
for(w in c(1:nrow(dane3)))
{
  record = dane3[w,columnBMI]
  record
  if(record < 18.5)
  {
    BMIlevel <- rbind(BMIlevel, -1)
  }
  else if(18.5 <= record & record <= 24.9)
  {
    BMIlevel <- rbind(BMIlevel, 0)
  }
  else if(24.9 < record & record <= 29.9)
  {
    BMIlevel <- rbind(BMIlevel, 1)
  }
  else if(29.9 < record & record <= 39.9)
  {
    BMIlevel <- rbind(BMIlevel, 2)
  }
}
```

Dalej używam funkcji oceniającej, która przedstawiana była podczas zajęć praktycznych.

Następnie stworzyłem parę wykresów przedstawiające zależność pomiędzy danymi obwodami oraz BMI.





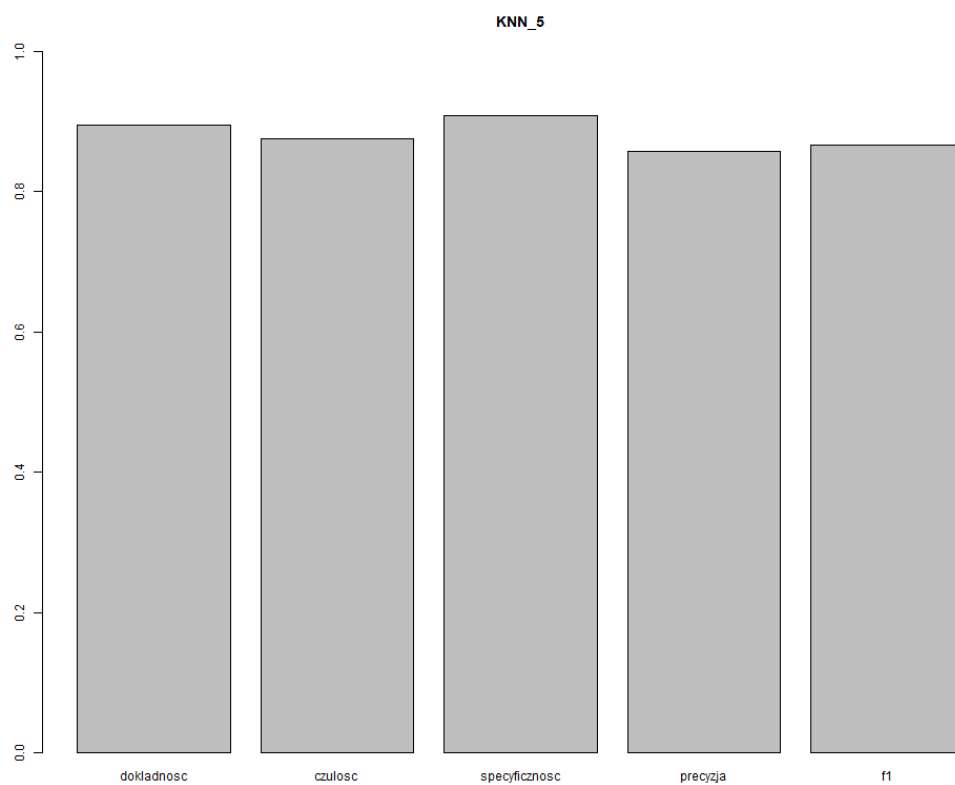


Z tych wykresów można łatwo zaobserwować, że BMI rośnie wprost proporcjonalnie do to coraz to większego wyniku obwodu danej części ciała. Widać też, że wiele wartości jest bardzo blisko siebie co oznacza, że obwód danej części ciała, nie ma aż tak wielkiego wpływu na BMI.

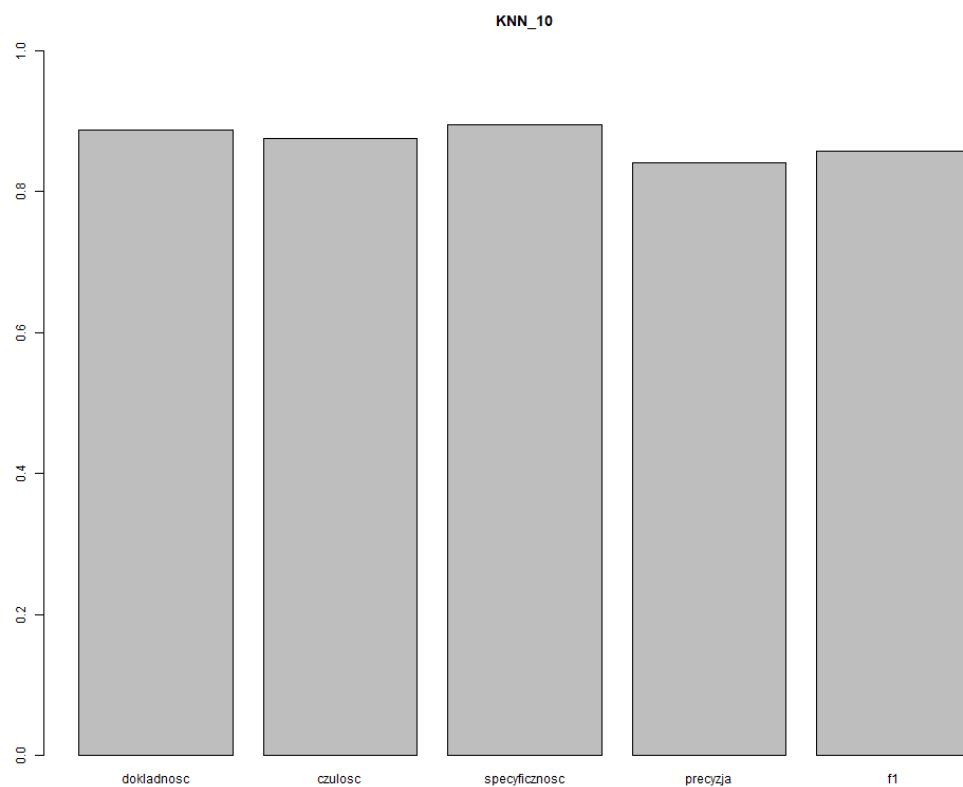
Na koniec rozpisałem klasyfikatory, KNN, Random Forest oraz Drzewo Decyzji.

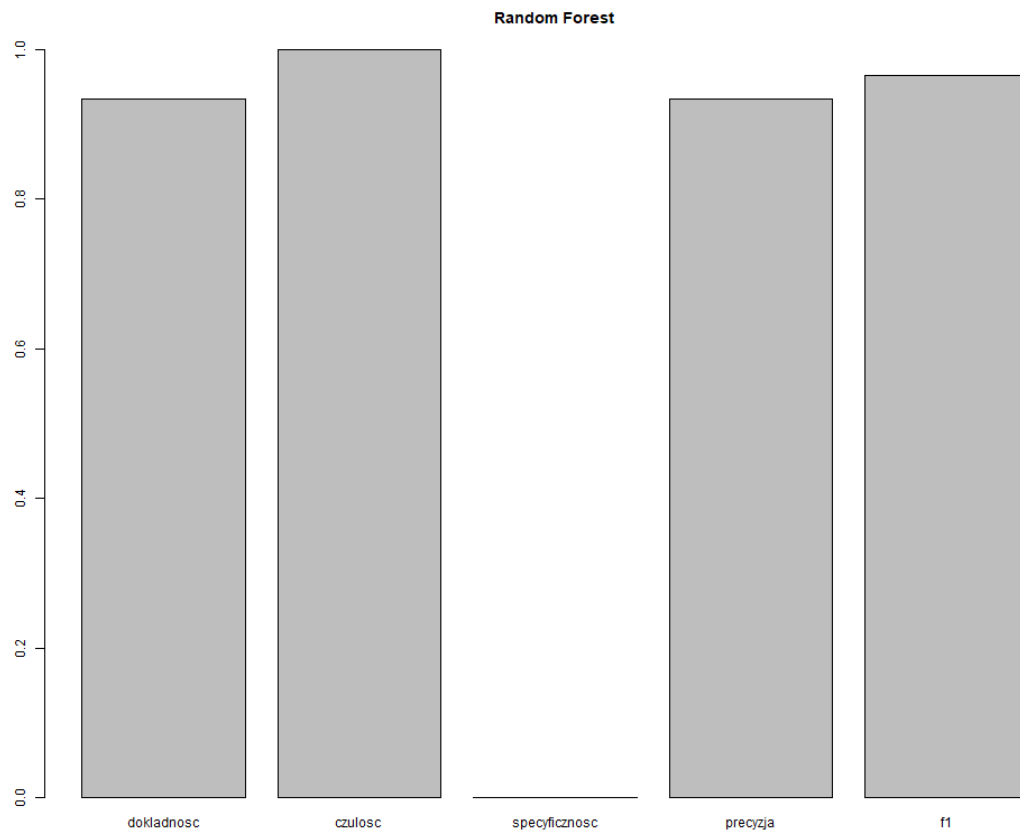
## 4 Wyniki

Poniższe dwa wykresy przedstawiają skuteczność działania klasyfikatora KNN. Testy w KNN były przeprowadzane dla k sąsiadów: 5,10. Różnice w obu tych wykresach są praktycznie znikome.

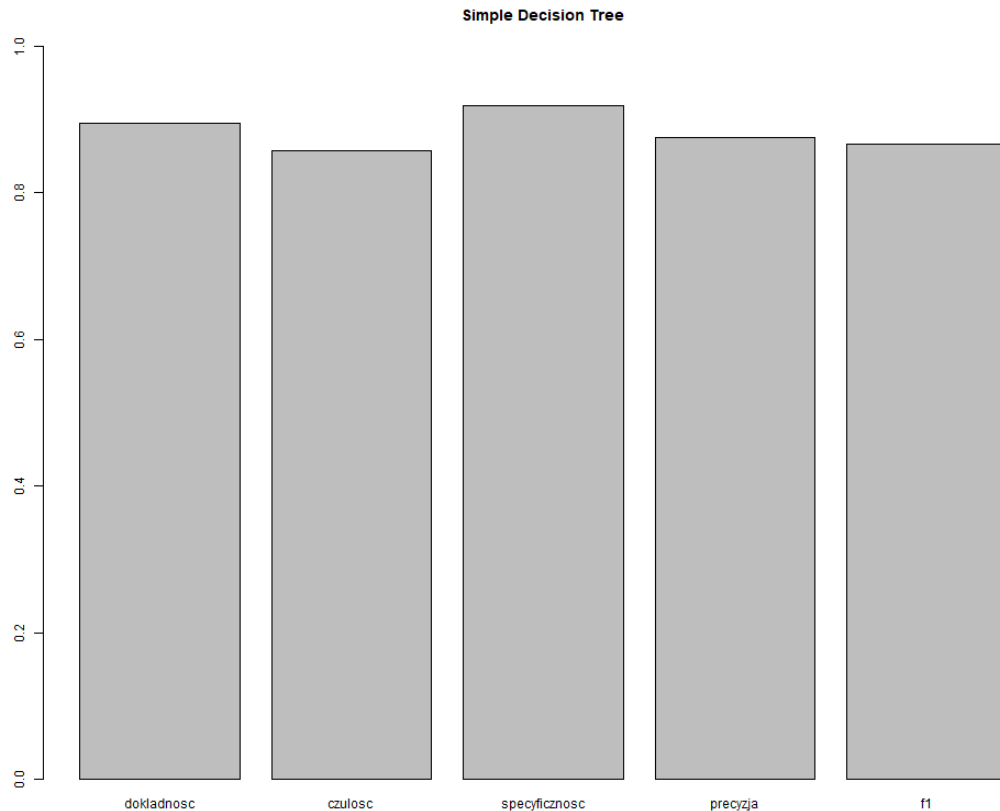








Jak widać na powyższym wykresie Random Forest w dokładności jest nieco lepszy niż KNN, a samą czułość ma na poziomie 100%.



Drzewo decyzyjne wypadło porównywalnie dobrze jak i również KKN ale pod względem dokładności jest nadal gorszy niż Random Forest.

## 5 Wnioski oraz podsumowanie

Głównym założeniem projektu było sprawdzenie z jaką dokładnością można określić czy dana osoba ma niedowagę, wagę prawidłową, nadwagę, czy otyłość na podstawie samych obwodów paru części ciała. Najlepszą dokładność w tym zagadnieniu udało się otrzymać korzystając z Random Foresta. KNN oraz Drzewo Decyzyjne miały niewiele różniącą się dokładność, więc były bardzo do siebie zbliżone, lecz nadal miały gorszą dokładność niż Random Forest.

Na podstawie wykresów można stwierdzić, że ze wzrostem obwodów takich części ciała jak: szyja, klatka piersiowa, pas, biodra, udo oraz ramię, BMI zwiększa się wprost proporcjonalnie.

## Literatura

- [1] <https://www.festiwalbiegowy.pl/trening/jak-dokonywac-pomiaru-obwodow-ciala-sledzic-postepy-w-treningu-0#.YzBz2nZBxE>
- [2] <https://aszokalski.github.io/AI/KNN.html>
- [3] <https://www.bmi-kalkulator.pl/wzor>
- [4] <https://www.ikard.pl/badanie-bmi.html>
- [5] <https://ekoj.pl/co-to-jest-to-bmi-i-do-czego-sluzyc-jak-czesto-trzeba-sobie-obliczac-bmi,87.html>
- [6] <https://docplayer.pl/9239024-Algorytm-random-forest.html>
- [7] <https://www.geeksforgeeks.org/decision-tree/>
- [8] <https://www.ibm.com/topics/decision-trees>