



Dokumentacja projektowa

Inżynieria Biomedyczna w Praktyce

Bartosz Rolnik

Kierunek: *Informatyka*

Wydział: *Matematyki Stosowanej*

Rok akademicki *2021/2022*

Tytuł: *Kalkulator masy ciała, masy mięśniowej, masy tłuszczu, procentu tłuszczu w organizmie oraz BMI na podstawie różnych obwodów ciała, wzrostu i płci*

1 Wstęp

Projekt zakłada utworzenie prostego graficznego kalkulatora danych inbody na podstawie danych antropometrycznych. Źródłem danych były dwa pliki opracowywane w przeciągu piętnastu zajęć laboratoryjnych. Całość zagadnienia została podzielona na dwie części:

- Dodatkowe edycje plików CSV oraz testowanie obliczeń matematycznych potrzebnych później. Ten fragment był pisany przy pomocy interaktywnego środowiska obliczeniowego Jupyter Notebook. W skład edycji plików wchodziło usuwanie zbędnych kolumn, łączenie plików z danymi antropometrycznymi oraz inbody, a następnie dzielenie ich według płci.

- Właściwa część projektu pisana w języku python. Wprowadzając odpowiednie wartości danych antropometrycznych w prostym menu, takich jak: płeć, wzrost oraz obwody: szyi, ramion, klatki, bioder, uda i pasa wyświetli nam się przewidywana waga osobnika oraz średni przewidywany błąd bezwzględny (wyrażony w tej samej jednostce co wynik). W każdej chwili liczenie wagi można zmienić na inne wartości inbody takie jak: procent tłuszczu w organizmie, masę tłuszczu w organizmie, masę mięśni w organizmie, BMI. Wszystkie wyniki są liczone przy użyciu regresji liniowej wielu zmiennych na zbiorze treningowym. Średni błąd bezwzględny sprawdzany jest na zbiorze walidacyjnym.

2 Analiza zagadnienia

Regresja - metoda pozwalająca na modelowanie związku pomiędzy zmiennymi i wykorzystanie tej wiedzy do przewidywania nieznanych wartości na podstawie znajomości innych. W przypadku mojej pracy zdecydowałem się na użycie regresji liniowej wielu zmiennych. Opis funkcji liniowej dla takiej regresji prezentuje się w następujący sposób:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + e$$

Gdzie:

i – numer rekordu

y – zmienna zależna,

β_0 – stała przecinająca y

β_n – współczynnik nachylenia dla zmiennej objaśniającej o indeksie n

x_n – zmienna objaśniająca o indeksie n

e – składnik błędu modelu

Obsługa plików CSV była możliwa dzięki bibliotece pandas. Pierwszym krokiem w sprawie przygotowywania bazy danych było odpowiednie sklejenie plików „inbody” i „antropo”, a następnie usuwanie kolejnych kolumn, które zostały uznane za zbędne. W danych antropometrycznych zdecydowano się zostawić jedynie płeć, wzrost oraz obwody: szyi, ramion, klatki, bioder, uda i pasa. Reszta statystyk została uznana za niezwiązaną z ideą projektu. Następny w kolejności był plik danych inbody, z którego usunięto wszystkie kolumny poza tymi zaznaczonymi w temacie, czyli: masa ciała, masa mięśniowa, BMI, masa tłuszczu i jego procentowy udział w organizmie. W kolejnym kroku usunięto wszelkie linie, w których pojawiły się braki danych. Nie zdecydowano się na zastąpienie poszczególnych komórek przy pomocy jakiejś konwencji, aby być pewnym, że ewentualne błędy obliczeniowe nie wynikają z wybranego sposobu uzupełniania danych.

Komunikacja między plikiem edytującym bazę oraz właściwą aplikacją odbywała się dzięki dwóm plikom zawierającym dane nastolatków. Zostały one podzielone według płci, jednak dane inbody i antropometryczne zostały sklejone dla danej płci do jednej paczki danych.

Posiadając oba pliki, można było rozpocząć budowanie modelu regresji liniowej wielu zmiennych przy pomocy biblioteki sklearn w pythonie. Zanim model został zbudowany, pliki zostały podzielone na zbiór treningowy i walidacyjny w proporcjach 70:30. Po zbudowaniu regresji można było przetestować o ile nasz model się myli względem realnych danych, które posiadano i dla każdego rodzaju zmiennej został wyliczony średni błąd bezwzględny sprawdzany na zbiorze walidacyjnym.ⁱ

Graficzne GUI zostało przygotowane dzięki bibliotece tkinter, która pokazały prosto, czytelnie i mile dla oka przygotowane grafiki dla programu. Pozwoliła ona na zbudowanie ośmiu wejść dla danych antropometrycznych, kilku przycisków oraz dwóch wyjść z wynikami, które są wyświetlane na podstawie naszego modelu. Każdorazowe włączenie programu spowoduje wylosowanie się zbiorów treningowych i walidacyjnych, więc każdorazowo wyniki mogą się nieznacznie różnić.

3 Specyfikacja wewnętrzna

Plik Jupyter Notebook:

Plik głównie do czyszczenia bazy danych i ewentualnych testów. Zawiera zaledwie jedną klasę o nazwie „Project”. Jest w niej funkcja do usuwania kolumn w pandas.dataframe po otrzymaniu listy z nazwami kolumn. Reszta funkcji służy do testowania metod matematycznych potrzebnych do pliku z właściwą częścią projektu. Całość tego programu do pewnego momentu to usuwanie niepotrzebnych kolumn oraz czyszczenie i analiza danych, później jest kilka próbnych linijek użycia metod z biblioteki sklearn.

Bazy danych:

Podstawowe informacje dla bazy osób płci żeńskiej i męskiej (różnica występuje tylko w liczbie rekordów)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 165 entries, 0 to 164
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Wysokosc                             165 non-null    float64
1   ObwodSzyi                             165 non-null    float64
2   ObwodKlatkiPiersiowej                 165 non-null    float64
3   ObwodPasa                             165 non-null    float64
4   ObwodBioder                           165 non-null    float64
5   ObwodUda                              165 non-null    float64
6   ObwodRamienia                         165 non-null    float64
7   Weight                                165 non-null    float64
8   BFM_BodyFatMass_                      165 non-null    float64
9   SMM_SkeletalMuscleMass_               165 non-null    float64
10  BMI_BodyMassIndex_                    165 non-null    float64
11  PBF_PercentBodyFat_                   165 non-null    float64
dtypes: float64(12)
memory usage: 15.6 KB
```

- Baza kobiet (164 rekordy):

Baza składa się z danych antropometrycznych: Wysokość, obwód szyi, obwód ramion, obwód klatki, obwód bioder, obwód uda, obwód pasa oraz danych inbody: masa ciała, masa mięśniowa, BMI, masa tłuszczu i procentowy udział tłuszczu w organizmie.

	Wysokosc	ObwodSzyi	ObwodKlatkiPiersiowej	ObwodPasa	ObwodBioder	ObwodUda	ObwodRamienia	Weight	BFM_BodyFatMass_	SMM_SkeletalMuscleMass_	BMI_BodyMassIndex_	PBF_PercentBodyFat_
count	165.000000	165.000000	165.000000	165.000000	165.000000	165.000000	165.000000	165.000000	165.000000	165.000000	165.000000	165.000000
mean	156.953939	30.370303	68.334545	64.157576	85.955758	48.464242	22.417576	47.886667	10.201212	20.352121	19.206667	20.399394
std	11.217123	2.105169	5.349249	5.304970	8.982560	5.238366	2.076405	10.236106	4.589968	3.928967	2.298333	6.018554
min	127.800000	25.000000	56.700000	52.000000	63.500000	36.000000	18.000000	25.700000	1.200000	11.300000	14.400000	3.600000
25%	148.900000	29.000000	65.000000	61.000000	79.000000	45.000000	21.000000	39.200000	7.000000	17.200000	17.700000	16.700000
50%	159.500000	30.500000	68.000000	65.000000	89.000000	49.000000	22.500000	50.400000	10.000000	21.300000	19.300000	20.200000
75%	165.700000	32.000000	71.000000	67.500000	93.000000	52.500000	24.000000	55.400000	13.000000	23.200000	20.600000	24.100000
max	185.000000	35.500000	92.000000	82.000000	105.000000	62.000000	30.000000	75.700000	29.500000	27.800000	29.400000	38.900000

- Baza mężczyzn (203 rekordy):

Pod względem struktury bazy danych nie różni się ona niczym od bazy kobiet.

	Wysokosc	ObwodSzyi	ObwodKlatkiPiersiowej	ObwodPasa	ObwodBioder	ObwodUda	ObwodRamienia	Weight	BFM_BodyFatMass_	SMM_SkeletalMuscleMass_	BMI_BodyMassIndex_	PBF_PercentBodyFat_
count	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000
mean	165.675490	33.374020	74.585784	69.114216	86.519608	47.627451	23.628431	54.968137	7.674510	26.152941	19.636765	14.128922
std	15.348355	3.398603	8.005001	7.956108	9.497192	5.938961	3.379360	14.823331	3.980858	8.162219	2.750508	6.177202
min	126.900000	26.000000	56.000000	50.000000	60.000000	32.500000	16.000000	24.100000	1.700000	10.000000	13.700000	3.700000
25%	155.550000	30.875000	69.000000	64.000000	81.000000	43.875000	21.000000	44.400000	4.875000	19.950000	17.800000	9.600000
50%	168.250000	33.000000	74.000000	68.000000	86.000000	48.000000	23.500000	54.000000	6.650000	25.450000	19.550000	12.800000
75%	177.200000	36.000000	80.625000	74.000000	94.000000	52.000000	25.500000	64.850000	9.800000	32.300000	21.225000	18.350000
max	196.000000	41.000000	94.000000	92.000000	106.500000	65.000000	32.500000	91.200000	22.000000	44.500000	27.400000	31.600000

Różnice liczbowe:

- Porównywalne: obwód klatki piersiowej, obwód bioder, obwód uda, obwód ramienia, BMI
- Zdecydowane: wysokość, obwód szyi, masa, procent tłuszczu w organizmie, masa tłuszczu, masa mięśniowa

Plik Python:

Program posiada cztery uporządkowane klasy: Check, Widgets, Predictor oraz ProcessingData.

- **Klasa Check:** posiada cztery metody, z czego dwie służą do poprawnej walidacji danych wpisywanych w wejścia:

- Walidacja danych liczbowych: nie pozwala na wpisywanie liczb ujemnych, liter oraz zbyt długich liczb,
- Walidacja płci: dopuszcza tylko wpisanie „k” lub „m”,
- Następna funkcja o nazwie „checkoptions” sprawdza jedynie, która z rubryk inbody została wybrana i zwraca nam jej index,
- Ostatnia metoda nazwana „isResultPossible” odpowiada za sprawdzenie, czy we wszystkich wejściach zostały wprowadzone dane. W przypadku braku danych program stosownie o tym poinformuje.

- **Klasa Widgets:** posiada pięć funkcji. Każda z nich odpowiada za inny typ widgetu, który jest wyświetlany na ekranie. Kolejno odpowiadają za zbudowanie: wejść na dane, etykiety dla wejść i wyjść, okno wyboru danych inbody do wyliczenia, potrzebne przyciski oraz etykiety wyjścia, które mają wstępne sprawdzenie wyniku.

- **Klasa Predictor:** Najważniejsza klasa projektu. Posiada metody odpowiadające za liczenie dopasowanej funkcji liniowej dla obu płci, wyliczanie wartości dla zmiennej zależnej oraz liczenie średniego błędu bezwzględnego na zbiorach walidacyjnych.

```
@staticmethod
def calculate(plec: str, wysokosc: float, obwodSzyi: float, obwodKlatkiPiersiowej: float, obwodPasa: float,
             obwodBioder: float, obwodUda: float, obwodRamienia: float):
    predicted = 'Brak danych'
    index = Check.checkoptions()
    if plec == 'k':
        predicted = regsFemale[index].predict(
            [[wysokosc, obwodSzyi, obwodKlatkiPiersiowej, obwodPasa, obwodBioder, obwodUda, obwodRamienia]])
    elif plec == 'm':
        predicted = regsMale[index].predict(
            [[wysokosc, obwodSzyi, obwodKlatkiPiersiowej, obwodPasa, obwodBioder, obwodUda, obwodRamienia]])
    if predicted >= 0:
        Widgets.resultLabel(str(predicted)[1:6], plec)
    else:
        Widgets.resultLabel('Złe dane', 'brak')

# zbiorzy dla regresji
@staticmethod
def makeregs(dataT):
    regs = [] # kolejność: Waga, Masa tłuszczu, Masa mięśniowa, BMI, Procent Tłuszczu

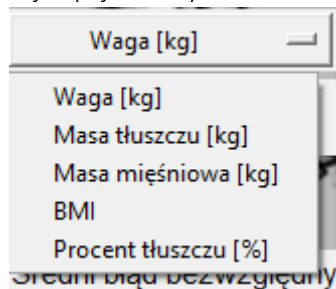
    # tworzenie modelu regresji liniowej wielu zmiennych dla kobiet
    for column in dataT.columns.tolist()[7:]:
        reg = linear_model.LinearRegression()
        # tworzenie zbioru treningowego z inbody kobiet dla danych zmiennych
        reg.fit(dataT[['Wysokosc', 'ObwodSzyi', 'ObwodKlatkiPiersiowej', 'ObwodPasa', 'ObwodBioder', 'ObwodUda',
                       'ObwodRamienia']], dataT[column])
        regs.append(reg)

    return regs
```

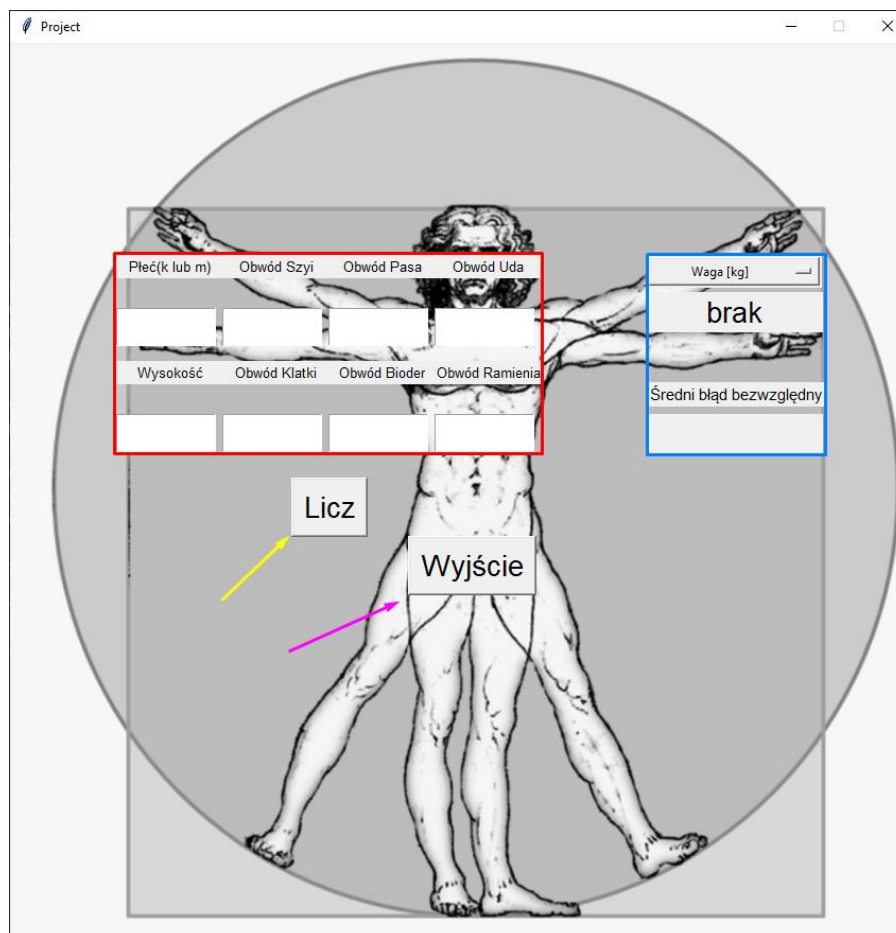
Inicjacja programu: ostatni fragment kodu odpowiada za uruchomienie programu oraz odpowiednich metod startowych.

4 Specyfikacja zewnętrzna

- Czerwony obszar: miejsce do wpisywania odpowiednich danych.
- Niebieski obszar: miejsce, w którym pojawia się wynik i średni błąd bezwzględny. W miejscu wyniku może pojawić się komunikat o jego braku (jak na ekranie poniżej), wprowadzeniu złych danych, niekompletnych danych lub właściwej liczby zmiennoprzecinkowej. Opcje do wyboru:



- Żółta strzałka: przycisk uruchamiający liczenie.
- Różowa strzałka: przycisk wyłączający aplikację.



5 Testowanie

- 1 W przypadku nieuzupełnienia jednej lub większej ilości komórek, program w miejscu wyniku wyświetli komunikat „Brak danych”:

The screenshot shows the application interface with the Vitruvian Man background. The input fields are populated with the following values:

Płeć(k lub m)	Obwód Szyi	Obwód Pasa	Obwód Uda	Waga [kg]
m	30	80	50	

The output field displays "Brak danych". The "Licz" button is visible, and the "Wyjście" label is present.

- 2 Program ma zaimplementowany prosty system służący do wychwytywania błędnych danych wejściowych:
 - W przypadku wprowadzania prawidłowych danych, ale pozostawienia co najmniej jednej komórki z zerową wartością program wyświetli komunikat „Złe dane”

The screenshot shows the application interface with the Vitruvian Man background. The input fields are populated with the following values:

Płeć(k lub m)	Obwód Szyi	Obwód Pasa	Obwód Uda	Masa Muszku [kg]
m	30	100	45	

The output field displays "Złe dane". The "Licz" button is visible, and the "Wyjście" label is present.

- W przypadku wprowadzenia nieprawdopodobnych liczb takich jak człowiek mający 4 cm wzrostu, obwód szyi mierzący 3 cm oraz obwód klatki mający 2 cm program również wyświetli komunikat „Złe dane”:

Płeć(k lub m)	Obwód Szyi	Obwód Pasa	Obwód Uda	Masa mięśniowa [kg]
m	3	80	50	Złe dane
Wysokość	Obwód Klatki	Obwód Bioder	Obwód Ramienia	Średni błąd bezwzględny
4	2	80	40	

Licz

Wyjście

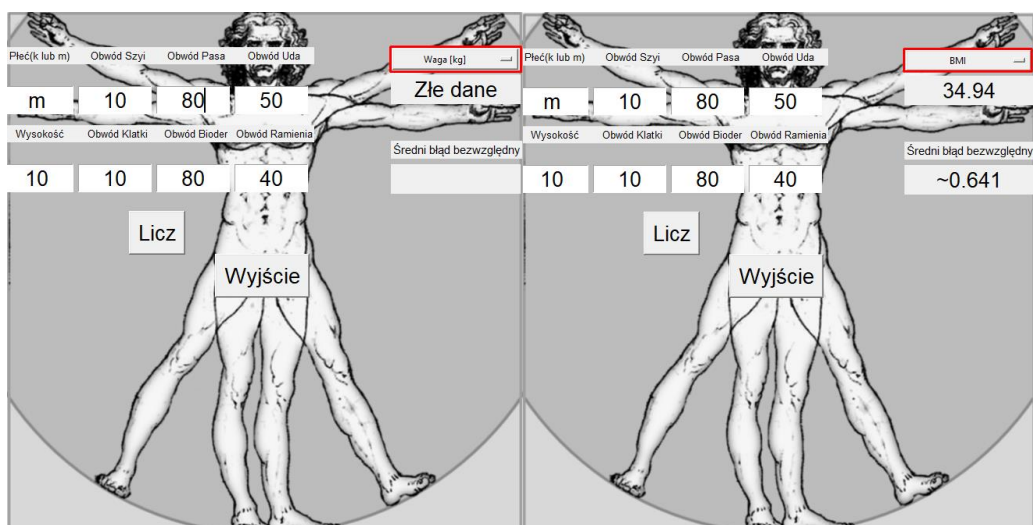
- Podczas testowania zostały odkryte mankamenty tak prostego systemu wykrywania błędów:
 - Często w przypadku wprowadzenia zaledwie jednej nieprawdopodobnej zmiennej program nie wychwytywał błędu:

Płeć(k lub m)	Obwód Szyi	Obwód Pasa	Obwód Uda	Waga [kg]
m	40	80	50	24.84
Wysokość	Obwód Klatki	Obwód Bioder	Obwód Ramienia	Średni błąd bezwzględny
10	90	80	40	~1.967

Licz

Wyjście

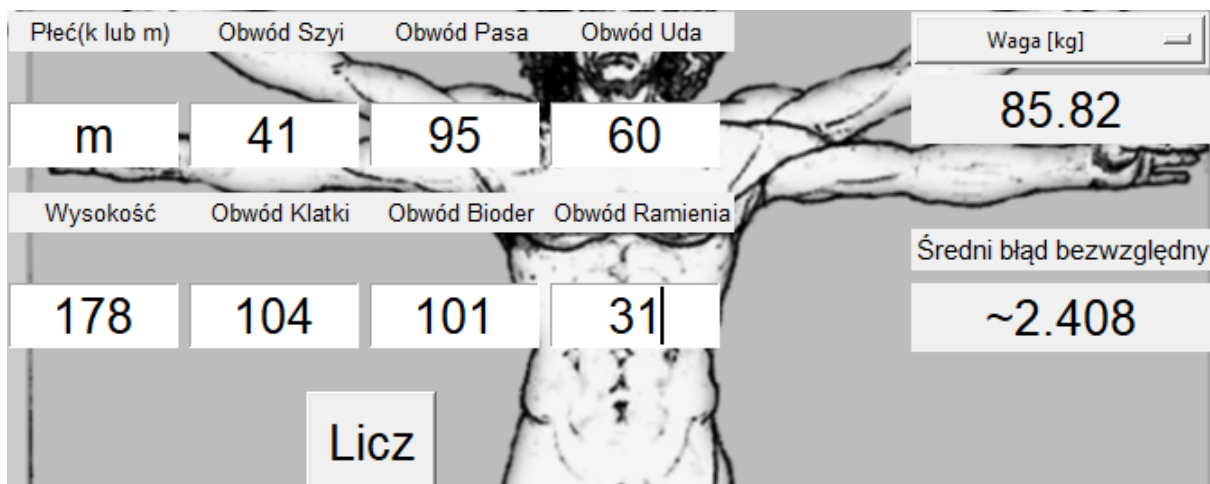
- Komunikat „Złe dane” odnosi się jedynie do szczególnego typu wyniku takiego jak „waga”. Dla innych danych typu inbody wynik może być odbiegający od normy:





6 Wyniki



Wyliczenia w tym rozdziale zostały wykonane na podstawie rzeczywistego obiektu i porównane przy pomocy szczegółowej elektronicznej wagi.

Dane wprowadzone do aplikacji:



Pozostałe wyniki dla tych samych danych:

Masa tłuszczu [kg]	Masa mięśniowa [kg]
16.84	39.25
	
Średni błąd bezwzględny	Średni błąd bezwzględny
~1.758	~1.677

BMI	Procent tłuszczu [%]
26.91	20.38
	
Średni błąd bezwzględny	Średni błąd bezwzględny
~0.659	~2.819

Realne wartości:

- Masa: 85.5 kg
- Masa mięśniowa: 38.1 kg
- Masa tłuszczu: 14.6 kg
- BMI: 26.99
- % tłuszczu w organizmie: 17.1 %

Błędy:

- Masa: $85.82 - 85.5 = 0.32\text{kg}$, co stanowi $\sim 0.38\%$ realnej wartości (znajduje się w granicach wyświetlanego błędu)
- Masa mięśniowa: $39.25 - 38.1 = 1.15\text{ kg}$ co stanowi $\sim 3\%$ realnej wartości (znajduje się w granicach wyświetlanego błędu)
- Masa tłuszczu: $16.84 - 14.6 = 2.24\text{ kg}$, co stanowi $\sim 15,3\%$ realnej wartości (nie znajduje się w granicach wyświetlanego błędu)
- BMI: $26.99 - 26.91 = 0.08$, co stanowi $\sim 0.29\%$ realnej wartości (znajduje się w granicach wyświetlanego błędu)
- % tłuszczu w organizmie: $20.38 - 17.1 = 3.28$, co stanowi 19.2% realnej wartości (nie znajduje się w granicach wyświetlanego błędu)

7 Wnioski

Analizując wyniki przedstawione w poprzednim rozdziale kalkulator radzi sobie różnie:

- Masa: w tym przypadku kalkulator wyliczył prawie perfekcyjnie masę obiektu myląc się o zaledwie 0.32 kg. Jak można zaobserwować po błędach wyświetlanych w programie jest to drugi najdokładniejszy pomiar.
- Masa mięśniowa: tutaj kalkulator napotkał lekkie problemy, ale wciąż były one zadowalające. Pomylił się o 1.15 kg, co stanowiło zaledwie około 3% realnej masy i mieściło się wciąż w granicach błędu.
- BMI: najłatwiejsza do wyliczenia szukana, ponieważ spora część danej jest już podana przy pomocy wzrostu, a dodatkowo jak można dostrzec w punkcie powyżej, masa jest również wyliczana bardzo dobrze, więc fenomenalny wynik nie dziwi.
- Masa tłuszczu: wynik zdecydowanie poniżej oczekiwań. Pomyłka zdecydowanie większa niż w poprzednich punktach, mimo mniejszej skali liczbowej.
- Procent tłuszczu w organizmie. Najgorzej liczona statystyka, co można zaobserwować już przy wyświetlanych błędach. Nie dość, że korzysta z drugiej najgorzej liczonej szukanej, to dodatkowo wkłada się tu błąd związany z wagą.

Elementy wpływające na błąd pomiarowy:

- niedokładność sprzętu pomiarowego (grupy badanej i obiektu)
- obiekt znacznie odbiegał wiekiem od grupy badanej, a co z tym się wiąże miał inną budowę i proporcje ciała
- rząd wielkości wszystkich danych był zdecydowanie mniejszy w przypadku grupy badanej i błąd wyświetlany w aplikacji mógłby być inny przy bardziej zróżnicowanej grupie,
- mały zbiór treningowy (dla mężczyzn wynosiło to zaledwie 70% z 203 rekordów)

8 Podsumowanie

Celem było wykonanie kalkulatora, który na podstawie grupy wyselekcjonowanych obwodów części ciała liczył wybrane cechy inbody. Dodatkowym zamiarem było stworzenie intuicyjnego interfejsu graficznego. Rezultatem jest aplikacja spełniająca wszystkie wyżej wymienione zamierzenia, która jednak ma problemy z dokładnością niektórych wyników.

Program można rozwijać w wielu kierunkach, w tym między innymi:

- Zwiększenie grupy badanej u ludzi w bardziej zróżnicowanym wieku i różnej budowie ciała, co spowoduje większą dokładność,

- Rozbudowanie algorytmu o inne cechy inbody takie jak: masa układu kostnego, masa poszczególnych części ciała oraz inne cechy inbody dla nich,
- Poprawienie strony graficznej programu,
- Wyeliminowanie danych abstrakcyjnych takich jak np. obwód szyi równy 10 cm,
- Dodanie różnych wykresów i tabel,
- Poprawa algorytmów liczenia błędów i rezultatów.

Źródła

- <https://matematyka.poznan.pl/artukul/regresja-liniowa-czyli-o-zastosowaniu-funkcji-liniowej-w-analizie-statystycznej/>
 - <https://www.investopedia.com/terms/m/mlr.asp>
 - https://pl.m.wikipedia.org/wiki/Regresja_liniowa
-