

POLITECHNIKA WROCŁAWSKA
WYDZIAŁ INFORMATYKI I TELEKOMUNIKACJI



Metody i Systemy Decyzyjne

Sprawozdanie z laboratorium

AUTOR

Jakub Roszkowski

nr albumu: **260282**

kierunek: **Informatyka Stosowana**

Analiza czy po rozmiarze dłoni można określić parametry fizyczne osoby

June 5, 2022

Contents

1	Wstęp	2
1.1	Wybór data setu oraz problemu do analizy	2
1.2	Źródło data setu	2
2	Preprocessing danych	3
2.1	Wybór danych poddanych analizie	3
2.2	Przetworzenie danych	3
2.3	Rozkłady danych	3
3	Rozwiązanie problemu	7
3.1	Opis problemu	7
3.1.1	Dopasowanie rozkładu do wartości długości dłoni	7
3.1.2	Dopasowanie modeli regresji: liniowej, GLM i SVR modelującego wzrost człowieka na podstawie jednego z parametrów dłoni	7
3.1.3	Dopasowanie modeli regresji: liniowej, GLM i SVR modelującego wagę człowieka na podstawie jednego z parametrów dłoni	7
3.1.4	Klasyfikacja płci na podstawie pomiarów dłoni	7
3.2	Podział danych testowych	7
4	Wykonywane eksperymenty z modelami / rozkładami	8
4.1	Dopasowanie rozkładu do wartości długości dłoni	8
4.2	Dopasowanie modeli regresji: liniowej, GLM i SVR modelującego wzrost człowieka na podstawie jednego z parametrów dłoni	9
4.3	Dopasowanie modeli regresji: liniowej, GLM i SVR modelującego wagę człowieka na podstawie jednego z parametrów dłoni	12
4.4	Klasyfikacja płci na podstawie pomiarów dłoni	14
5	Wnioski	15

1 Wstęp

1.1 Wybór data setu oraz problemu do analizy

Poszukując zbioru danych do analizy, zainteresował mnie ten dotyczący pomiarów dłoni oraz parametrów fizycznych osób. Celem analizy jest zbadanie czy wielkość dłoni ma jakiś związek ze wzrostem i wagą

1.2 Źródło data setu

Data set analizowany w dalszej części raportu został pobrany ze strony:

<https://www.kaggle.com/code/seshadrikolluri/anthropometric-data-analysis-and-visualization/data?select=ANSUR+II+FEMALE+Public.csv>

2 Preprocessing danych

2.1 Wybór danych poddanych analizie

Wybrany data set posiadał na wstępie 108 kolumn danych. Spośród nich wyłoniłem 6 które moim zdaniem mogły wpłynąć w pewnym stopniu na rozpatrywane wyniki.

- hand breadth - kolumna zawierające dane na temat szerokości dłoni w milimetrach.
- hand circumference - kolumna zawierające dane na temat obwodu dłoni w milimetrach.
- hand length - kolumna zawierające dane na temat długości dłoni w milimetrach.
- Gender - kolumna zawierająca dane na temat płci
- Weight - kolumna zawierająca dane na temat wagi w kilogramach
- Height - kolumna zawierająca dane na temat wzrostu w centymetrach

2.2 Przetworzenie danych

Zmieniłem wartości wagi z funtów na kilogramy oraz wartości wysokości z cali na centymetry. Musiałem również połączyć dwie bazy danych w jedną ponieważ były one rozdzielne na mężczyzn i kobiety. Postanowiłem również zmienić wartości płci z Female na 0 oraz z Male na 1

Nieznacznie zmieniłem również nazwy kolumn na takie które wydają mi się bardziej przejrzyste.

2.3 Rozkłady danych

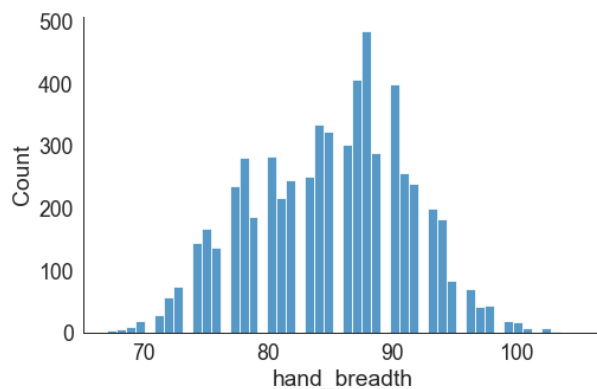


Figure 1: Rozkład danych - hand breadth

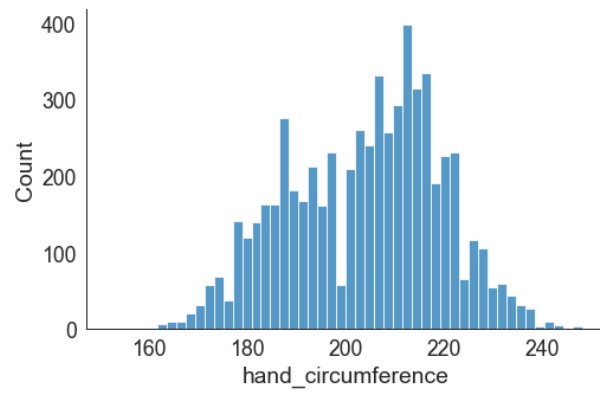


Figure 2: Rozkład danych - hand circumference

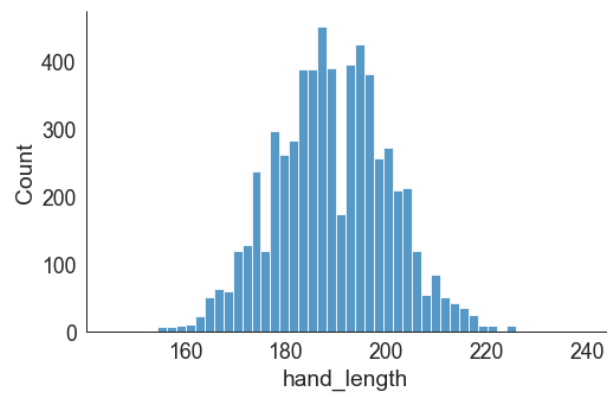


Figure 3: Rozkład danych - hand length

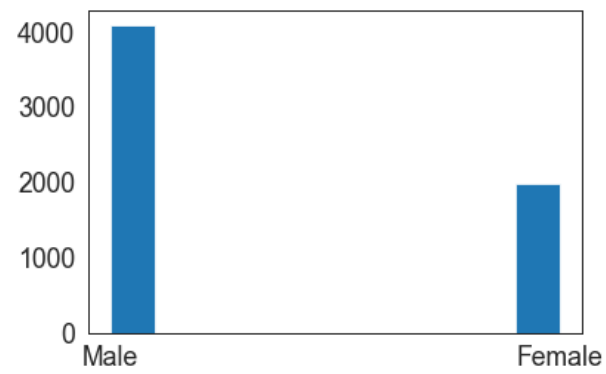


Figure 4: Rozkład danych - Gender

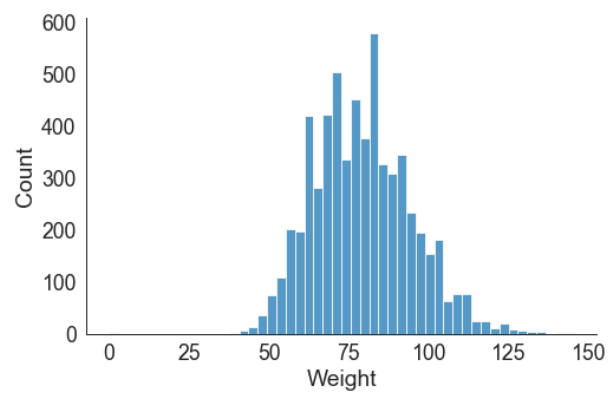


Figure 5: Rozkład danych - Weight

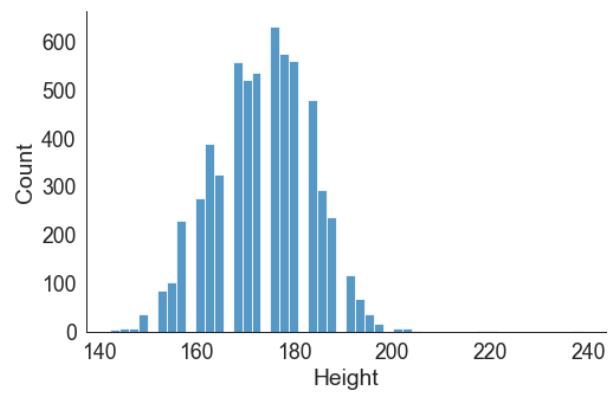


Figure 6: Rozkład danych - Height

	hand_breadth	hand_circumference	hand_length	Gender	Weight	Height
0	90	214	193	1	81.65	180.34
1	86	203	195	1	72.57	172.72
2	93	220	203	1	92.99	172.72
3	91	217	194	1	79.38	167.64
4	94	222	218	1	96.62	195.58

Figure 7: Rozkład danych - przykładowe wartości danych

	data
hand_breadth	84.953362
hand_circumference	203.861404
hand_length	189.286585
Gender	0.672709
Weight	79.264771
Height	173.373418

Figure 8: Rozkład danych - średnie wartości danych

3 Rozwiązanie problemu

3.1 Opis problemu

Celem analizy jest próba:

- 3.1.1 Dopasowanie rozkładu do wartości długości dłoni
- 3.1.2 Dopasowanie modeli regresji: liniowej, GLM i SVR modelującego wzrost człowieka na podstawie jednego z parametrów dłoni
- 3.1.3 Dopasowanie modeli regresji: liniowej, GLM i SVR modelującego wagę człowieka na podstawie jednego z parametrów dłoni
- 3.1.4 Klasyfikacja płci na podstawie pomiarów dłoni

3.2 Podział danych testowych

Wszystkie dane użyte do wyznaczenia regresji jak i klasyfikatora w kolejnych etapach, dzieli na dane trenujące jak i testowe w proporcji 75:25 z ustawionym ziarnem losowości na 143 aby dane były losowane zawsze w ten sam sposób. Analizować będę modele stworzone na podstawie estymatorów dostępnych w scikit-learn.

4 Wykonywane eksperymenty z modelami / rozkładami

4.1 Dopasowanie rozkładu do wartości długości dłoni

Na początku za pomocą modułu fitter opartym na module SciPy określiłem wartości błędu kwadratowego dla kilku popularnych rozkładów tj. Normalnego, Gamma, Chi2 oraz Cauchy'ego

Jakość każdego dopasowania określamy błędem średnio-kwadratowym

Rozkład	Błąd
Normalny	0.003132
Gamma	0.003271
Chi2	0.003427
Cauchy'ego	0.004453

Najlepszym rozkładem okazał się rozkład normalny

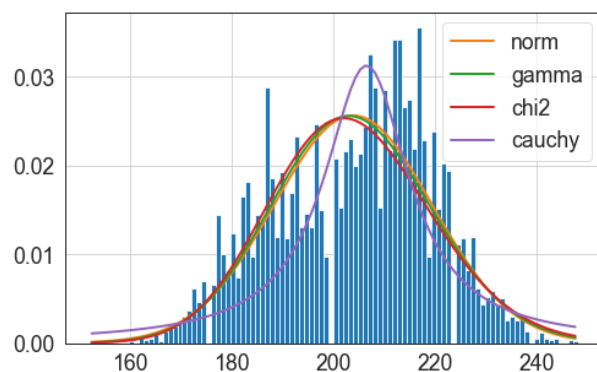


Figure 9: Próba dopasowania kilku popularnych rozkładów do danych długości dłoni

4.2 Dopasowanie modeli regresji: liniowej, GLM i SVR modelującego wzrost człowieka na podstawie jednego z parametrów dłoni

Po analizie współczynnika korelacji możemy wywnioskować że atrybutem najlepiej odwzorowującym wartość wzrostu jest pomiar długości ręki. Dlatego też wybieramy ją na daną naszego modelu.

	hand_breadth	hand_circumference	hand_length	Gender	Weight	Height
Weight	0.708388	0.738392	0.599202	0.564082	1.000000	0.659984
Gender	0.747338	0.776501	0.497361	1.000000	0.564082	0.660288
hand_breadth	1.000000	0.963017	0.704189	0.747338	0.708388	0.714933
hand_circumference	0.963017	1.000000	0.703555	0.776501	0.738392	0.722392
hand_length	0.704189	0.703555	1.000000	0.497361	0.599202	0.737603
Height	0.714933	0.722392	0.737603	0.660288	0.659984	1.000000

Figure 10: Tabela korelacji

Za pomocą danych testujących zostały stworzone 3 modele: regresji liniowej, GLM jak i SVR. Parametry modeli, błąd średnio-kwadratowy (Mean Square Error (MSE)) jak i uśredniony błąd bezwzględny (Mean Absolute Error (MAE)) tych modeli został ukazany w Tabeli

Model	Parametry	MSE	MAE
Liniowy	[0.63239], 53.74442	44.8	5.09
GLM	[0. 2.0747 -0.0038], -82.48953	44.3	5.07
SVR	—	44.6	5.08

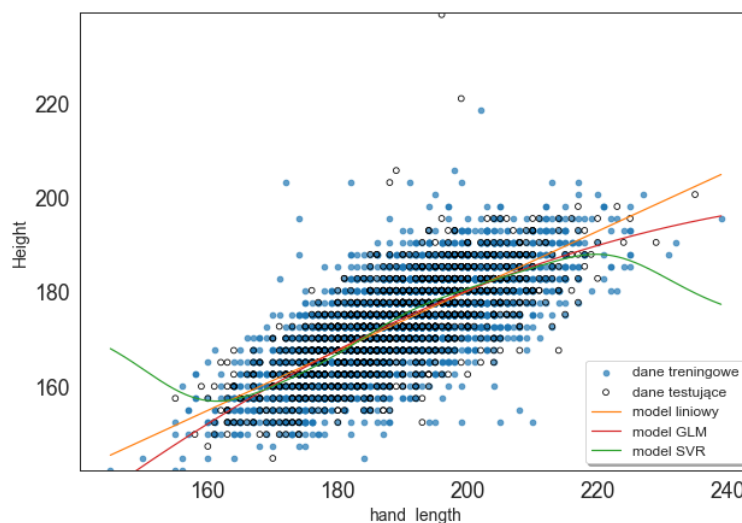


Figure 11: Modele regresji próbujące opisać wzrost człowieka względem długości dłoni

Dodatkowo warto sprawdzić czy wartość błędu zmieni się gdy uwzględnimy dane mężczyzn i kobiet osobno.

tabela kobiet

Model	Parametry	MSE	MAE
Liniowy	$[0.42635], 86.89839$	31.7	4.37
GLM	$[0.14223 -0.0027], -3.7936, -82.48953$	31.6	4.36
SVR	————	32.5	4.4

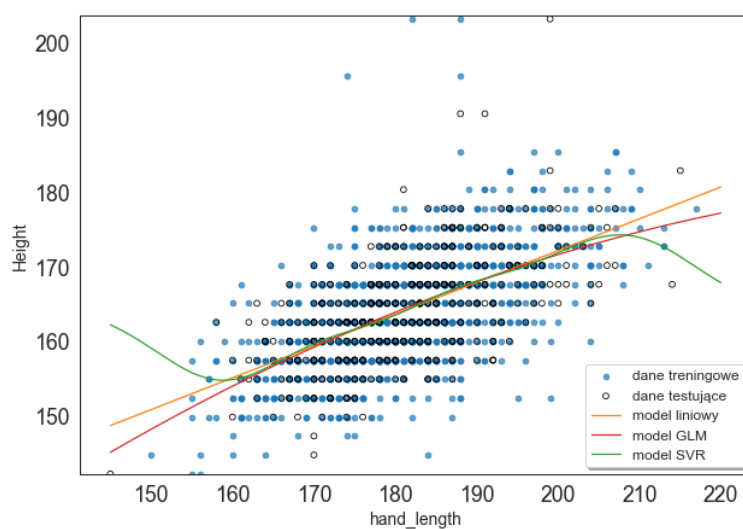


Figure 12: Modele regresji próbujące opisać wzrost kobiety względem długości dłoni

tabela mężczyzn

Model	Parametry	MSE	MAE
Liniowy	$[0.48744], 83.68982$	35.9	4.51
GLM	$[0. 1.2108 -0.0019], 13.53194$	35.8	4.5
SVR	————	36.0	4.52

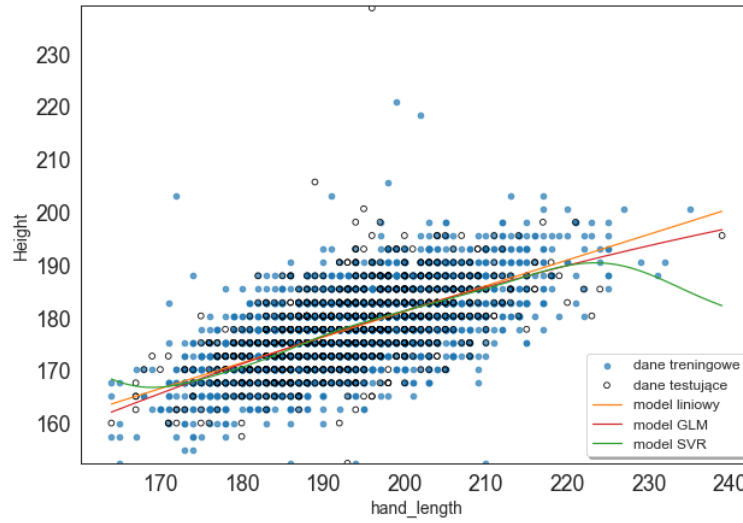


Figure 13: Modele regresji próbujące opisać wzrost mężczyzny względem długości dłoni

Z doświadczenia wynika że można określić wzrost człowieka na podstawie długości ręki, już dla ogólnej bazy wartości są dosyć dokładne, jednak przy odróżnieniu płci wartości są jeszcze dokładniejsze.

4.3 Dopasowanie modeli regresji: liniowej, GLM i SVR modelującego wagę człowieka na podstawie jednego z parametrów dłoni

Po analizie współczynnika korelacji możemy wywnioskować że atrybutem najlepiej odwzorowującym wartość wzrostu jest pomiar obwodu ręki. Dlatego też wybieramy ją na daną naszego modelu.

	hand_breadth	hand_circumference	hand_length	Gender	Weight	Height
Gender	0.747338	0.776501	0.497361	1.000000	0.564082	0.660288
hand_length	0.704189	0.703555	1.000000	0.497361	0.599202	0.737603
Height	0.714933	0.722392	0.737603	0.660288	0.659984	1.000000
hand_breadth	1.000000	0.963017	0.704189	0.747338	0.708388	0.714933
hand_circumference	0.963017	1.000000	0.703555	0.776501	0.738392	0.722392
Weight	0.708388	0.738392	0.599202	0.564082	1.000000	0.659984

Figure 14: Tabela korelacji

Za pomocą danych testujących zostały stworzone 3 modele: regresji liniowej, GLM jak i SVR. Parametry modeli, błąd średnio-kwadratowy (Mean Square Error (MSE)) jak i uśredniony błąd bezwzględny (Mean Absolute Error (MAE)) tych modeli został ukazany w Tabeli

Model	Parametry	MSE	MAE
Liniowy	[0.73261], -70.04801	1.04e+02	7.99
GLM	[0. -0.2449 0.0024], 28.19685, -82.48953	1.04e+02	7.98
SVR	—	1.05e+02	7.98

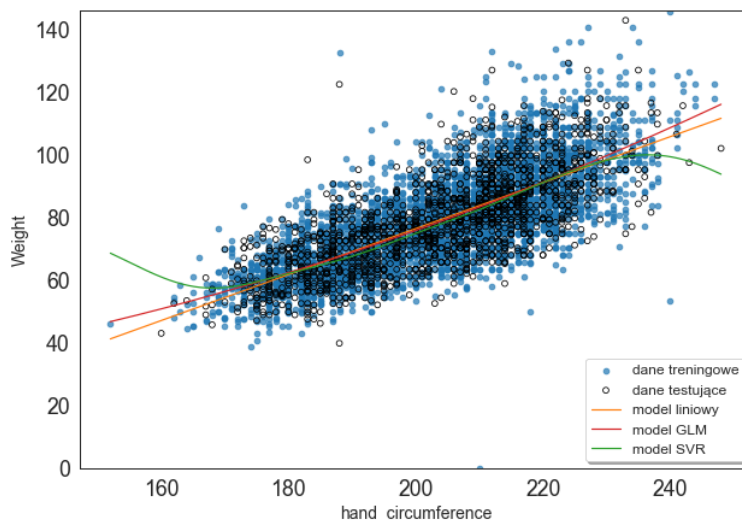


Figure 15: Modele regresji próbujące opisać wagę człowieka względem obwodu dłoni

Dodatkowo warto sprawdzić czy wartość błędu zmieni się gdy uwzględnimy dane mężczyzn i kobiet osobno.

tabela kobiet

Model	Parametry	MSE	MAE
Liniowy	[0.66795], -57.68512	73.3	6.82
GLM	[0. -0.1457 0.0022], 18.18156	73.6	6.82
SVR	————	73.6	6.77

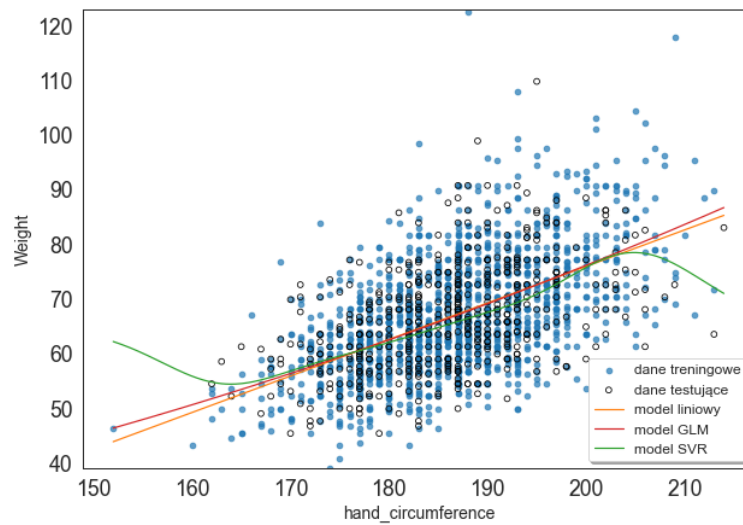


Figure 16: Modele regresji próbujące opisać wagę kobiety względem obwodu dłoni

tabela mężczyzn

Model	Parametry	MSE	MAE
Liniowy	[0.79419], -83.30662	1.08e+02	8.23
GLM	[0. 0.4009 0.0009], -41.56403	1.08e+02	8.23
SVR	————	1.08e+02	8.19

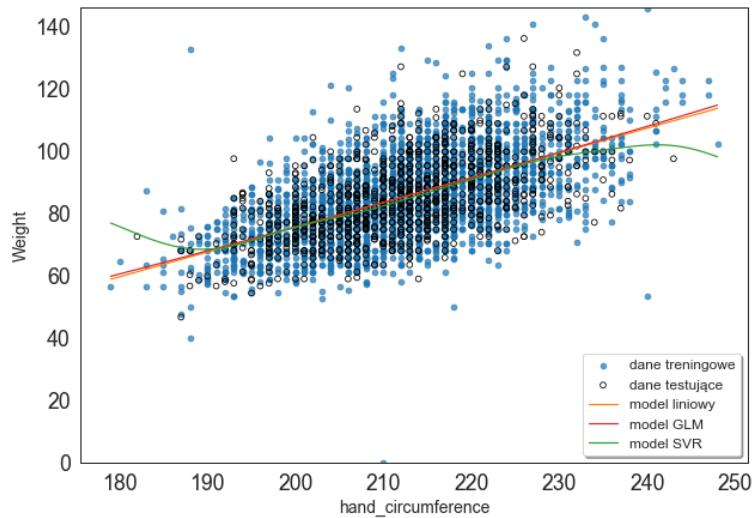


Figure 17: Modele regresji próbujące opisać wagę mężczyzn względem obwodu dłoni

Z doświadczenia wynika ciekawy wynik, co prawda błędy zarówno u mężczyzn jak i kobiet są dosyć duże, jednak u kobiet błędy są znacznie mniejsze niż u mężczyzn.

4.4 Klasyfikacja płci na podstawie pomiarów dłoni

Jak mogliśmy zauważyć dzięki znajomości płci jesteśmy w stanie lepiej określić pewne wartości fizyczne osób. Więc postaramy się ją wyznaczyć. Za pomocą danych opisujących wartości parametrów odcisku palca jak i płci uczestników badania tworzymy klasyfikator. Naszym klasyfikatorem będzie KNN (K-Nearest Neighbours) z biblioteki scikit-learn, określający klasę/wartość(u nas płeć) na podstawie odległości od K -sąsiadów. Nasze dane na tych samych zasadach dzielimy na testowe i trenujące co w poprzednich eksperymentach. Natomiast K ustawiamy równe 7. Dokładność rozwiązania tego klasyfikatora wynosi 94% czyli z całkiem dużą dokładnością możemy określić płeć osoby na podstawie jej odcisku palca.

5 Wnioski

Wnioskując z wykonanych eksperymentów można dowiedzieć się, że można przewidzieć wzrost jak i wagę człowieka za pomocą dłoni poprzez modele regresji. Całkiem ważnym aspektem jest zaznaczenie tego że dokładniej można określić wzrost, gdy wiemy jaka jest płeć poszukiwanej osoby. Natomiast gdy nie mamy takiej informacji, możemy ją łatwo uzyskać. Dostaniemy ją za pomocą klasyfikatora KNN (Eksperyment III-D) którego danymi są pomiary dłoni. Jesteśmy w stanie z 94 % poprawnością określić czy dana osoba jest kobietą czy mężczyzną tylko za pomocą kilku parametrów dotyczących ręki. Natomiast określenie wagi osoby nie jest aż tak właściwe za pomocą modeli regresji ponieważ jest obciążone dość dużym błędem. Zważając nawet na te błędy modeli. Wartości pomiarów ręki, mogą służyć do całkiem trafnego określenia wzrostu oraz wagi. Jednak musimy się liczyć że nasz model jest obciążony pewnym błędem w przypadku wzrostu mniejszym a wagi większym.