

Wykorzystanie algorytmów ML do określenia predyspozycji do otyłości na podstawie danych z analiz GWAS

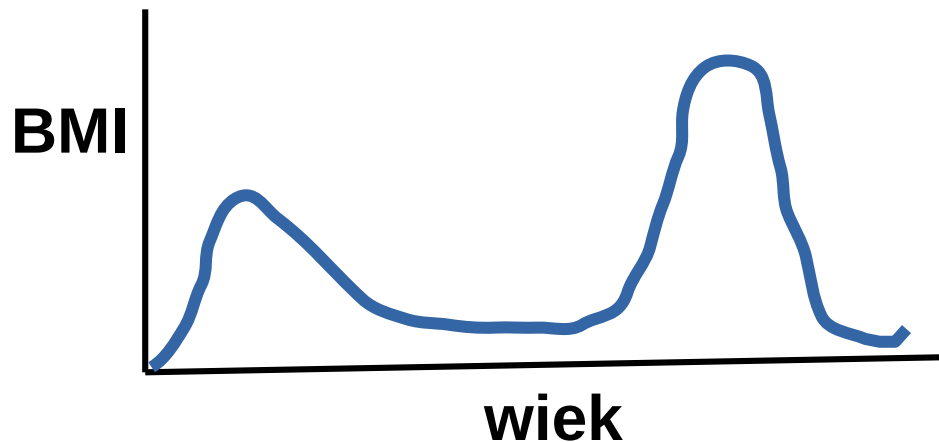
Drużyny:

BinaryMudTurtles
+
ParaData

Jacek Stańdo
Artur Jurgas
Monika Krzyżanowska
Artur Bąk
Bartłomiej Kantor
Grzegorz Migdałek

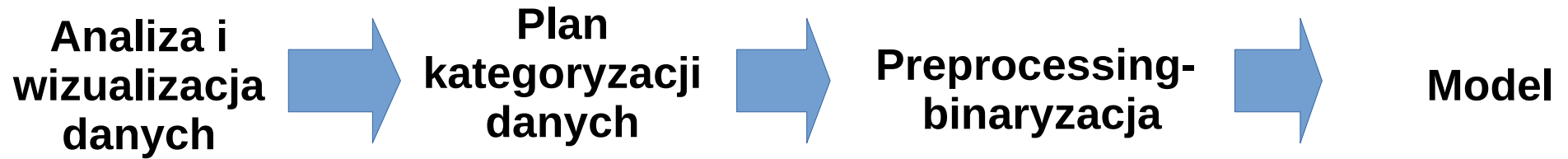
Cel 1: spersonalizowany profil predyspozycji do otyłości zależnie od płci i SNP

Dane:
- Profil SNP
- płeć

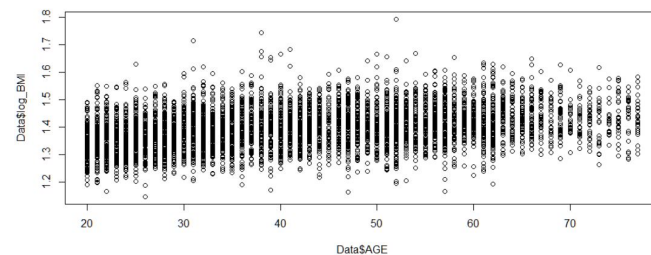
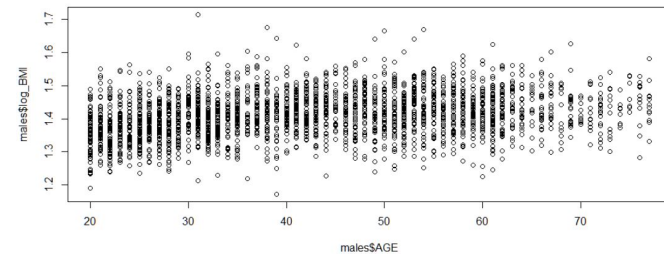
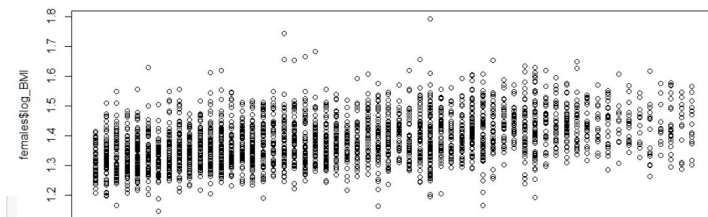
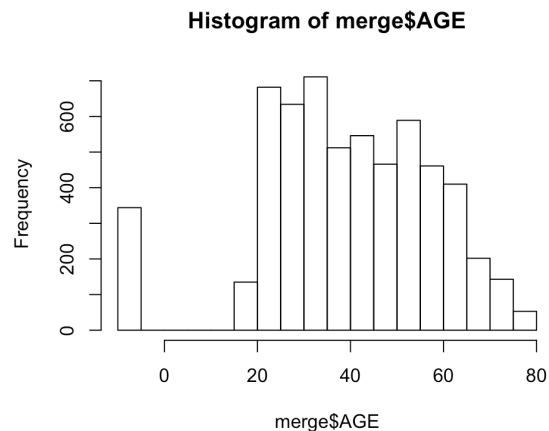
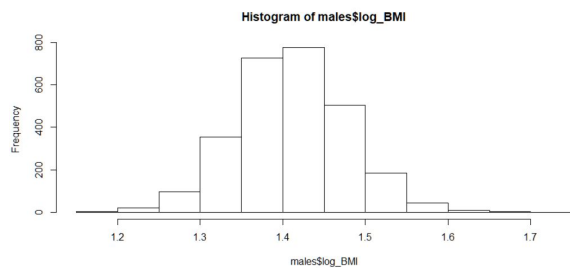
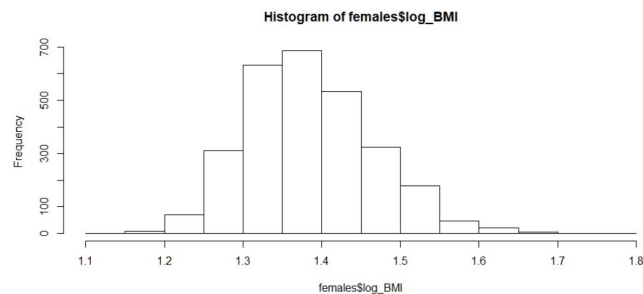
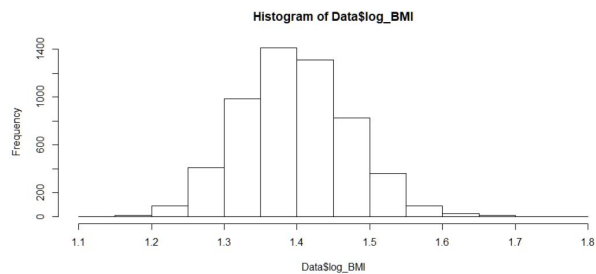


**Cel 2: Wyznaczenie najistotniejszych SNP,
wpływających na predyspozycje do otyłości
za pomocą algorytmów uczenia maszynowego**

Workflow:



1. Analiza i wizualizacja danych



2. Plan kategoryzacji danych

Age_group

60 - 69

20 - 29

60 - 69

20 - 29

30 - 39

40 - 49

50 - 59

20 - 29

40 - 49

20 - 29

Zbyt mało
danych

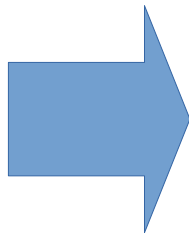
Uproszczenie:

20 – 40

40 +

Binaryzacja danych

SNP325.2	SNP326.1	SNP326.2	SNP327.1	SNP327.2	SNP328.1	SNP328.2	SNP329.1
G	T	T	G	G	G	G	G
G	T	T	G	G	G	G	G
G	T	T	G	G	G	G	G
G	T	T	G	G	G	G	G
G	T	T	G	G	G	G	G
G	T	T	A	G	A	G	G
G	T	T	G	G	G	G	G
G	T	T	G	G	G	G	G
G	T	T	G	G	0	0	0
G	T	T	G	G	G	G	G



SNP325.1	SNP325.2	SNP326.1	SNP326.2	SNP327.1	SNP327.2
0	0	0	0	0	0
1	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	1	0
0	0	0	0	0	0
0	0	0	0	0	0
1	0	0	0	0	0

Model

XgBoost – zastosowanie optymalizacji

doi: [10.1002/gepi.21000](https://doi.org/10.1002/gepi.21000)

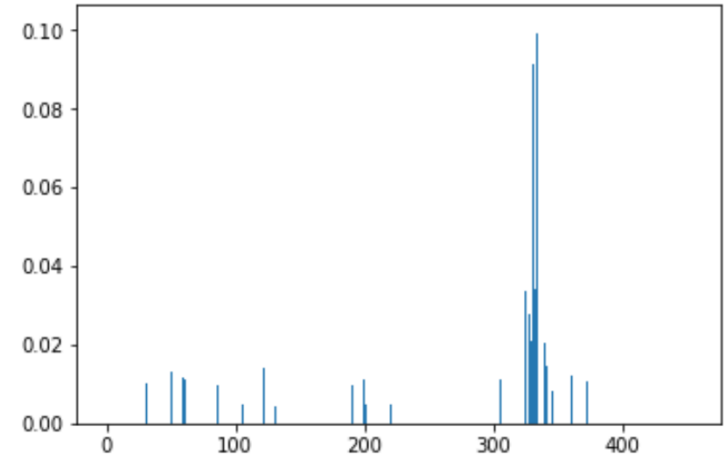
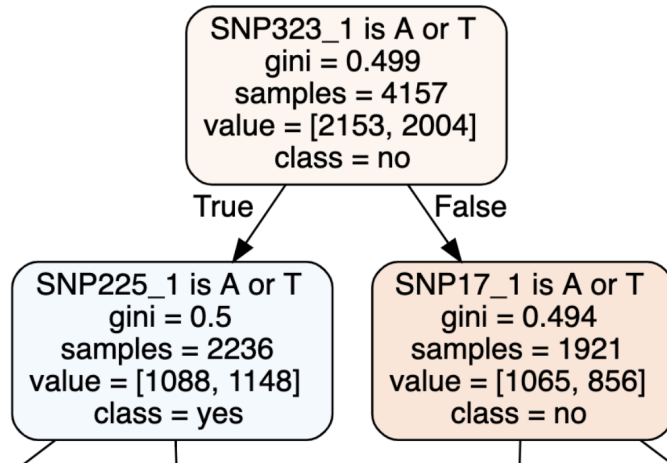
Bootstrap Aggregating of Alternating Decision Trees to Detect Sets of SNPs that Associate with Disease

[Richard T. Guy](#),^{1,4} [Peter Santago](#),^{2,3} and [Carl D. Langefeld](#)¹

► [Author information](#) ► [Copyright and License information](#) [Disclaimer](#)

Drzewo decyzyjne

Najistotniejsze SNP



XBoost

Drzewo decyzyjne

Perspektywy zastosowania prototypu

- Po optymalizacji
- Jakie SNP są istotne przy diagnostyce predyspozycji do chorób – diagnostyka na podstawie wybranych SNP

Wątpliwości, co ulepszyć

- Użycie Age + Sex w algorytmie
- BMI – czy dobry wyznacznik?
 - Zawartość tłuszczu w organizmie
- Otyłość – nie tylko genetyka ...
 - Warunki środowiska

Zaplecze techniczne

Python, R, ggplot, scikit-learn, Xgboost

Dziękujemy za uwagę

