
Wyjaśnialny ML w naukach medycznych

Katarzyna Kobylińska

Bazy danych:

- MIMIC
 - <https://mimic.physionet.org/about/mimic/>
- Operacyjny rak płuc w polskiej populacji
 - Instytut Gruźlicy i Chorób Płuc w Warszawie

MIMIC III (Medical Information Mart for Intensive Care III):

- rejestr pacjentów z Oddziału Intensywnej Terapii (Beth Israel Deaconess Medical Center)
- dane pochodzą z 2001-2012
- dostępna (kurs online)
- ~ 40 000
- 26 tabel zawierających: informacje demograficzne, pomiary wykonywane np. co godzinę, testy laboratoryjne, metody leczenia, śmiertelność



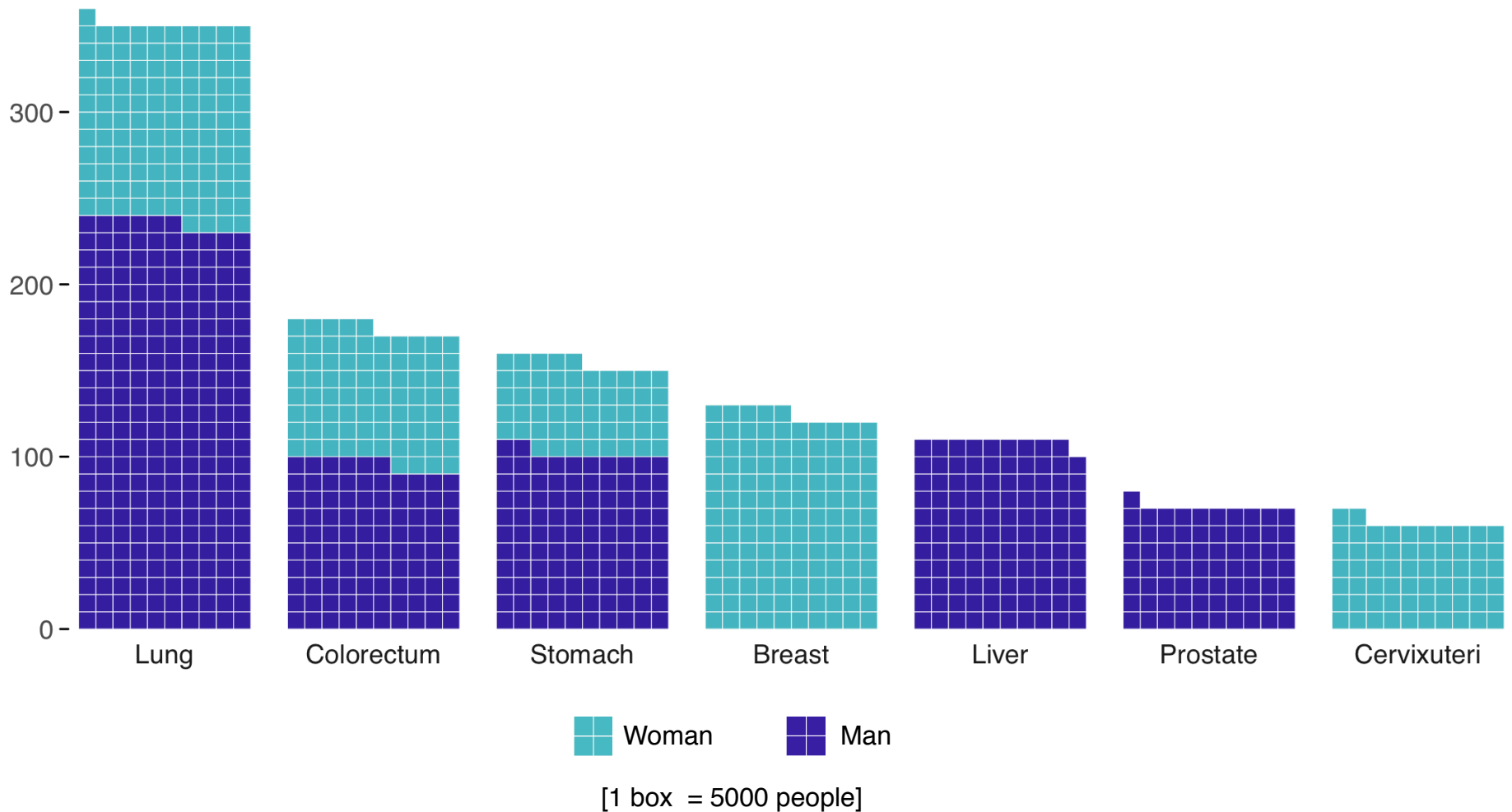
MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35). Available from: <http://www.nature.com/articles/sdata201635>

Operacyjny Rak Płuc w Polsce:

- centralny rejestr skupiający wszystkie ośrodki w Polsce
- każdy pacjent zoperowany w Polsce
- 30k+ przypadków
- ~200 zmiennych
- dane pochodzą z okresu: 2002 - 2016
- Czynniki ryzyka (e.g. narażenie na azbest, palenie papierosów), wyniki histopatologiczne (wymiar guza), wyniki badań, symptomy (astma, kaszel), wywiad rodzinny



Number of deaths in 2018 by cancer type



AIME 2019, 17-th Conference on Artificial Intelligence in Medicine, Poznań

- **Warsztat:** Transparent, Explainable and Affective AI in Medical Systems (TEAAM)
 - **Publikacja:** [Lecture Notes in Computer Science](#)
[Link do artykułu](#)
-

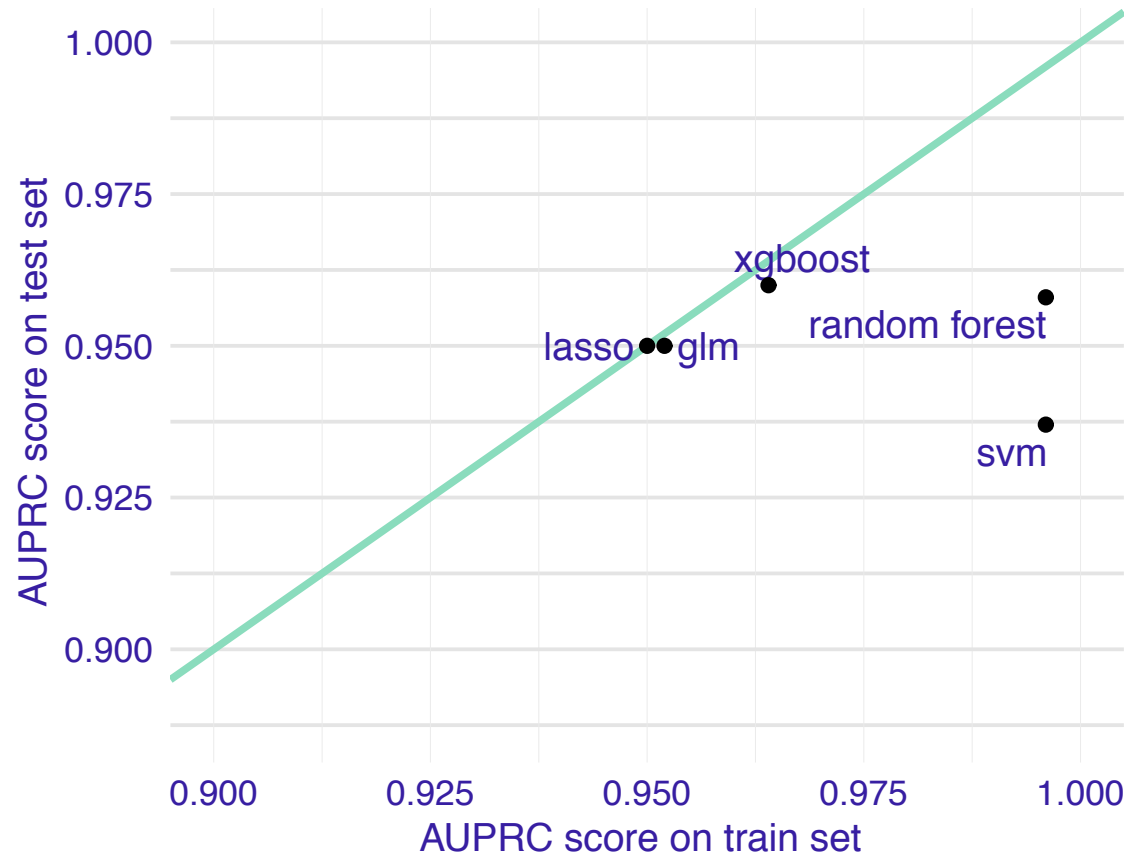
Modelowanie:

Predykcja przeżycia pierwszych
3 miesięcy po operacji na podstawie
informacji dostępnych przed operacją

- silnie niezbalansowane dane
- AUC ~ 68%
- AUPRC ~ 95% (area under the precision recall curve)
 - Precision = $TP / (TP + FP)$
 - Recall = $TP / (TP + FN)$ (sensitivity)

<http://pages.cs.wisc.edu/~jdavis/davisgoadrichcamera2.pdf>

Modelowanie:



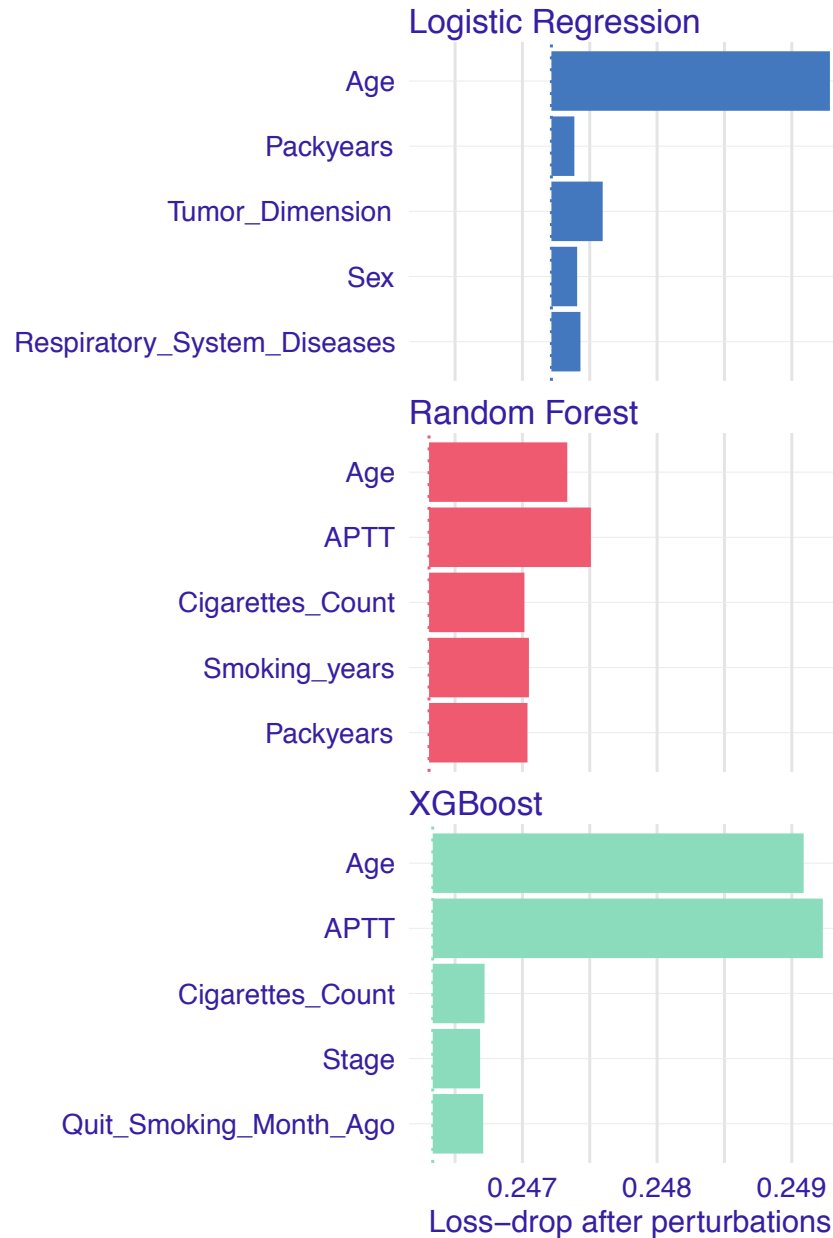
Las losowy vs Regresja Logistyczna vs XGboost

Wyjaśnienie modeli

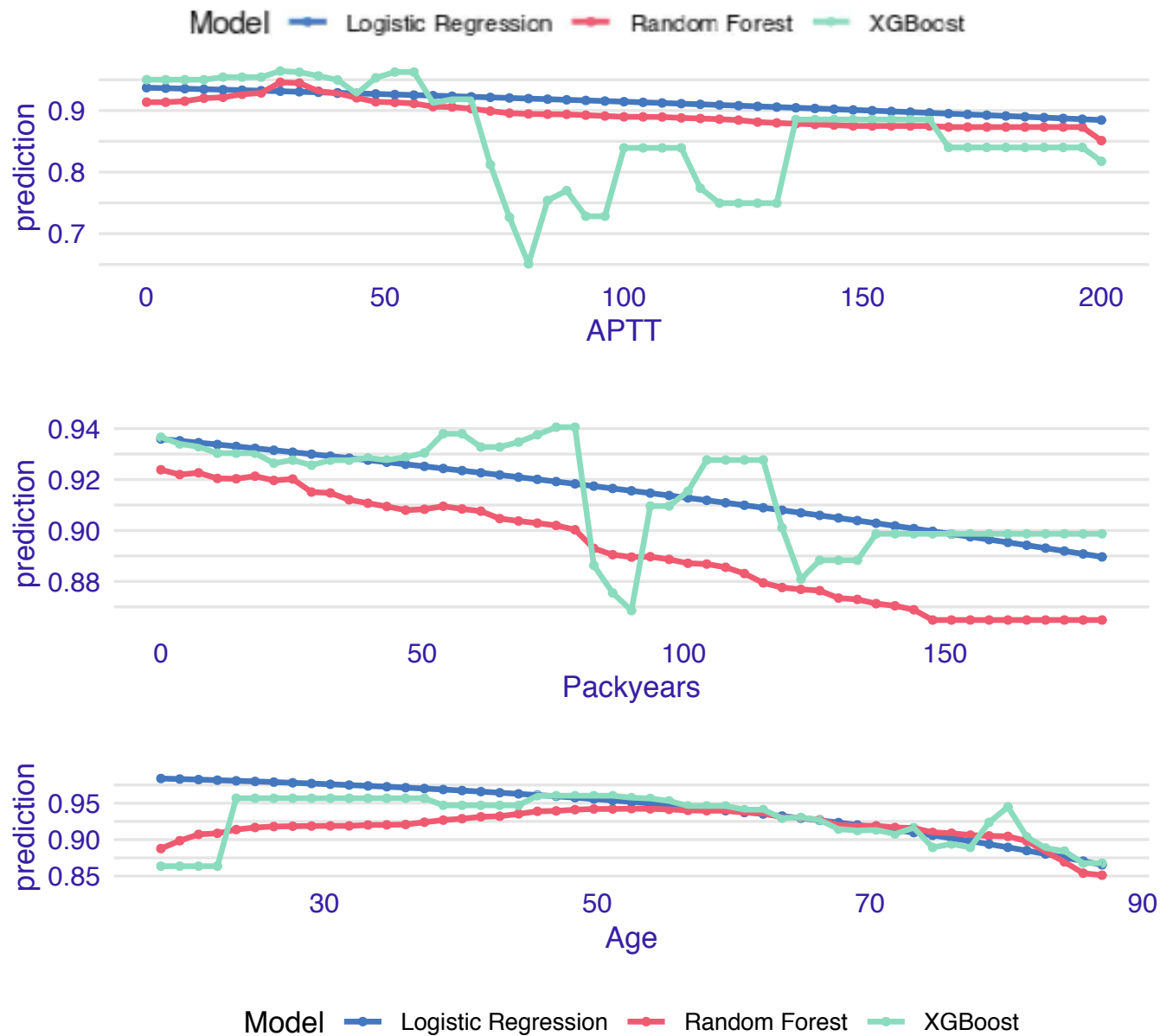
1. Model performance
2. Variable importance
3. Partial dependence plots
4. Sequential variable attribution
5. What-if analysis



Variable Importance



Partial Dependence Plots



Następne kroki:

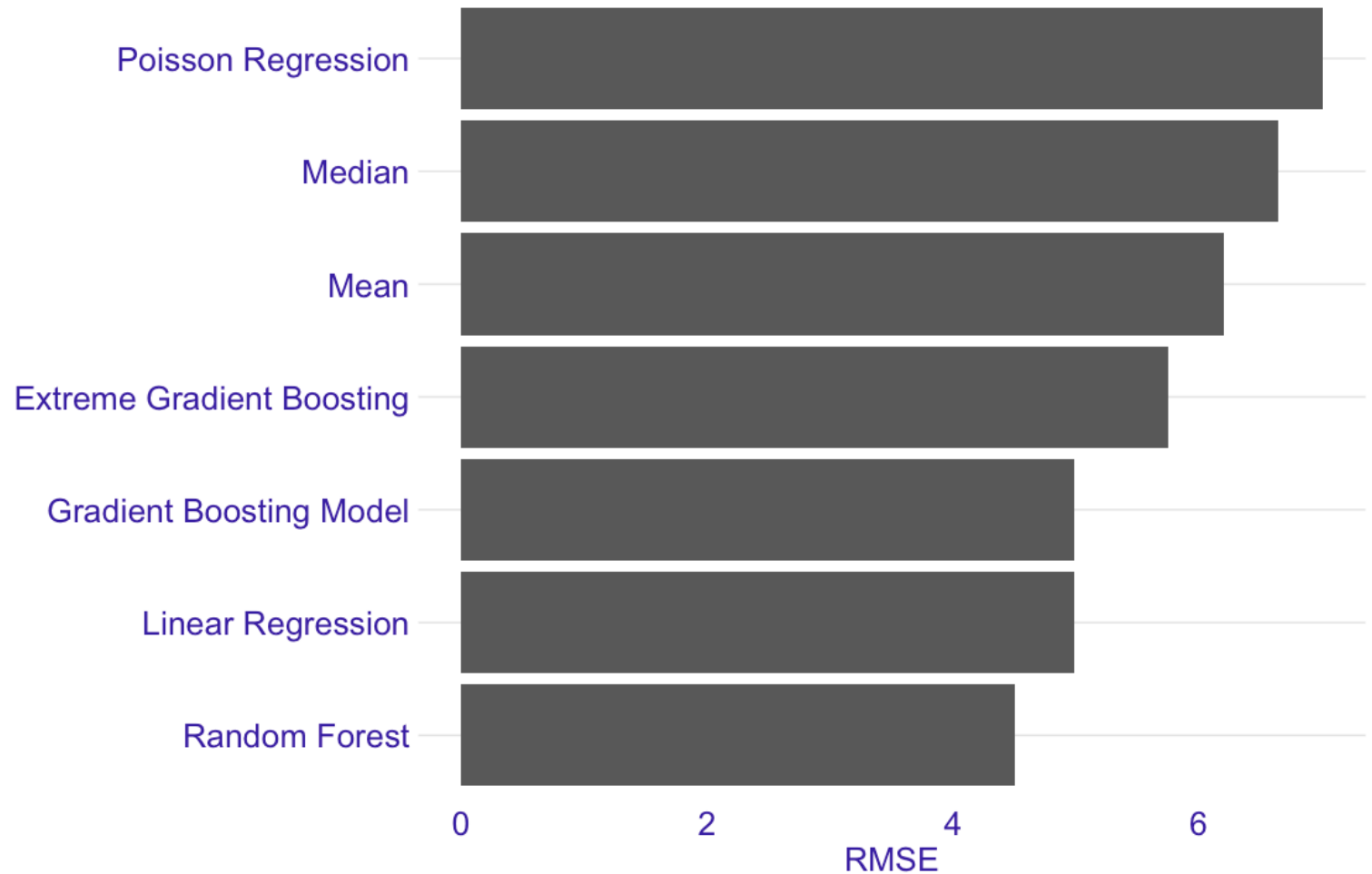
- porównanie istniejących modeli przewidyujących ryzyko zachorowania na raka płuc**

MIMIC modelowanie:

Prognoza czasu pobytu w szpitalu noworodków

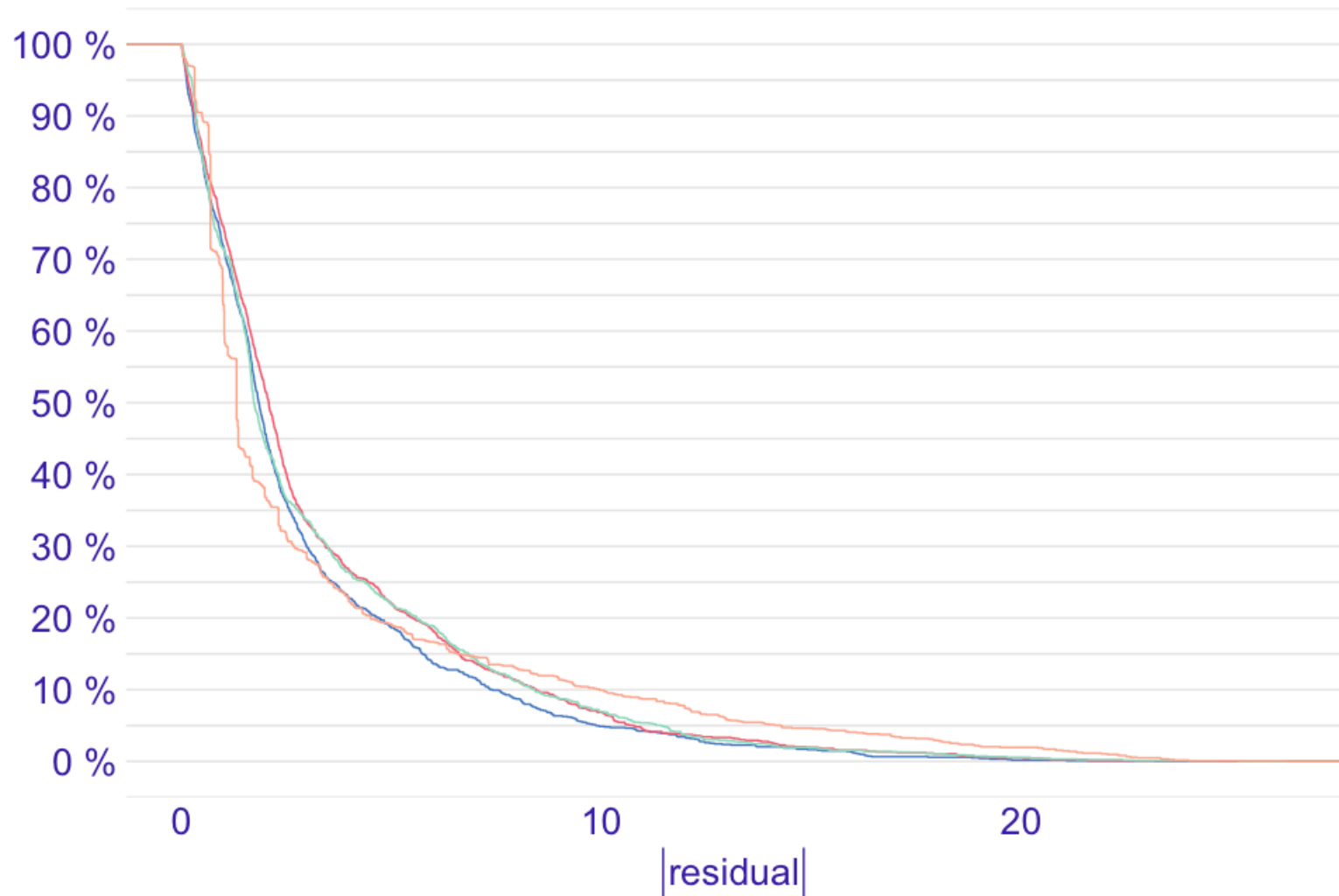
- Wybór zmiennych do modelu na podstawie [publikacji](#)
- Porównanie szeregu modeli

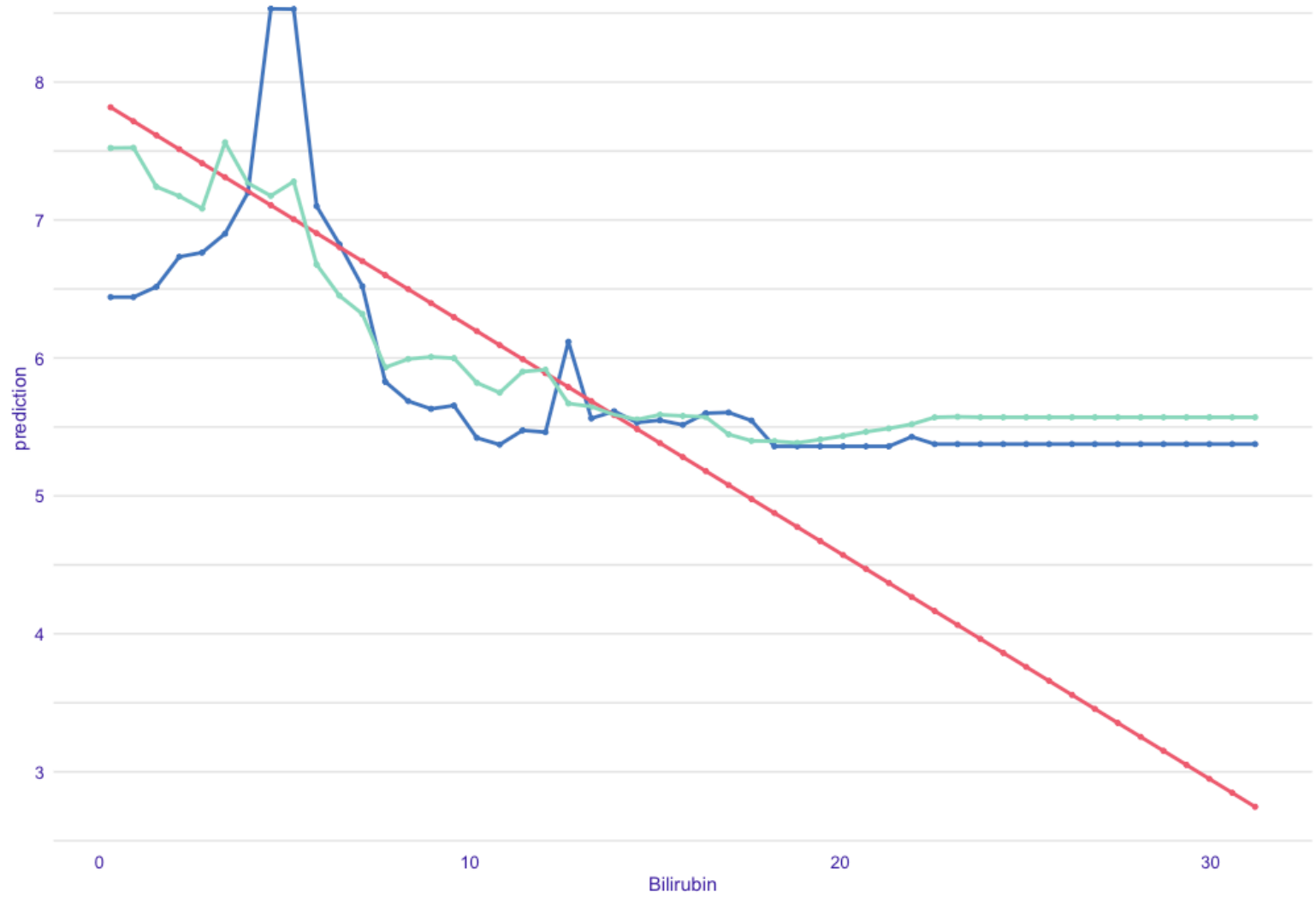
MIMIC modelowanie:



Distribution of $|\text{residual}|$

Model — Random Forest — Linear Regression — Gradient Boosting — Extreme Gradient Boosting





Następne kroki:

- bilirubina?**
- artykuł**