



PORÓWNIANIE TRANSFORMACJI ZMIENNYCH W REGRESJI LINIOWEJ

Łukasz Brzozowski, Wojciech Kretowicz, Kacper Siemaszko

- Zbudować model regresji liniowej tak dobry, jak czarna skrzynka
- Porównać skuteczność różnych metod transformacji zmiennych w regresji liniowej
- Porównać skuteczność czarnych skrzynek po tych samych transformacjach
- Zinterpretować modele po transformacjach

Nasze cele

Nasze dane

- Zbiór danych *Concrete_Data* z bazy *OpenML*
- 9 zmiennych, w tym zmienna celu
- 1030 obserwacji
- Brak braków danych, brak zmiennych kategoriycznych

- Transformacja zmiennych metodą prób i błędów
- Brute Force
- Optymalizacja Bayesowska
- Transformacje genetyczne



Nasze
metody

Nasz pomysł

Podstawowa regresja:

$$a_0 + \sum_{i=1}^n a_i \cdot x_i = \hat{y}$$

Wskaźnik wpływu zmiennej:

$$\left| \frac{\partial \hat{y}}{\partial x_i} \right| = |a_i|$$

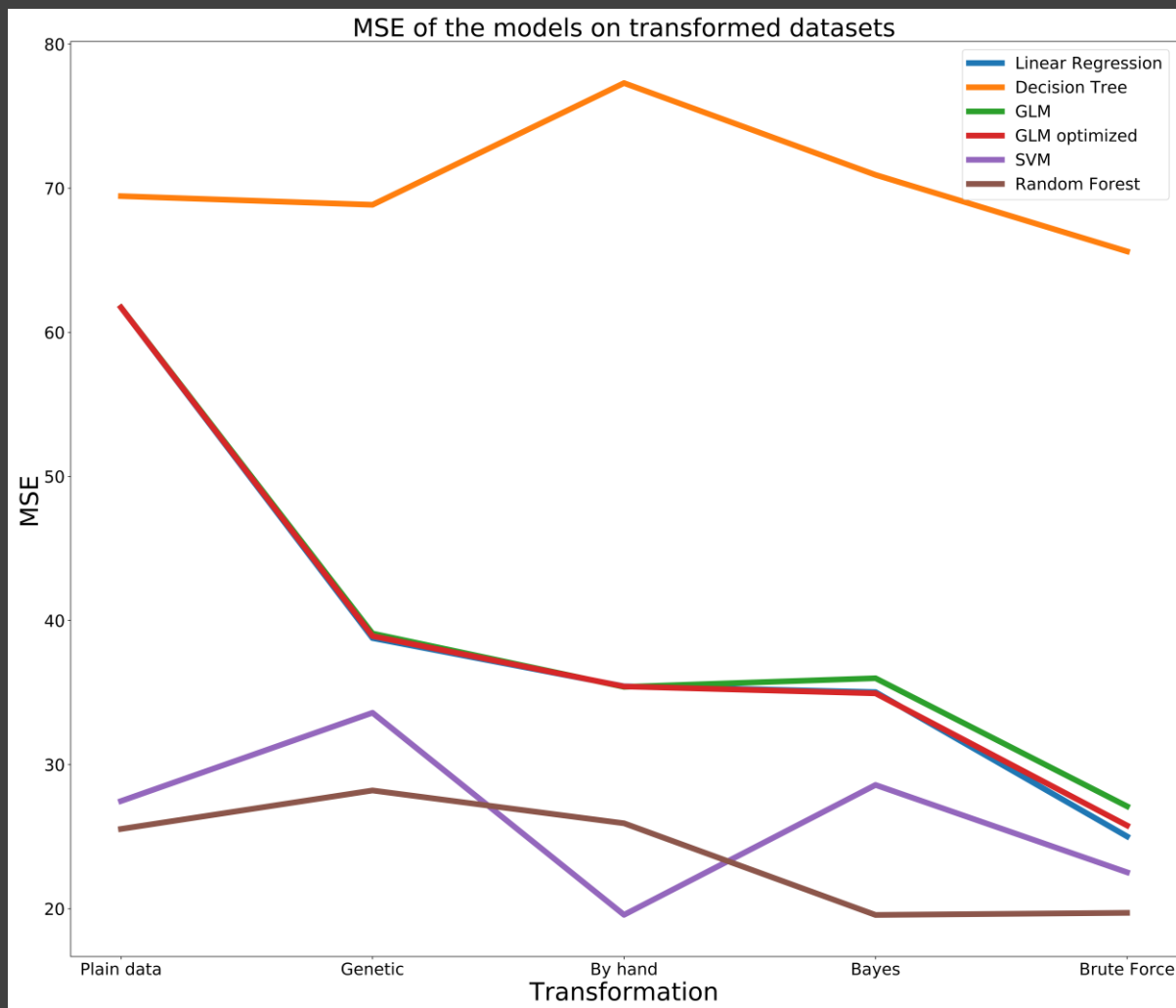
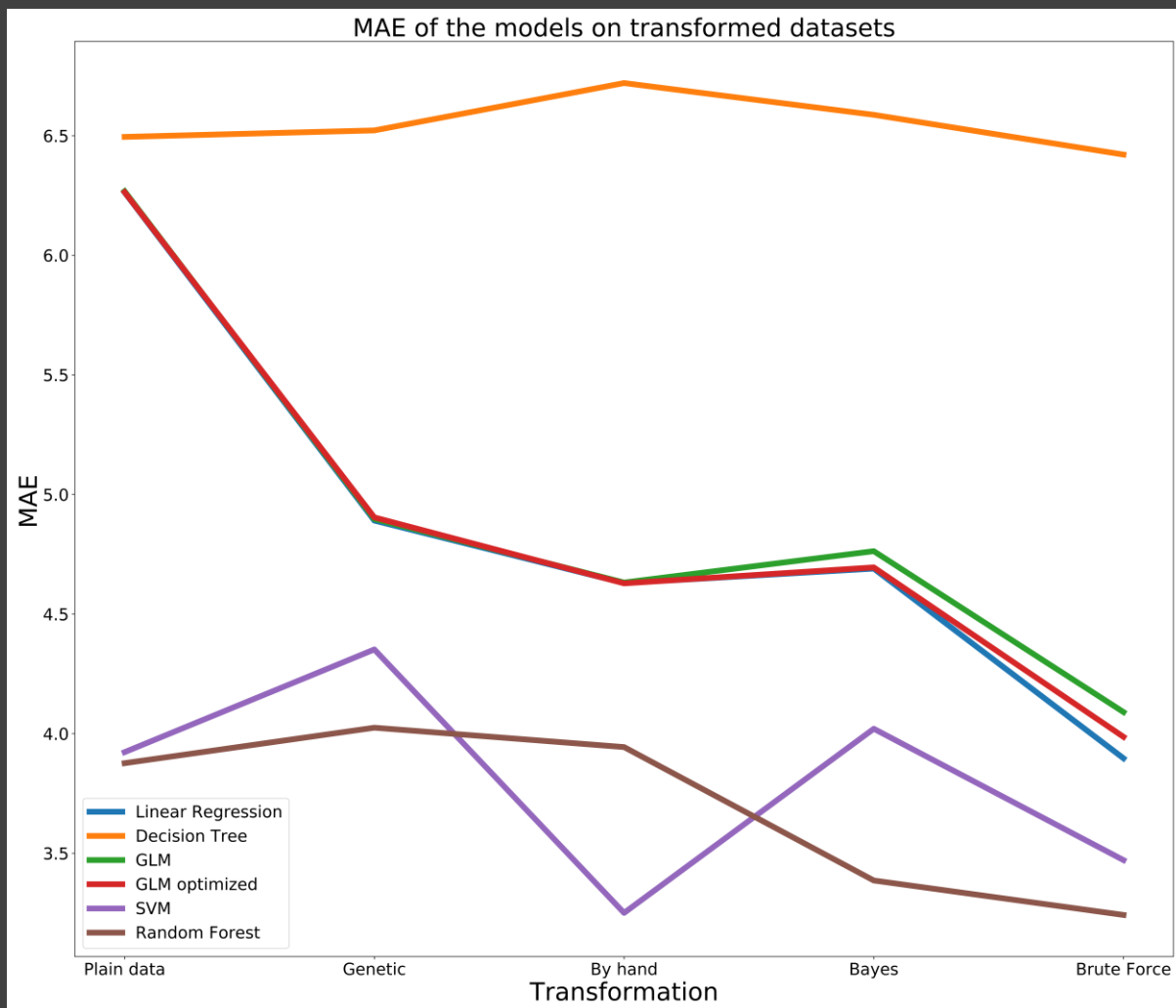
Regresja po transformacjach:

$$a_0 + \sum_{i=1}^N a_i \cdot f_i(x_1, \dots, x_n) = \hat{y}$$

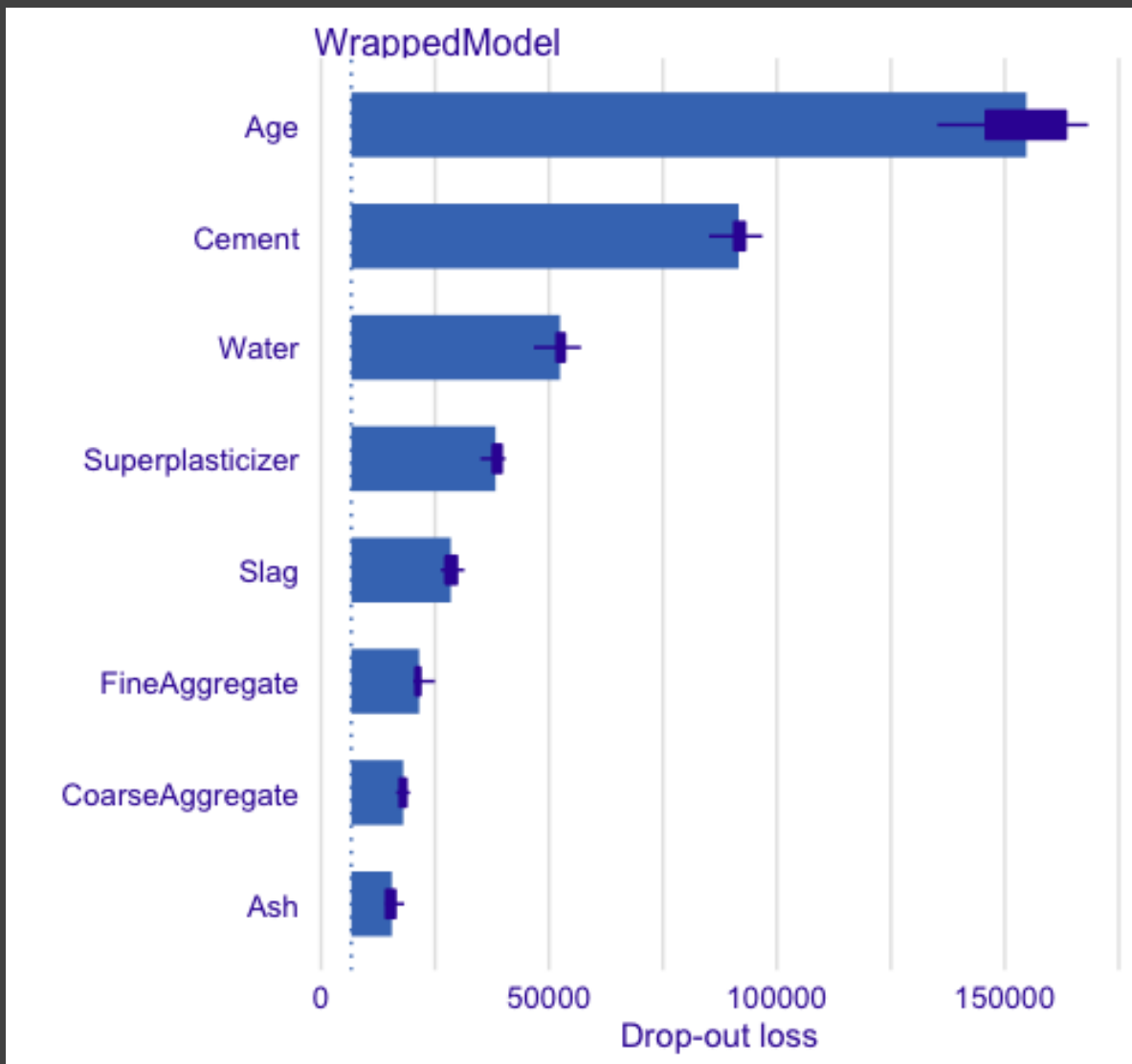
Wskaźnik wpływu zmiennej:

$$\left| \frac{\partial \hat{y}}{\partial x_i} \right|$$

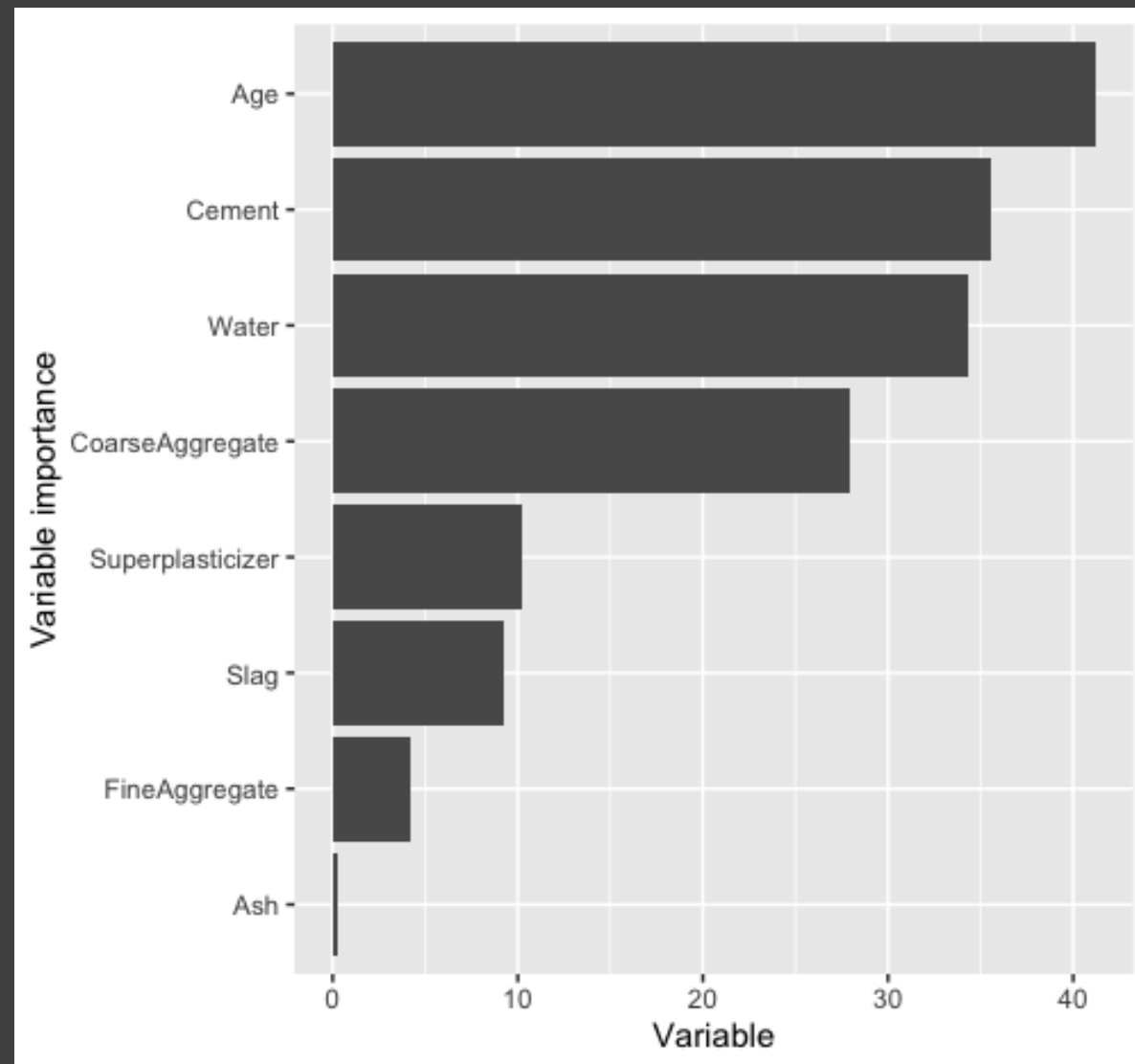
Nasze wyniki



Ranger FE



Nasza metryka



Dziękujemy za uwagę