

Важность признаков на деперсонализированном наборе данных




Входные данные

Строк 200967, столбцов 88

TARGET = 1, строк 751

TARGET = 0, строк 200216

Этапы

- 1) Удаляем строки с выбросами $\text{target}==0$: 2760 , $\text{target}==1$: 7
 - 2) Заполняем пропущенные значения медианой или не заполняем если два значения
 - 3) Убираем чрезмерную размерность у столбцов
 - 4) Кодлируем
- 

Очистка

- Столбцы с одним или пропущенным значением ['2.19', '2.2', '2.909', '2.91', '2.92', '2.926', '2.927', '2.93', '2.937', '2.94', '2.95']

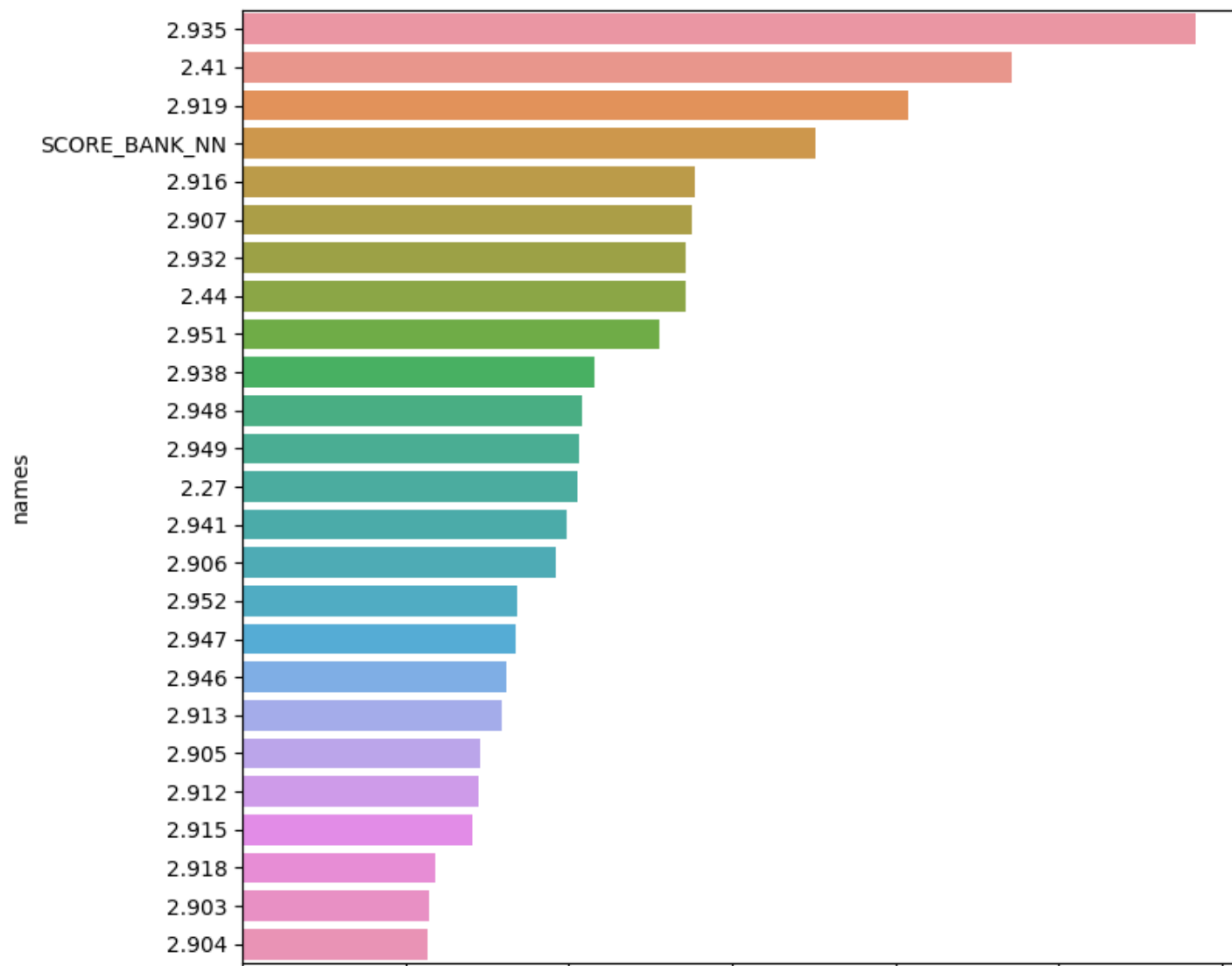
-

Удаляем столбцы, которые не показывают вклад в модель и являются шумом

- '1.18', '2.16', '2.42'

Важность вместе со столбцом SCORE_BANK_NN

Важность признаков сумма XGBoost и RandomForest



Важность вместе со столбцом score_bank

Важность признаков сумма XGBoost и RandomForest

