

Agro Data Science Cup

Максим Хандусь

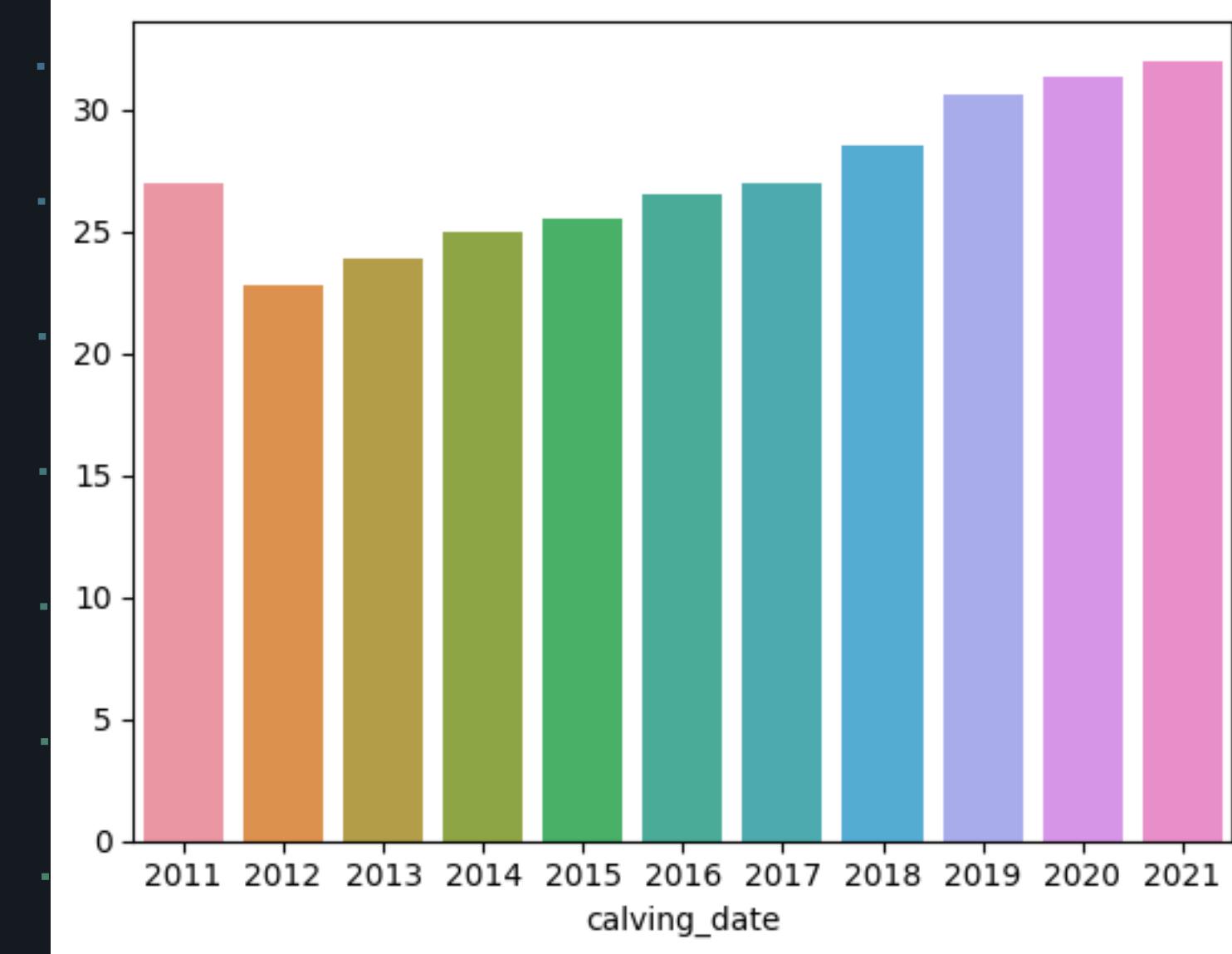
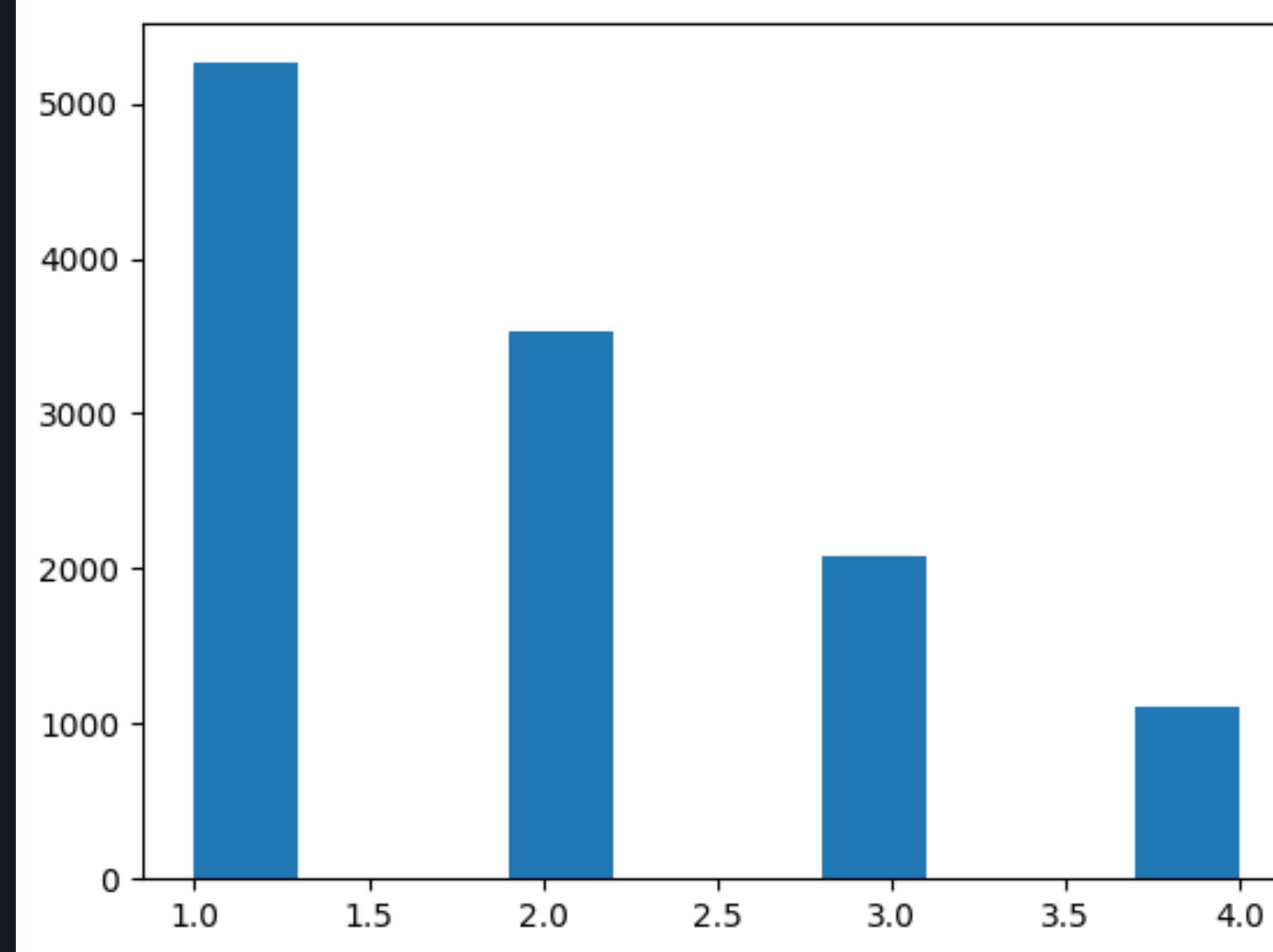
Private score - 7.0614354643665 RMSE

5-th place

Валидационный сет

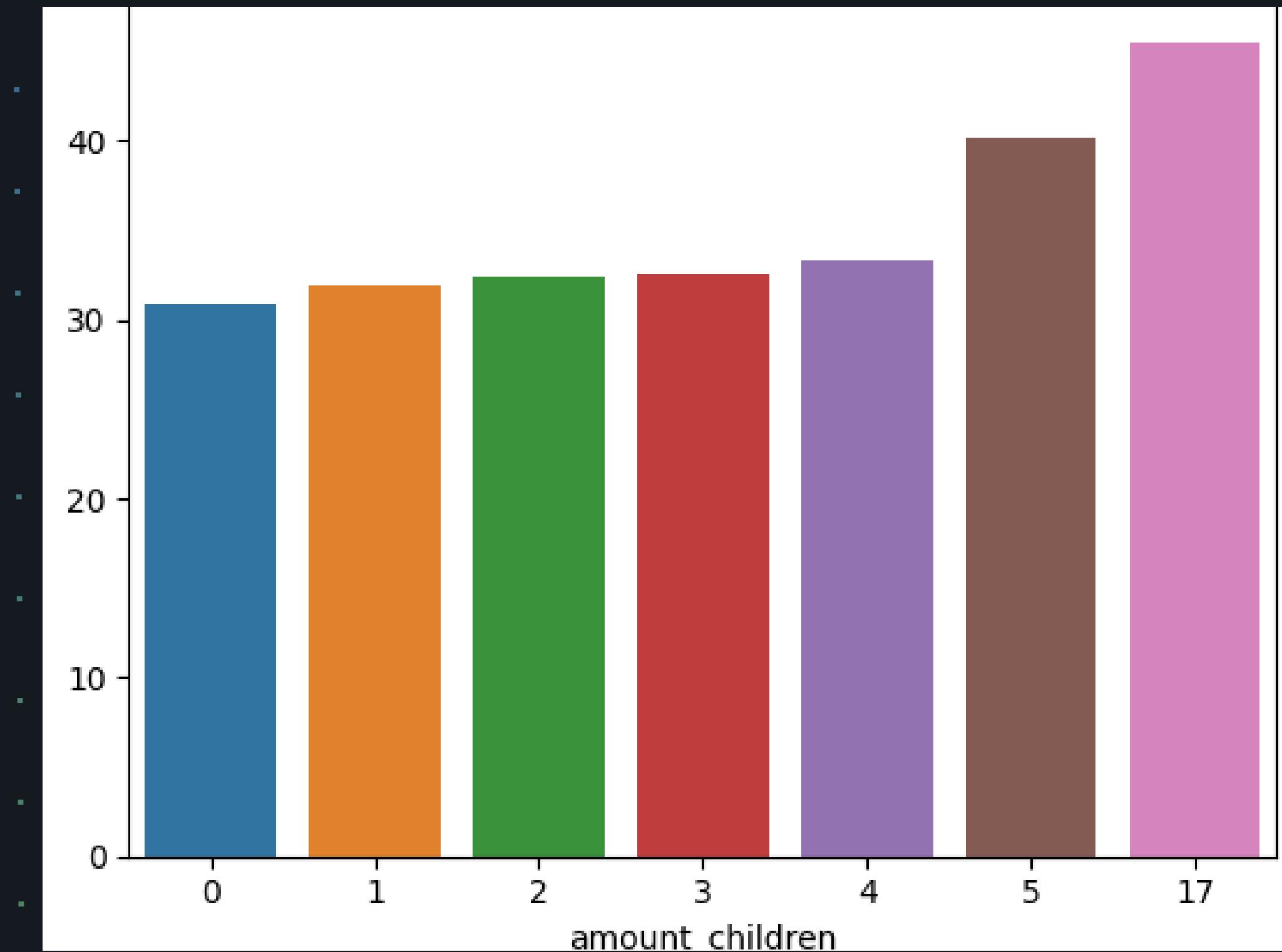
Валидационная выборка, отражающая тестовую. Одинакова по лактациям и по остальным параметрам

В тесте данные 2022 & 2021 годов, на валидации будут 2021
Обучать будем на данных после 2020



Полезные фичи

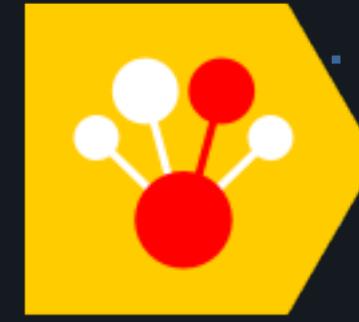
- 1 - Возраст коровы
- 2 - Разница дат между рождением теленка и началом лактации
- 3 - Количество дней, прошедших с предыдущей лактации
- 4 - Данные с предыдущих лактаций
- 5 - Кол-во телят
- 6 - Месяц рождения и лактации
- 7 - Средний удой матери
- 8 - Средний удой бабушек



Не такие полезные фичи

- 1 - Количество сестер / братьев по отцу
- 2 - Количество сестер / братьев по матери
- 3 - средний удой сестер

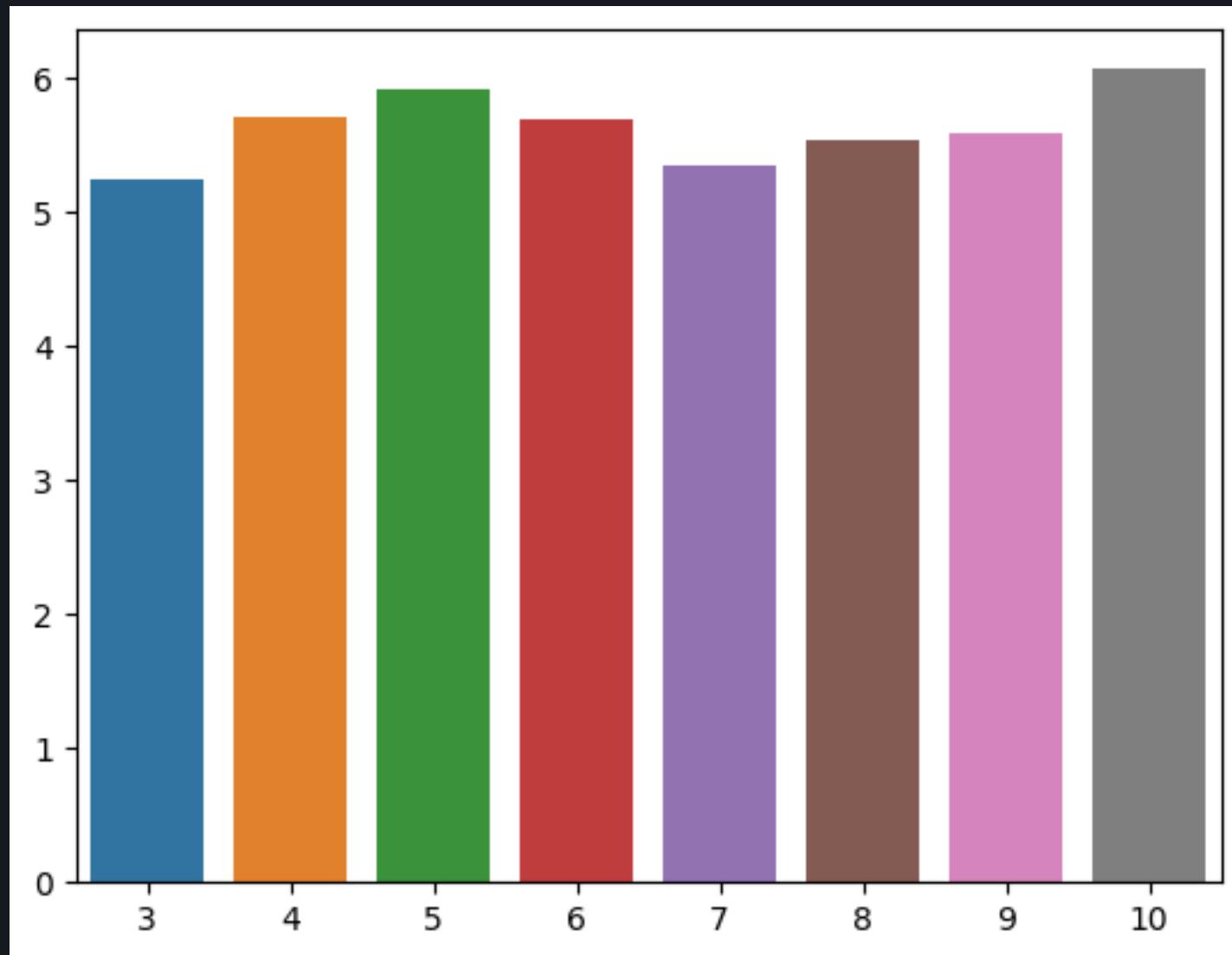
Модель



Регуляризованный CatBoost
для каждого таргета

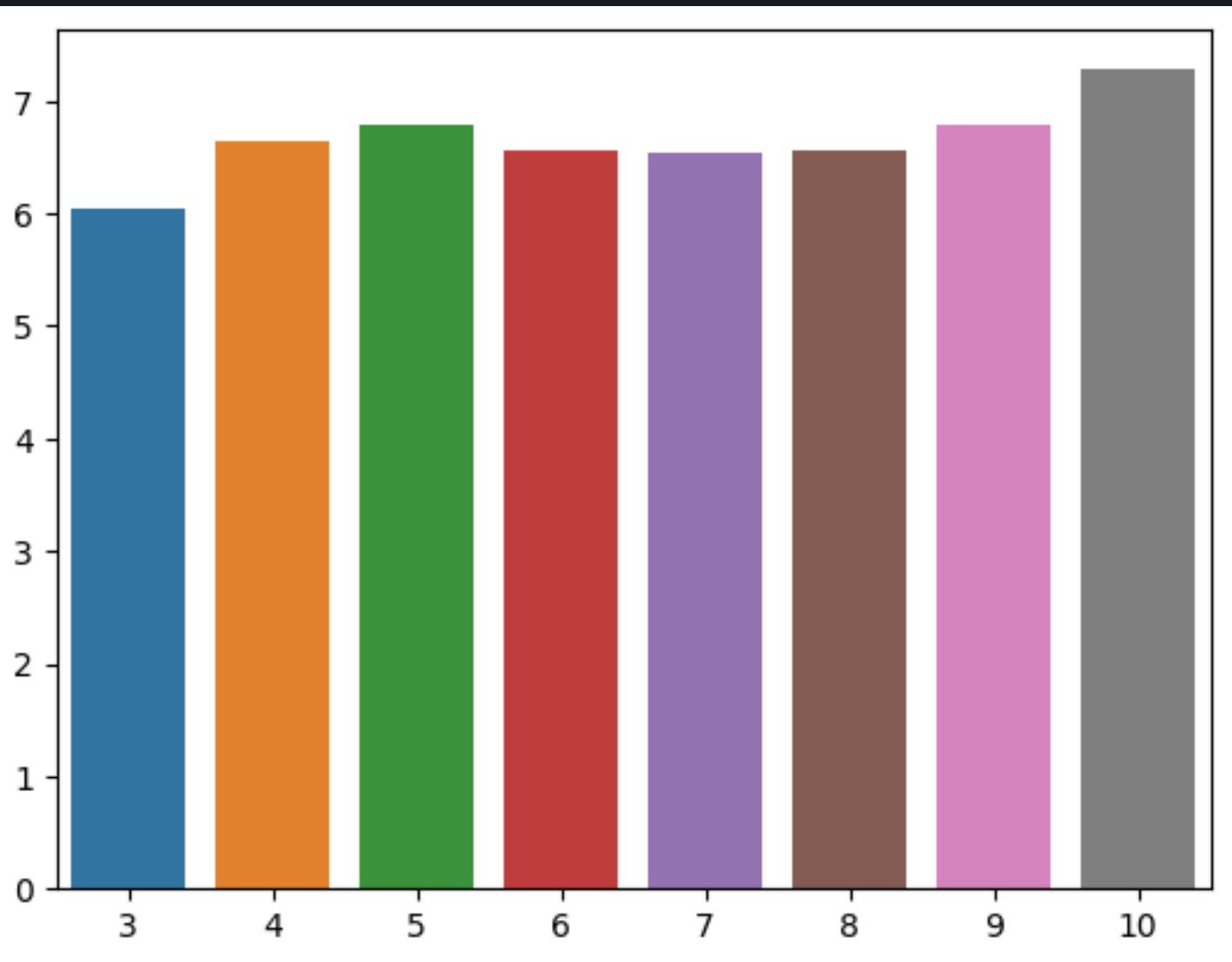
Train

Mean RMSE - 6,05



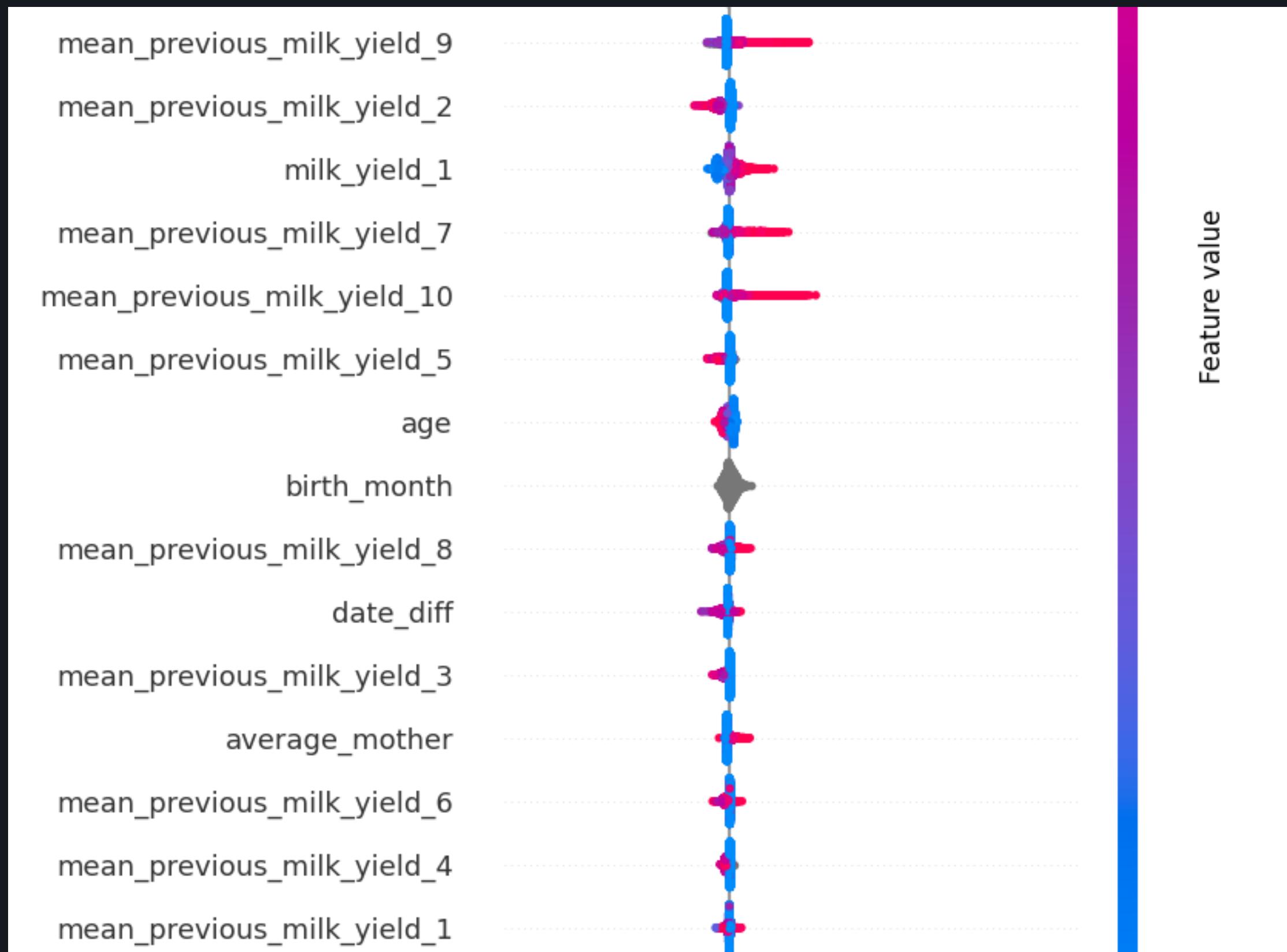
Validation

MEAN RMSE - 6,8



Интерпретация

SHAP & Permutation importance



SHAP

