Как подружить бизнес и GBDT: интерпретация моделей

Данила Савенков, МТС
dssavenk@mts.ru
@danila_savenkov

Мой путь к Data Science

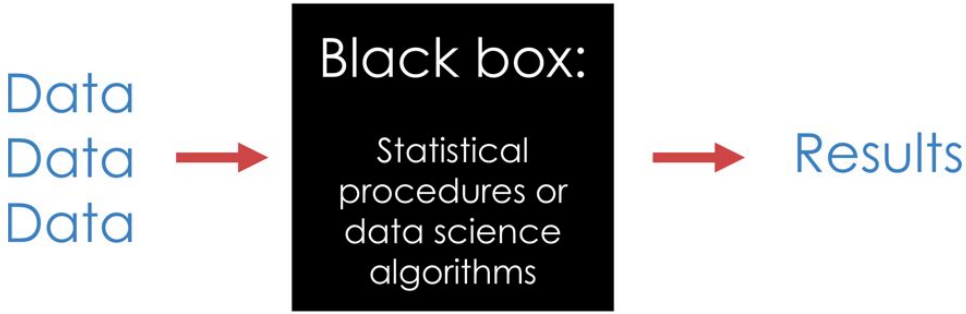
- Физический факультет СПбГУ
- Московская Консерватория
- Оркестр Московской Филармонии
- Язык программирования C++ (Б. Страуструп)
- Stepik.org
- Coursera: Введение в машинное обучение
- WorldQuant
- ODS Community (ods.ai)
- Kaggle
- МТС
- ...

Data Science: с чего начать

- Специализация от Яндекса и МФТИ “Машинное обучение и анализ данных” на coursera
- OpenDataScience Community (ods.ai)
- Kaggle

Бизнес и black box

Почему модель выдает такое предсказание?



LIME

KDD 2016 Paper

Local Interpretable Model-Agnostic Explanations

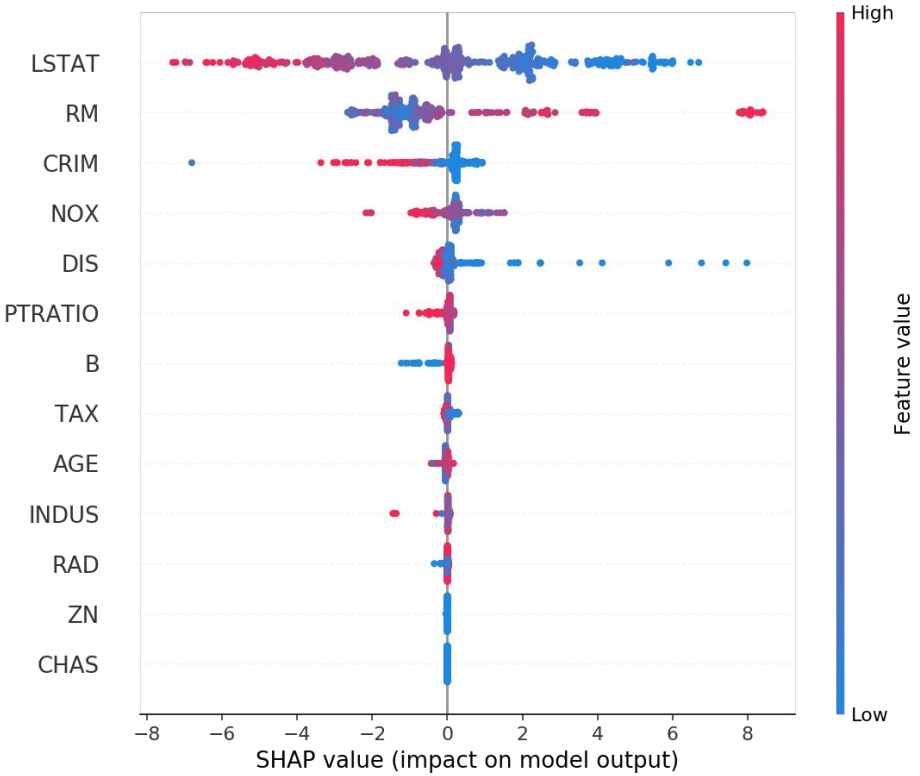
A Unified Approach to Interpreting Model Predictions

NIPS 2017 Paper

Интерпретация, основанная на теории игр

Объединяет в себе предыдущие подходы: LIME, eli5, feature importances

Реализована в пакетах xgboost, lightgbm, catboost

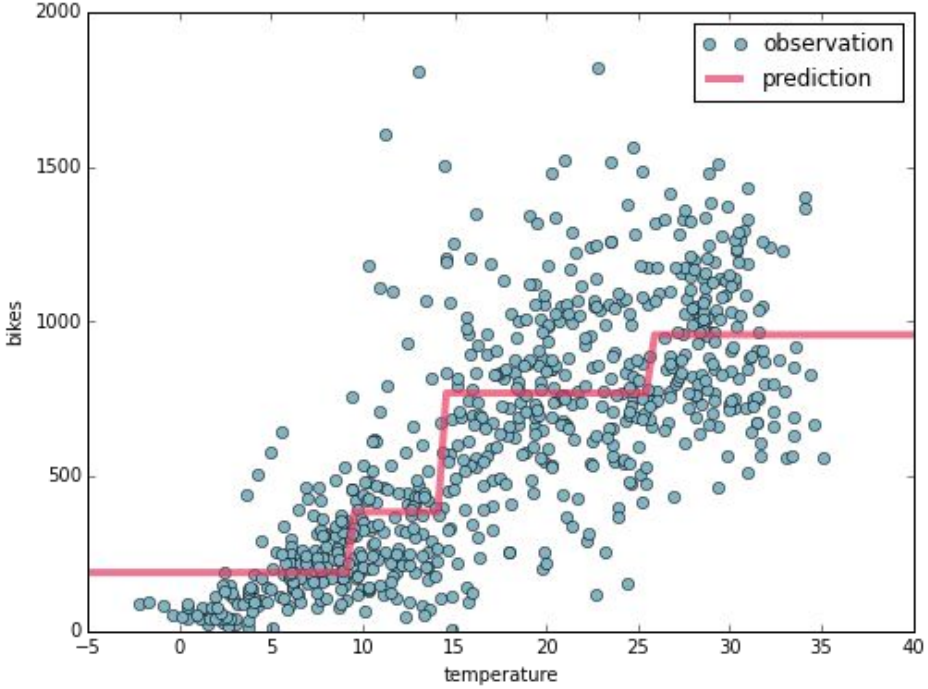


dssavenk@mts.ru

Feature importances: weight

Больше значений признака - больше сплитов

Misleading



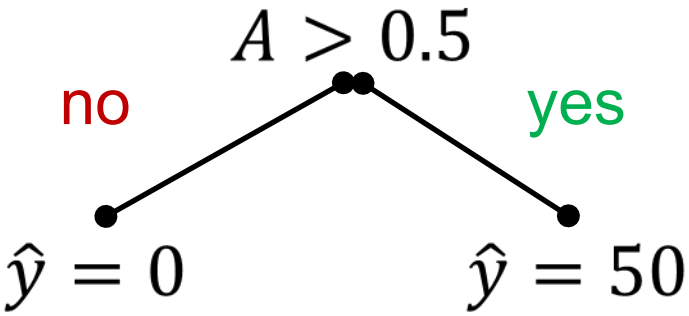
Feature importances: gain

$$\sum (y - \bar{y})^2 = 15000$$

Ошибка предсказания средним

A	B	Y
0	0	0
0	0	0
0	1	0
0	1	0
1	0	0
1	0	0
1	1	100
1	1	100

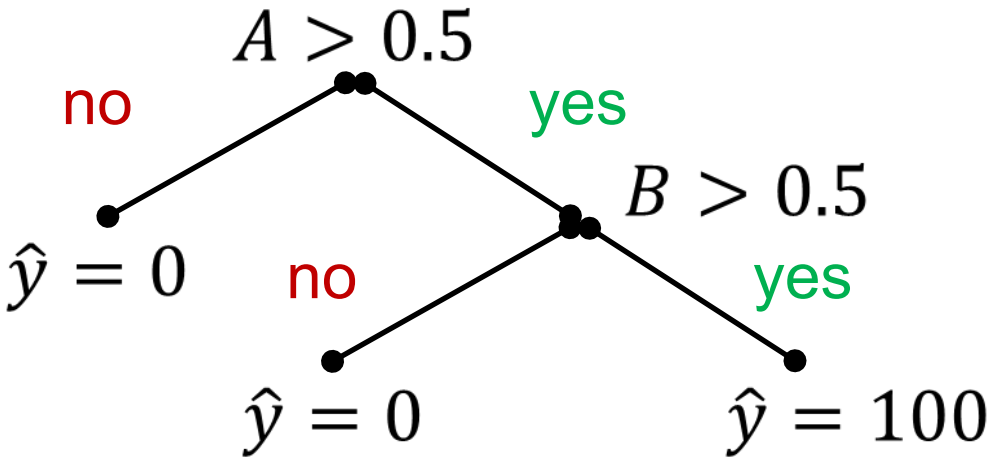
Feature importances: gain



A	B	Y
0	0	0
0	0	0
0	1	0
0	1	0
1	0	0
1	0	0
1	1	100
1	1	100

- Ошибка = 10000
- Gain(A) = 5000

Feature importances: gain

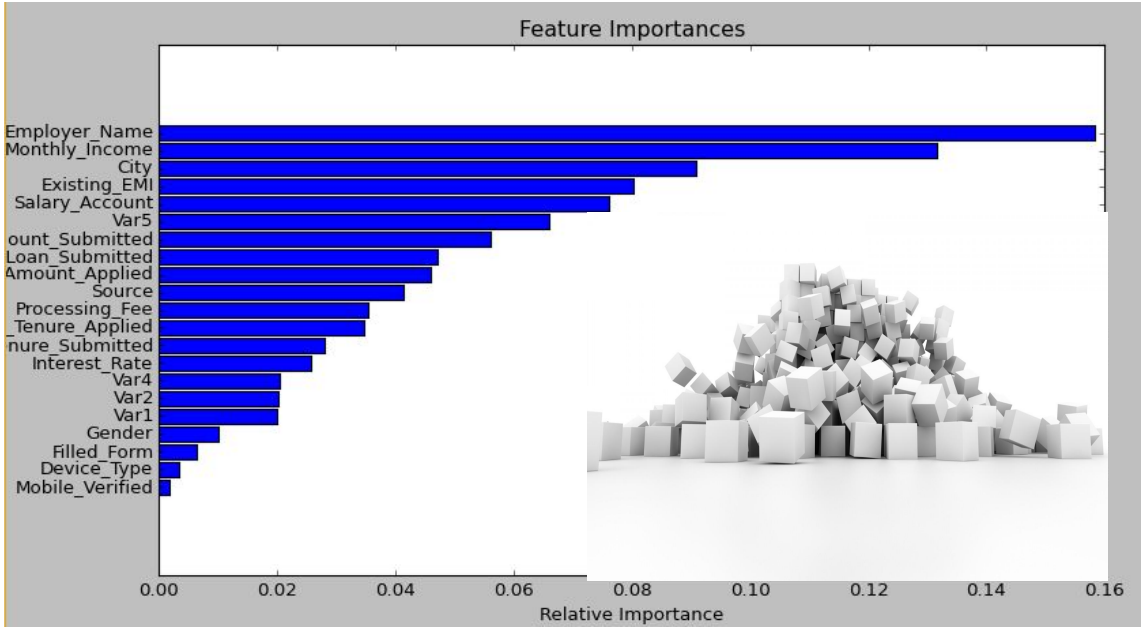


A	B	Y
0	0	0
0	0	0
0	1	0
0	1	0
1	0	0
1	0	0
1	1	100
1	1	100

- Ошибка = 0
- Gain(A) = 5000
- Gain(B) = 10000

Feature importances: gain

- Оценки по gain смещенные
- Не понятен характер зависимости

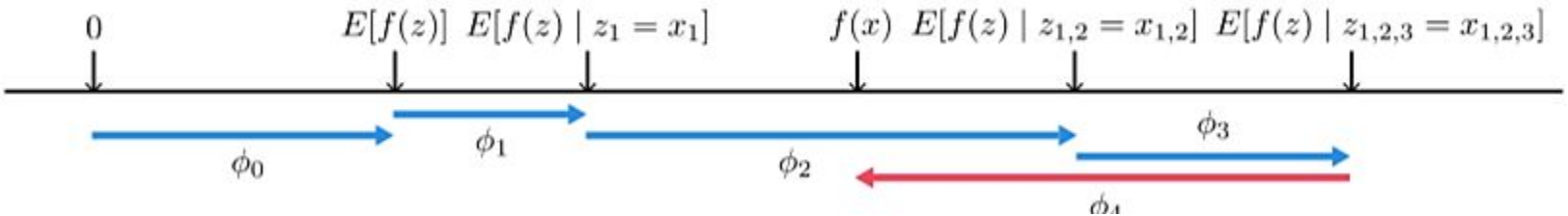


Feature importances: другие подходы

Permutation importance

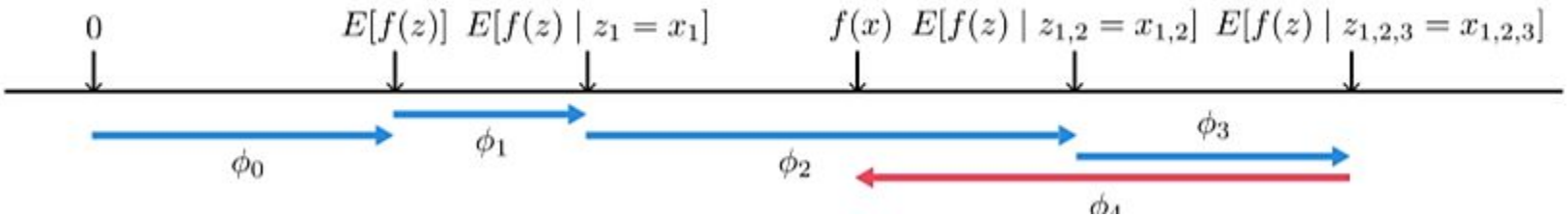
LIME, Eli5

SHAP - SHapley Additive Explanation

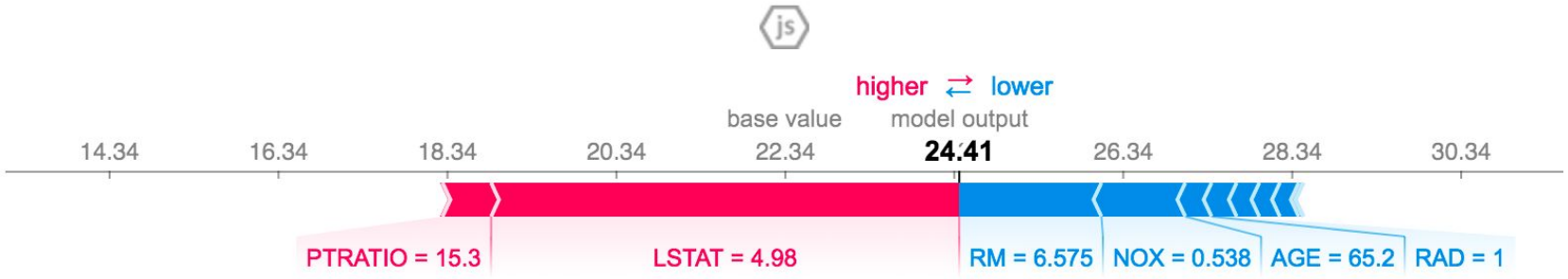


SHapley Additive Explanation

- Оцениваем вклад каждой фичи в итоговое предсказание на каждом объекте
- Shapley Sampling позволяет сделать это за приемлемое время

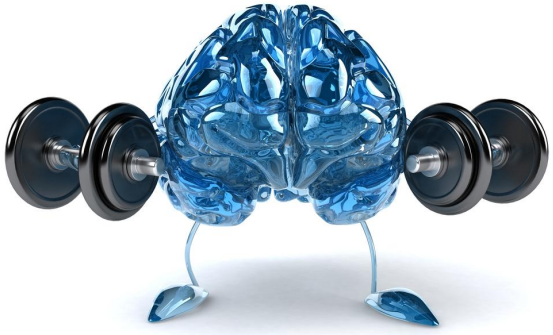


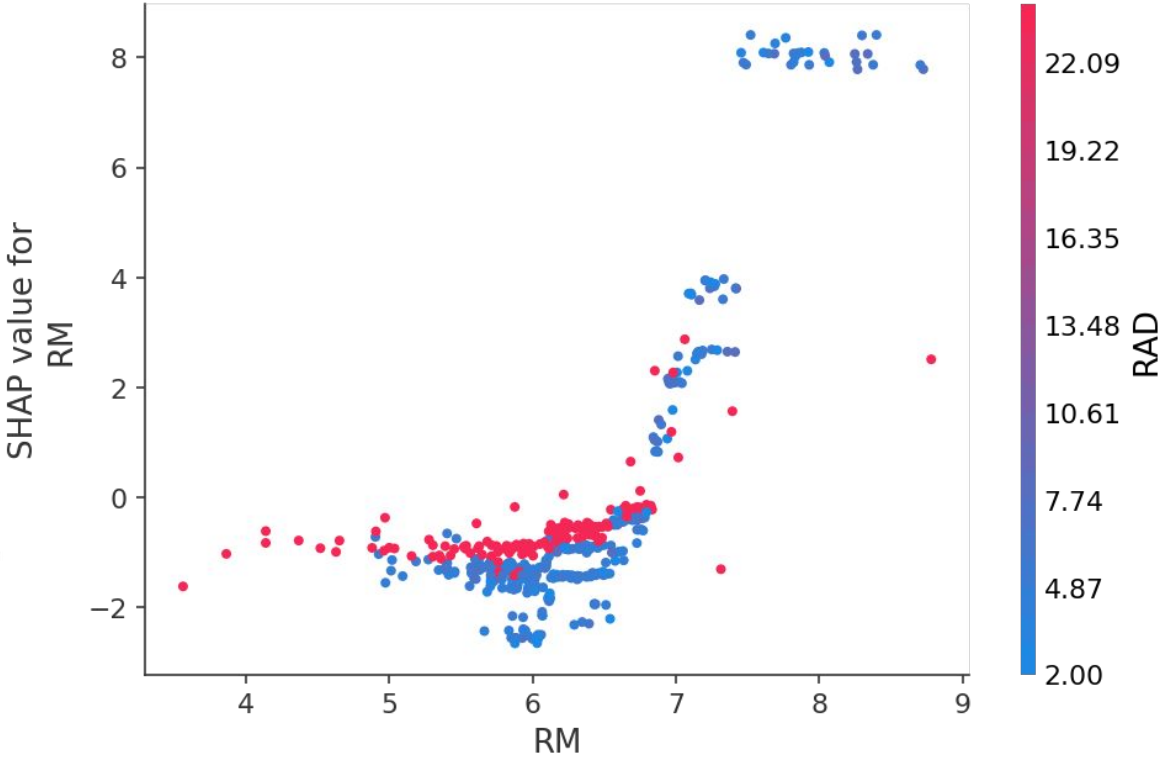
SHAP: на одном сэмпле



- Знание о том, что LSTAT = 4.98 увеличивает предсказание модели ~ на 5
- Знание о том, что RM = 6.575 уменьшает предсказание модели ~ на 1.5
- SHAP values: [n_samples, n_features]

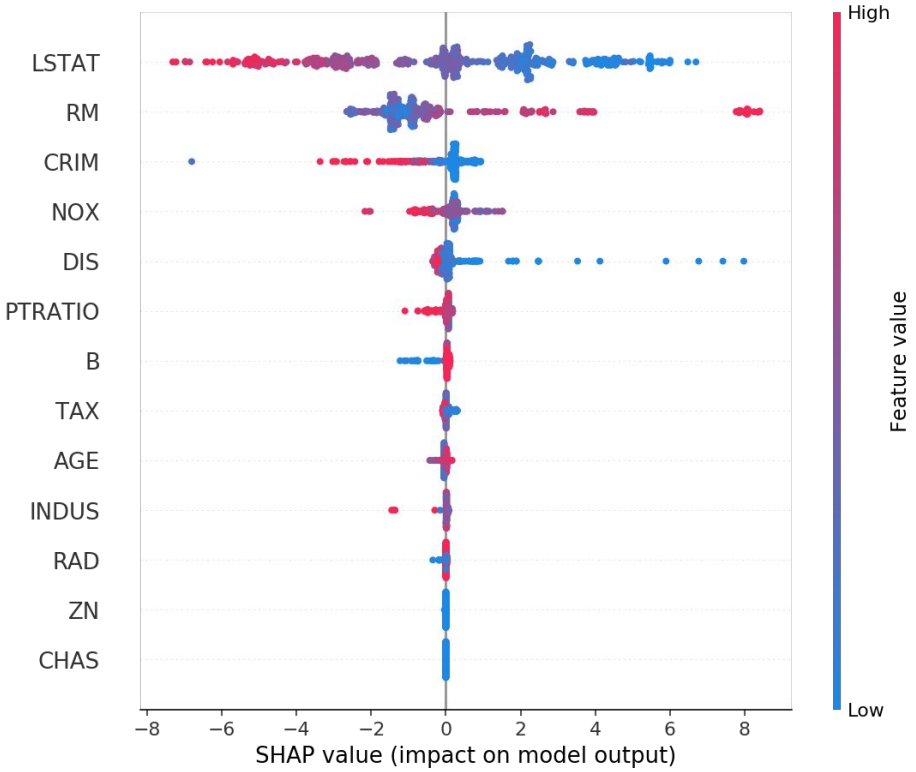
SHAP: возможности





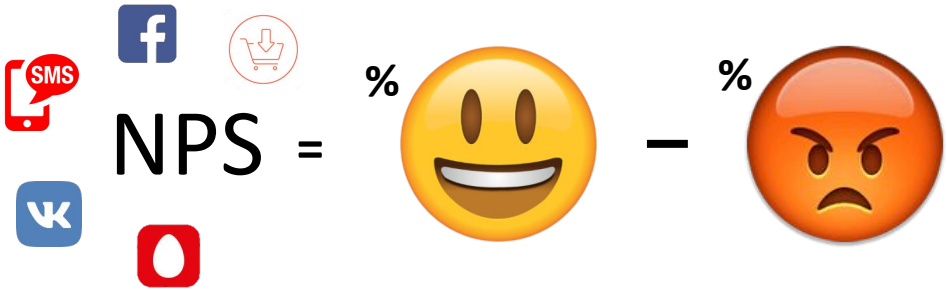
SHAP: возможности

Виден характер зависимости предсказания от признаков



CUSTOMER JOURNEY АБОНЕНТА НА ОСНОВЕ ОЦЕНОК NPS

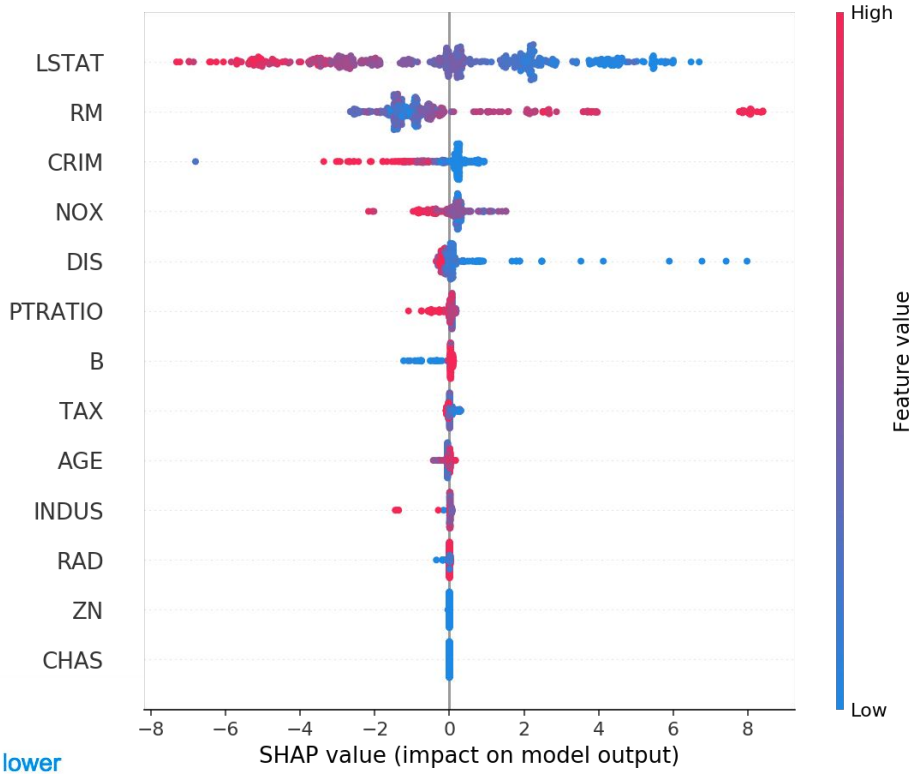
- Замер индекса лояльности в каждой точке контакта
- Классификация абонентской базы на критиков, нейтралов и промоутеров
- Customer Journey абонента до активации критика или промоутера

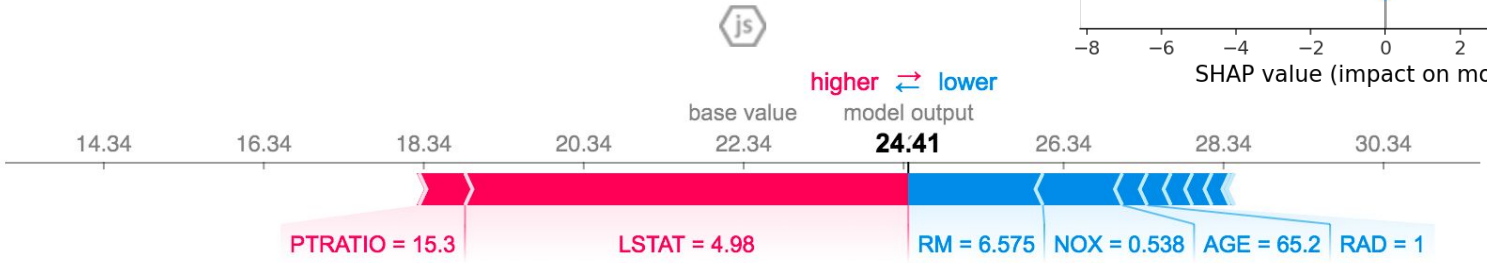


dssavenk@mts.ru

Предсказание NPS

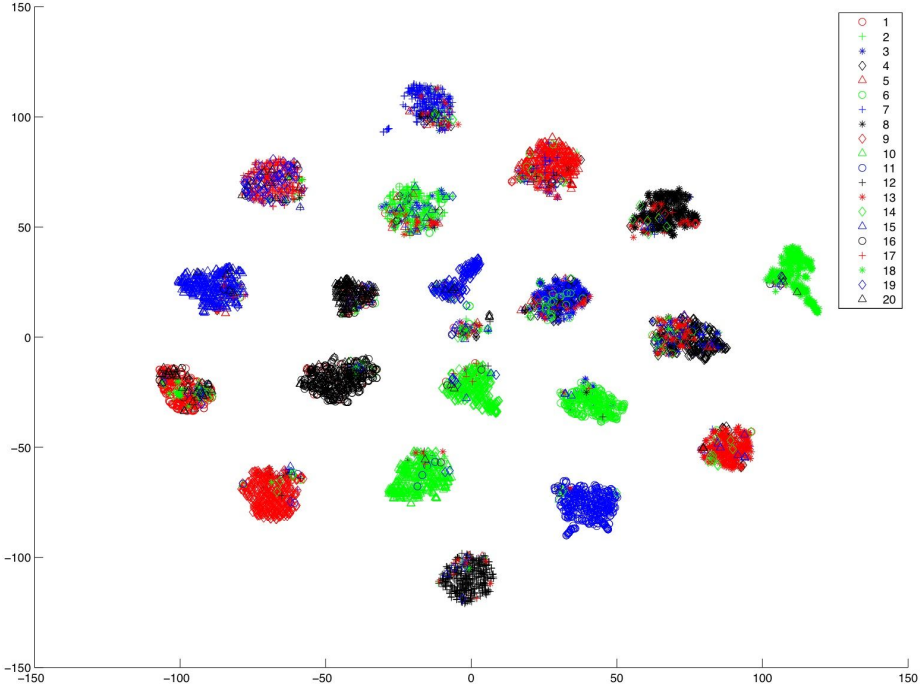
- Модель - GBDT
- Понимаем и объясняем бизнесу почему абонент “критик”





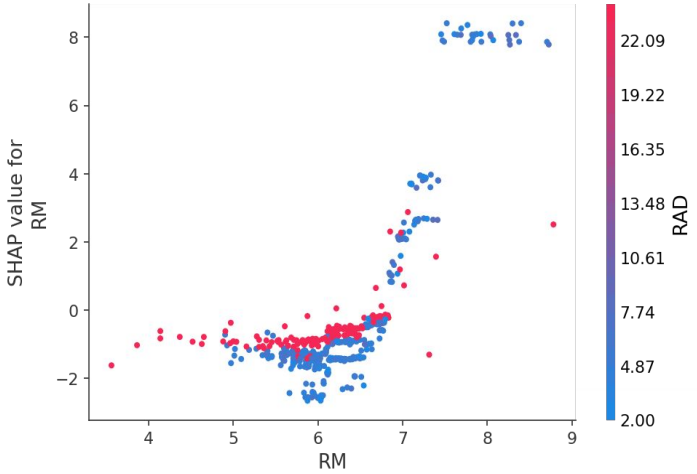
Кластеризация абонентов SHAP

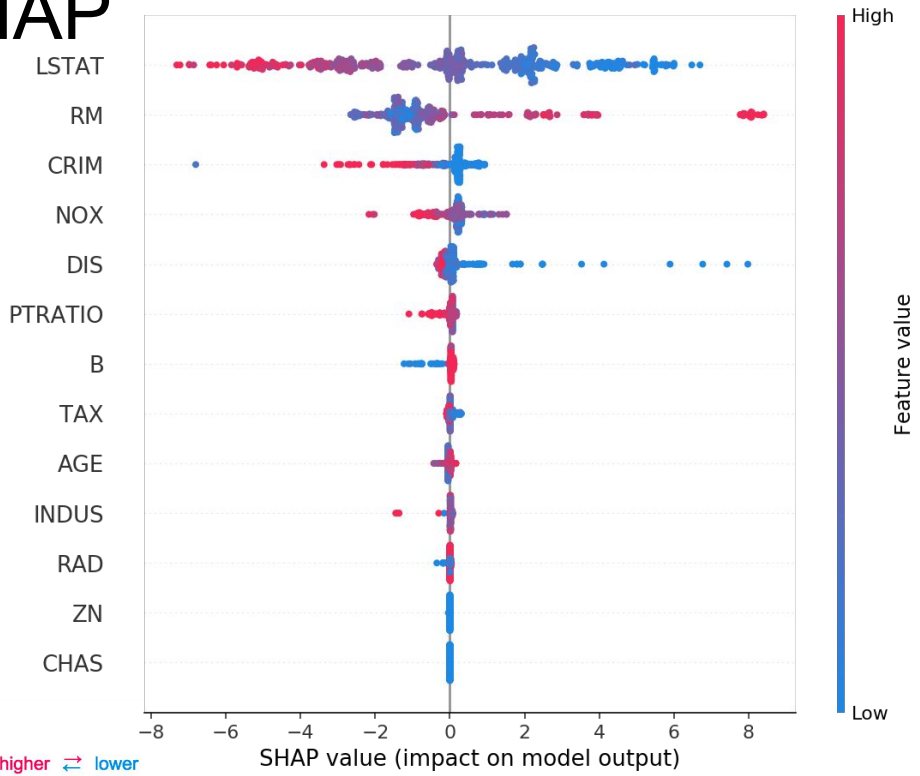
Выделяем кластеры “критиков” по SHAP

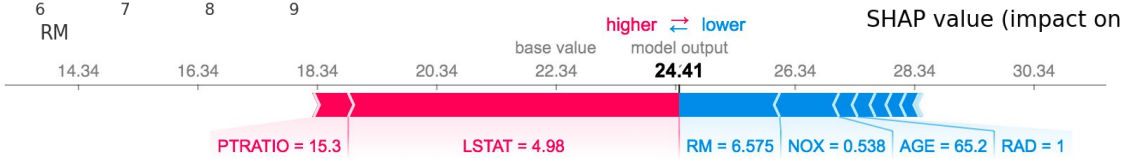


Градиентный бустинг + SHAP

- Почти интерпретируемая модель
- Дает новый взгляд на данные
- Понятна бизнесу







Спасибо за внимание!

Данила Савенков, МТС
dssavenk@mts.ru
@danila_savenkov