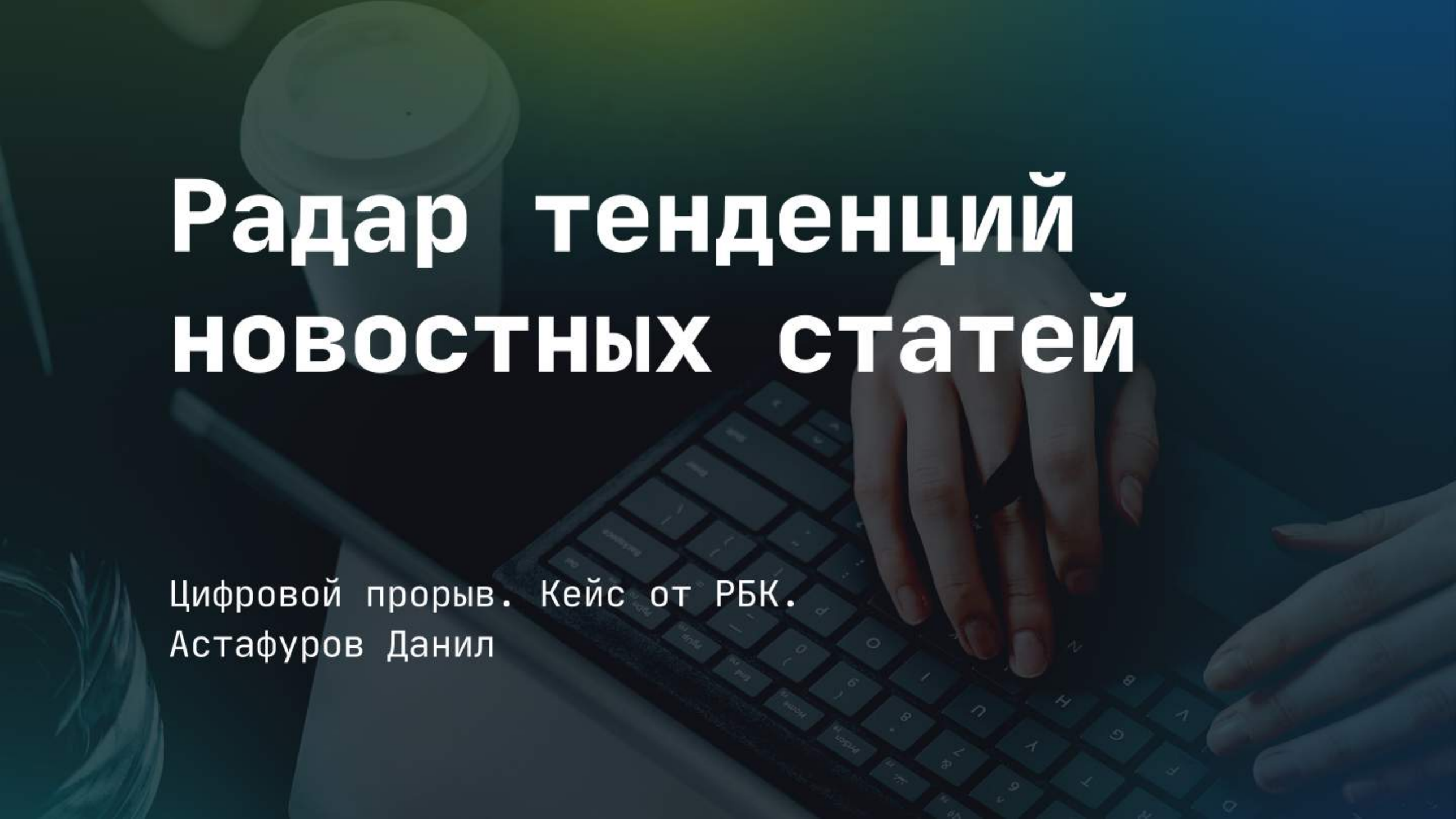

Радар тенденций новостных статей

Цифровой прорыв. Кейс от РБК.

Астафуров Данил

Парсинг

Текст статьи

Наличие картинки

неТекстовые фичи

длина текста/заголовка в символах/словах

год, месяц, день, сезон, время суток

количество дней с публикации

теги - one-hot + отбор (встречается > 25)

авторы - one-hot

Текстовые фичи

sberbank-ai/ruRoberta-large
sberbank-ai/sbert_large_nlu_ru
sberbank-ai/sbert_large_mt_nlu_ru
sberbank-ai/ruBert-large
sberbank-ai/ruBert-base
cointegrated/rubert-tiny2
DeepPavlov/rubert-base-cased-conversational
cointegrated/LaBSE-en-ru
microsoft/mdeberta-v3-base
vicgalle/xlm-roberta-large-xnli-anli
MoritzLaurer/mDeBERTa-v3-base-mnli-xnli
facebook/bart-large-mnli

Обучение

3 CatBoostRegressor с разными гиперпараметрами
обучение каждой на 5 фолдах
отбор лучших фич с помощью feature_importances_