

# Принятие законопроектов

## Команда DarkSide

	<b>Danil Vagapov</b>	Member
	<b>Danie1Nash</b> (you)	Member
	<b>Nikitochka Hrum</b>	Leader

**Вагапов Д.Д.**

**Челпанов В.А.**

**Понькин Н.А.**

# ПЛАН ДОКЛАДА

1. Анализ данных
2. Спроектированная модель
3. Полученные результаты

Анализ

Модель

Результат



# 1. АНАЛИЗ

0,698% данных в обучающей выборке содержали данные об удаленных проектах

	id	act_title	publication_date	developer	okved_list	views_num	comments_num	likes_num	dislikes_num	regulatory_impact
16195	23312	Проект удален	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16226	23374	Проект удален	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16834	24276	Проект удален	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16885	24351	Проект удален	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16919	24396	Проект удален	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Анализ

Модель

Результат



# 1. АНАЛИЗ - feature engineering

- разделение списка ОКВЭДов с помощью Label Encoding'a
- преобразование строчной информации о дате публикации к числовому виду

okved_list	okved_list_0	okved_list_1	okved_list_2	okved_list_3
государственное управление	1	0	0	0
сельское хозяйство; охота; лесное хозяйство	15	16	17	0
образование	5	0	0	0
государственное управление	1	0	0	0
образование	5	0	0	0



Анализ

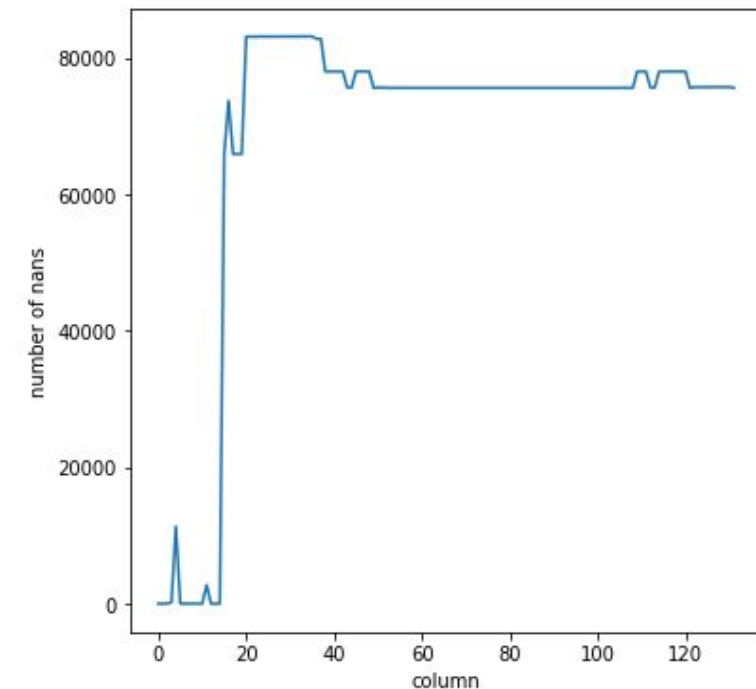
Модель

Результат



# 1. АНАЛИЗ - таблицы `ria_reports`.

- объединение всех таблиц в одну при помощи данного уникального идентификатора для каждого закона
- пропуски были заполнены самыми частыми значениями по столбцу



Анализ

Модель

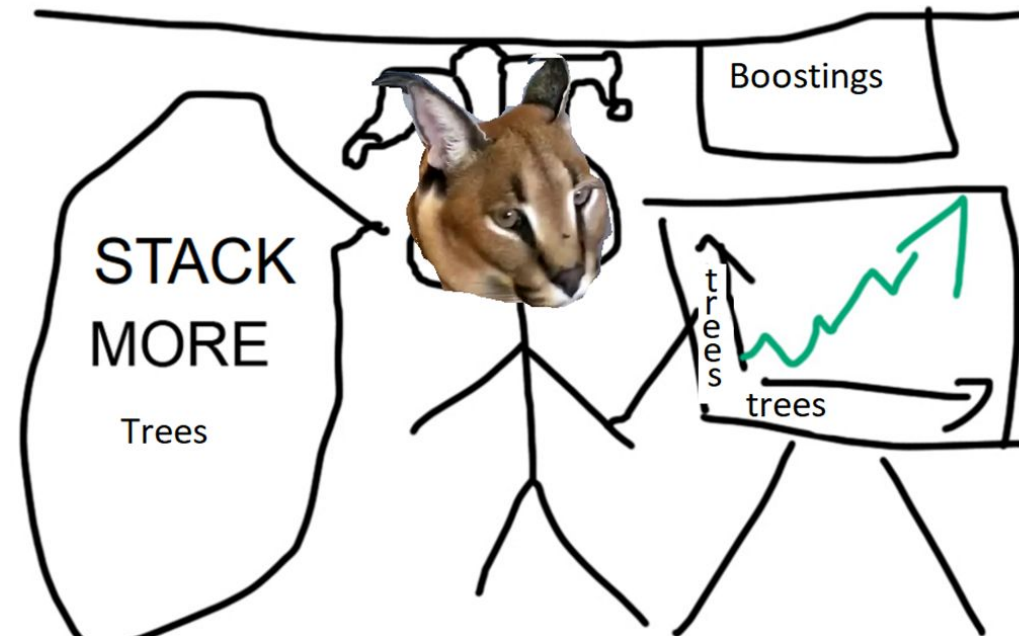
Результат

## 2. МОДЕЛЬ [baseline]

```
cat_cols = [col for col in df.columns if df[col].dtype=='object'] # brute
```



Yandex  
CatBoost



[https://catboost.ai/en/docs/concepts/python-reference\\_catboostclassifier](https://catboost.ai/en/docs/concepts/python-reference_catboostclassifier)

## 2. ДРУГИЕ МОДЕЛИ

В ходе решения поставленной задачи были использованы следующие подходы:

- Bert as feature extractor

Layer (type:depth-idx)	Output Shape
-----	
BertFeatureExtractor	--
└BertModel: 1-1	[24, 768]
└Dropout: 1-2	[24, 768]
└Linear: 1-3	[24, 32]
└Linear: 1-4	[24, 1]

- TabNet embeddings
- CatboostClassifier на новых признаках

<https://habr.com/ru/post/436878/>

Анализ

Модель

Результат

## 3. РЕЗУЛЬТАТ

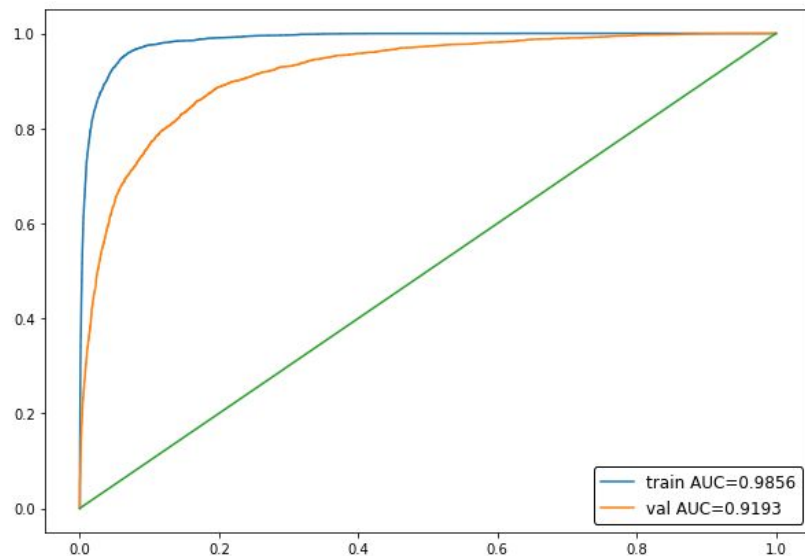
“Базовое” решение

Public Score

0,91052

Private Score

0.89204



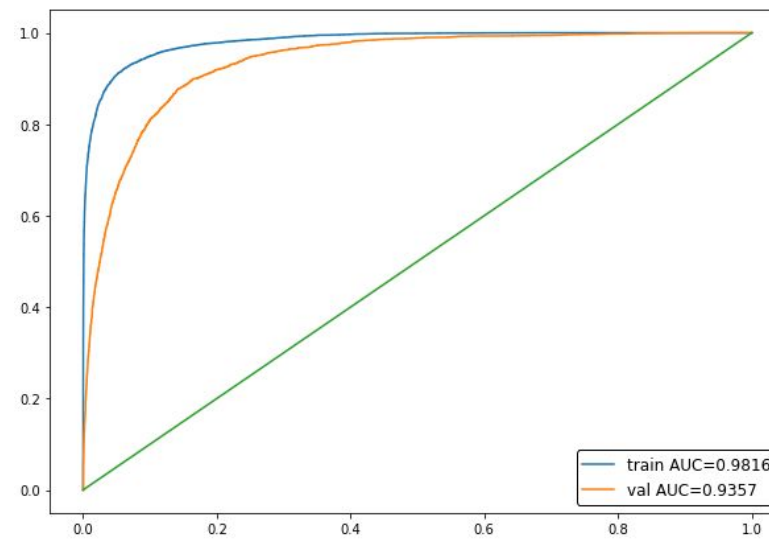
“Продвинутые” модели

Public Score

0.89426

Private Score

0.87304





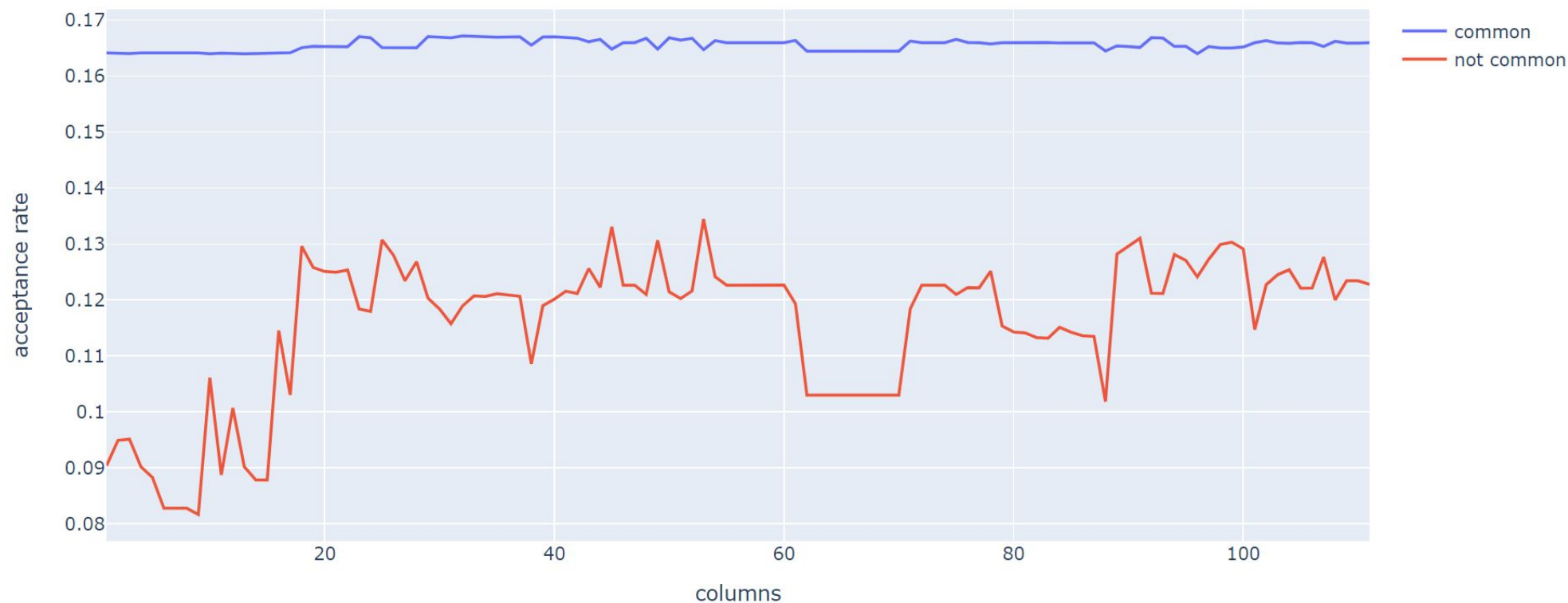
Анализ

Модель

Результат



## 3. РЕЗУЛЬТАТ [интерпретация]



**СПАСИБО ЗА ВНИМАНИЕ!**

**ГОТОВЫ ОТВЕТИТЬ НА ВАШИ  
ВОПРОСЫ!**



Состав команды:

Вагапов Данил  
tg:@addast

Понькин Никита  
tg:@ponkinikita

Челпанов Василий  
tg:@danielnash