

# Классификация новостных текстов с присвоением тегов

Команда:

**Schneeflocken nur essen**

Участники:

Васильев Михаил - [t.me/LaHundo](https://t.me/LaHundo)

Евдокимов Александр - [t.me/snakerzr](https://t.me/snakerzr)

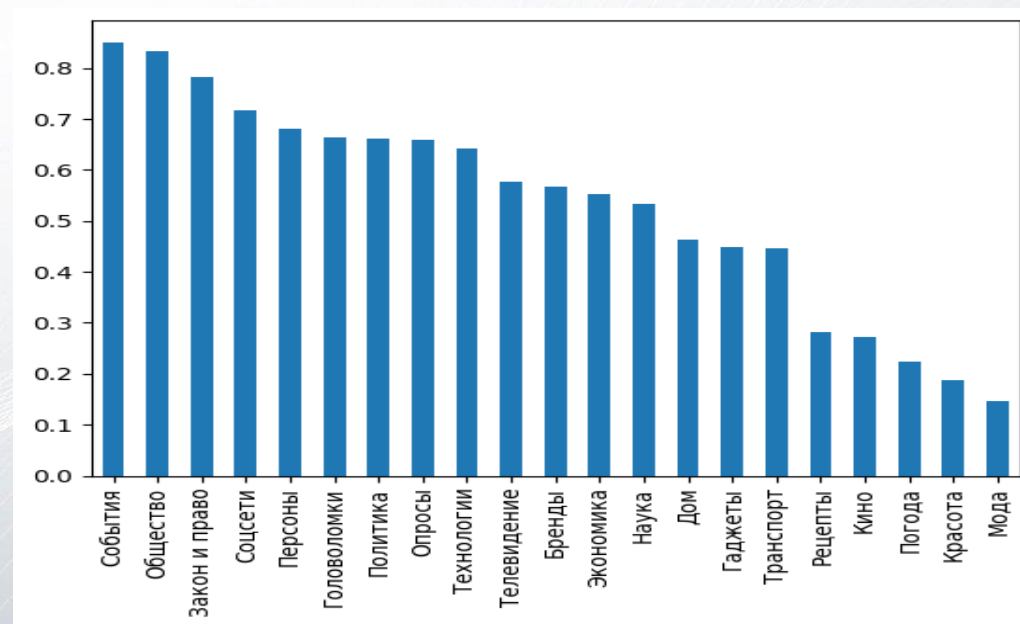
Иванова Ольга - [t.me/oakarabut](https://t.me/oakarabut) open to work

Метелёв Никита - [t.me/SciManNik](https://t.me/SciManNik) open to work

Скорина Валентина - [t.me/valentina\\_lineyka](https://t.me/valentina_lineyka) open to work

## Поиск данных для обучения

- интерфакс 56 000 статей



- данные из Notion 56 000 статей (4 тега из 21)



# Первая модель на LLM

Использовали модель **Saiga3 8b** для назначения тегов:

- модель по каждому из 21 тегов даёт соответствие статье
- выделяет из найденных тегов наиболее релевантные
- создает дополнительные уточняющие подтеги

**ИТОГ: Первая медленная модель готова!**

- Используем эту модель для разметки новостей интерфейса

# Schneeflocken nur essen

VK x HSE  
Data Hack

Модель **zero-shot**

Использовали мультиклассовую классификацию **cointegrated/rubert-base-cased-nli-threeway** для назначения тегов

Catboost – дал недостаточное качество

Итоговая модель 2

**cointegrated/rubert-tiny2** на размеченных данных

F1 micro - **0.91**

ROC AUC score - **0.95**

# Schneeflocken nur essen

VK x HSE  
Data Hack

## Визуализация данных эмбединги

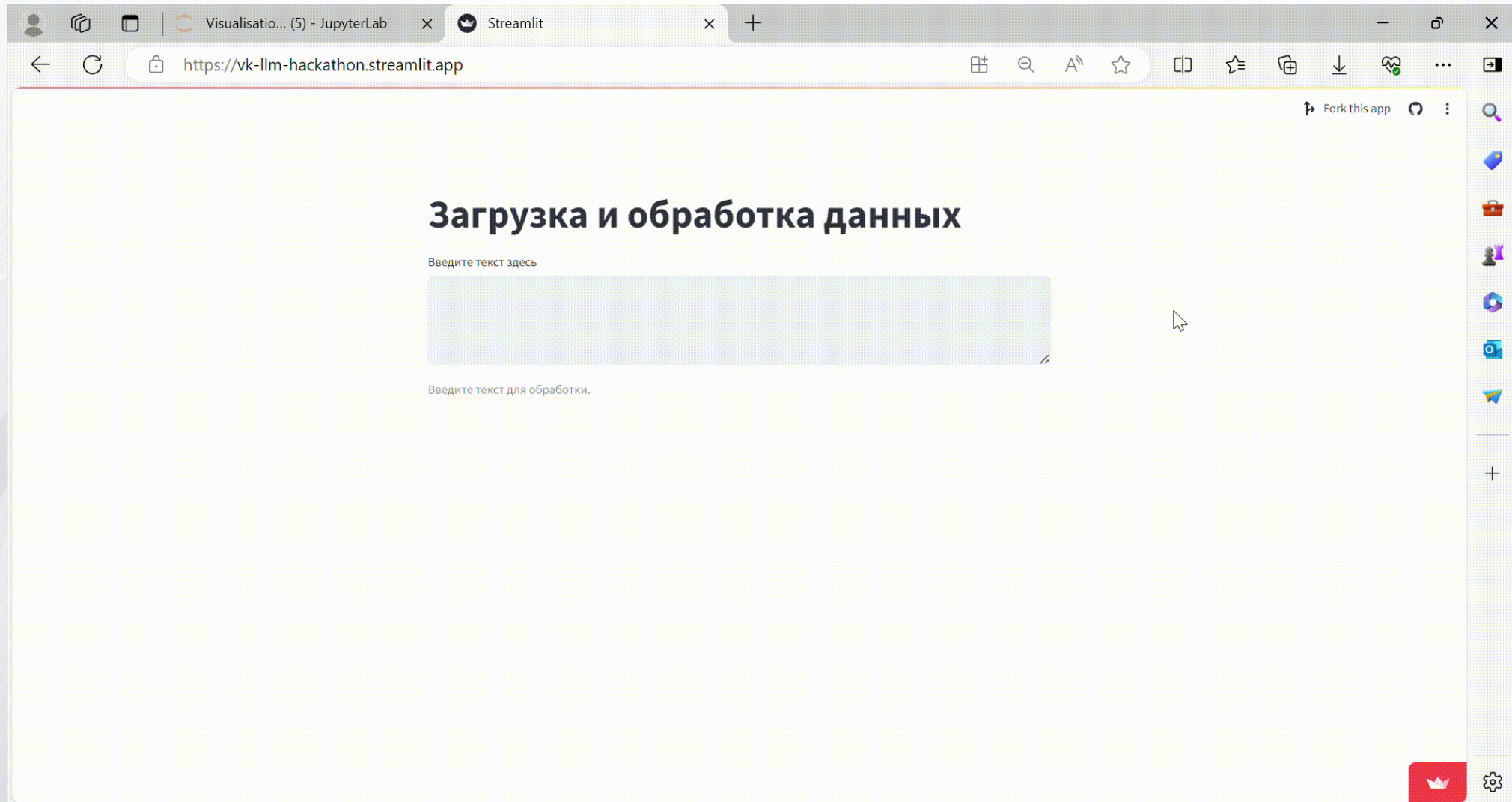




# Schneeflocken nur essen

## Как это работает

VK x HSE  
Data Hack



Наш сервис: <https://vk-llm-hackathon.streamlit.app/>

- Результаты: гибкое присвоение тегов, скорость работы, хорошие метрики, расширяемость
- Точки роста: повышение качества разметки, уточнение списка тегов

Schneeflocken nur essen

VK x HSE  
Data Hack

# Ваши вопросы?