



# BASNya

Basketball Auto-  
generated  
Statistical News

# Суть проекта

Автоматическая генерация твитов на основе аномальных\интересных выступлений игроков НБА

Демо-стенд проекта: <https://basnya.streamlit.app>  
Пароль: basnya\_weak\_password

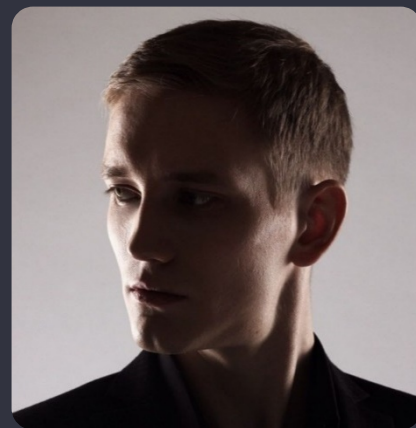
Код проекта: <https://github.com/no-one2k/basnya>

# Алексей

MANAGER

Lorem ipsum sed dolor sit  
sed amet, consectetur  
adipiscing incidunt ut  
labore et dolore nostrud  
exercitation ullamco

MORE INFORMATION



# Богдан

MANAGER

Lorem ipsum sed dolor sit  
sed amet, consectetur  
adipiscing incidunt ut  
labore et dolore nostrud  
exercitation ullamco

MORE INFORMATION



# Борис

MANAGER

Lorem ipsum sed dolor sit  
sed amet, consectetur  
adipiscing incidunt ut  
labore et dolore nostrud  
exercitation ullamco

MORE INFORMATION

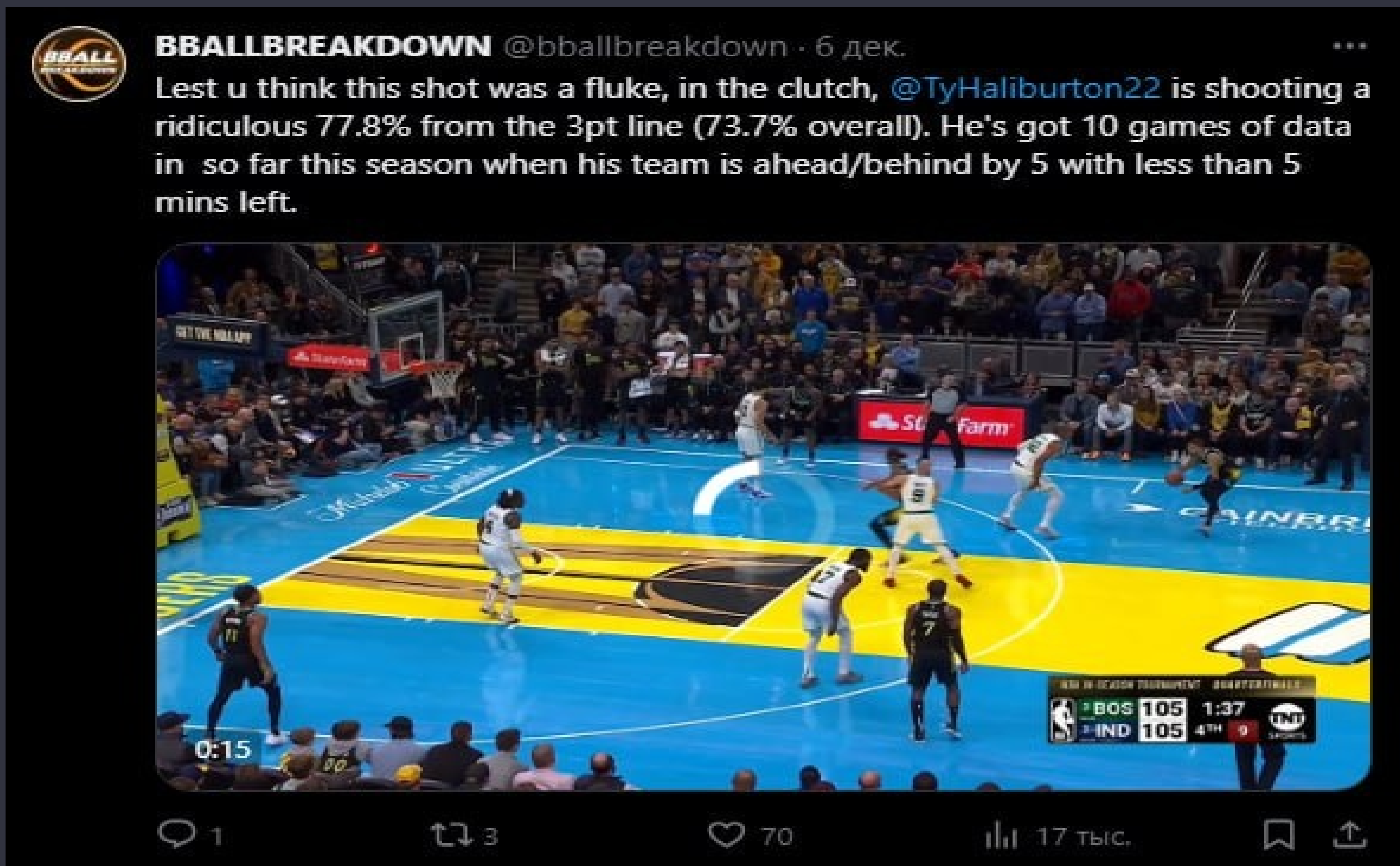


# Для кого\чего мы делали проект



Coach Nick

- Nick Hauselman
- **Твиттере**: 144.5K
- **Ютубе**: 866K



# Цели проекта

- Автоматизировать создание контента на стыке баскетбола и статистики
  - } Меньше трудозатрат
  - } Чаше публикации
  - } Меньше пауза между окончанием игры и выходом контента
  - } Больше разнообразия
  - } Шире охват команд и игроков
- Освоение техник и инструментов построения ML систем
  - } ML Design Doc как методологический подход
  - } LLM
  - } Инструменты для ETL, ML Observability и прототипирования UI

# Основной сценарий использования

(ЗДЕСЬ НУЖНЫ СКРИНШОТЫ или на отдельном слайде)

- 1) Пользователь авторизуется в аппе и выбирает дату, за которую он хочет проанализировать игры
- 2) Игры для этой даты загружаются в хранилище
- 3) Пользователь выбирает какие игры ему хочется проанализировать
- 4) Для выбранных игр происходит поиск потенциально интересных фактов
- 5) По этим фактам генерируются твиты
- 6) Пользователь публикует твит(ы) в исходном виде или проводит дополнительное ручное исследование на их основе

# Что мы строили

Система сбора баскетбольной статистики

+

Анализ и выявление потенциально интересных фактов

+

Генерация контента

# Предположения и ограничения

## 1. Знания о пользователе

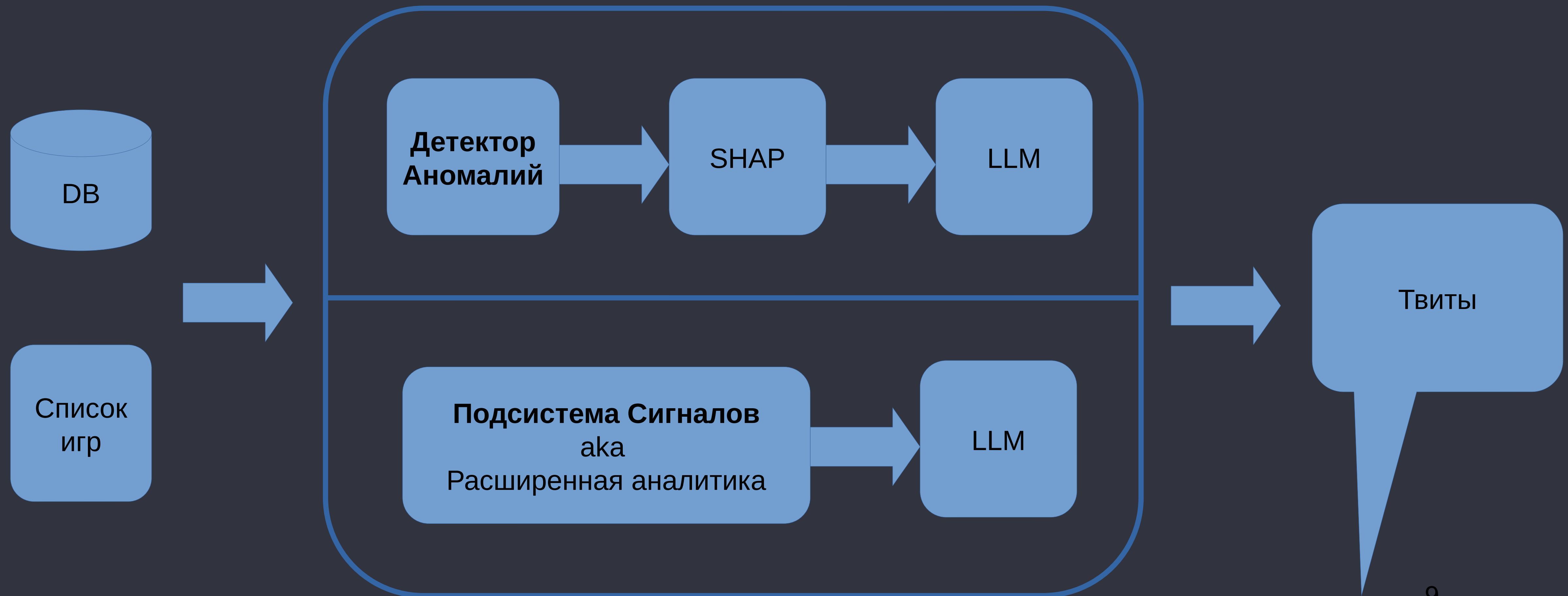
- англоязычный
- владеет python и средствами анализа данных

## 2. Знания о задаче:

- разработка и эксплуатация должны быть дешёвыми
- 1 игровой день обрабатывается менее, чем за 1 час
- интересность контента сложно оценить => нужен отбор человеком



# Карточка модели. Схема



# Карточка модели. Детали

1. Данные получены с портала `stats.nba.com` с помощью пакета ``nba_api``
  - для обучения детектора аномалий — протоколы игр регулярных сезонов НБА 2021-22 и 2022-23
  - для подсистемы сигналов — протоколы игр текущего сезона НБА (2023-24)
2. стек: `python + prefect + sklearn + shap + langchain + openAI + streamlit`
3. Fairness
  - не используем пол\расу\национальность\возраст
  - сторонняя LLM как чёрный ящик

# Какие были сложности

1. Слишком широкая постановка задачи в самом начале
2. Отсутствие прямых метрик для более узкой задачи: как измерить и сравнивать интересность?
3. PYOD и SHAP не дружат между собой
4. Отсутствие документации на данные
- 5\*. Галлюцинации и некорректное поведение LLM
- 6\*. Малое разнообразие и «искусственность» контента

Спасибо за  
внимание

