

**EMERGENCY  
DATAHACK**



ИНИД



МЧС  
РОССИИ

# Команда “Звёздочка”

Решение Трека №1 предсказание заторов льда  
на реке Лене



- Десятая река в мире по протяженности - более 4400 км
- Полностью замерзает зимой
- Бассейн реки занимает более 50% азиатской части России
- Мало населенных пунктов вдоль реки - непредсказуемая
- Ледяные заторы и сильные разливы
- Задача:
  - Предсказывать образование ледяных заторов на гидропостах в период весеннего ледохода по дням
  - Метрика F1

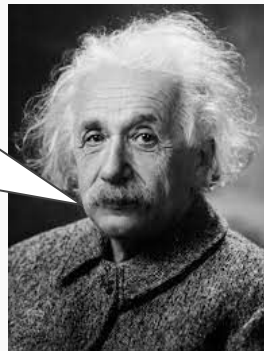
# Вызовы задачи

- Мало данных
  - ежесуточные и интервальные гидро и метео измерения измерения за 35 летний период (не слишком много с точки зрения задач машинного обучения)
- Сложные данные
  - Много отдельных наблюдений интуитивно связанных с задачей
  - Географически слабо связанные объекты
  - Относительная редкость заторов

# Как решали задачу

- Выявляем зависимости с помощью ML моделей
  - Основной вызов - подготовить данные на вход NN модели
    - События за 4х месячный интервал на каждом посту подавали на вход сети
    - Добавили географическую связность гидропостов как дополнительную ось
  - Попробовали рекуррентную сеть (GRU)
  - **Остановились на двойном трансформере на основе Albert**
- В чем преимущества подхода
  - Выжимаем из существующих данных максимум
  - В МЛ основная боль - это сбор и подготовка данных
  - Потенциал роста точности модели

Используйте  
трансформеры -  
это SOTA





- LenaBiTrans - трансформер в трансформере
  -

# Что подсказала модель

- Время
  - Самые важные периоды для предсказания заторов - ноябрь и февраль
- Толщина льда - самый важный признак
- Тепловой баланс для долгосрочного прогноза - второстепенная роль

# Рефлексия

- Модель легко расширять (добавляем новые признаки, не усложняя саму модель)
- Модель полностью воспроизводима (никакого шаманства над данными)
- Анализ важности признаков подтверждает обоснованность применения существующих эмпирических методов

**EMERGENCY  
DATAHACK**

Сделано с любовью  
на открытом ПО



ИНИД



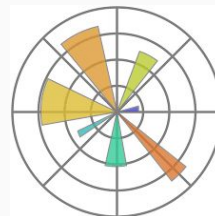
МЧС  
РОССИИ



OPTUNA



GitHub

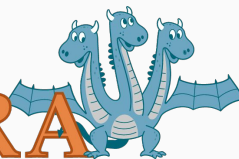


Poetry

pre-commit



HYDRA



archlinux



**EMERGENCY  
DATAHACK**



ИНИД



МЧС  
РОССИИ

# Команда Звёздочка



Сергей  
Lead ML Сбермаркет



Даня  
ML Сбермаркет



Глеб  
Lead ML Сбермаркет

