

ОНЛАЙН-ХАКАТОН

# EMERGENCY DATAHACK

ТРЕК 1



МЧС  
РОССИИ



ИНИД



# ТРЕК 1

Определение вероятности появления заторов льда на реке Лена в весенний период

## ОБЩЕЕ ОПИСАНИЕ ЗАДАЧИ

Построить предиктивную модель, которая позволит **за 60 дней рассчитать вероятность возникновения заторов льда на каждые сутки** в период ледохода около населенных пунктов: Киренск, Витим, Пеледуй, Крестовский Лесоучасток, Ленск, Олёкминск, Покровск, Якутск, Батамай, Сангар.

# ЗАЧЕМ МЫ ЭТО ДЕЛАЕМ?

Из-за долгой, холодной и малоснежной зимы в Средней Сибири на реке Лена зимой образуется чрезвычайно толстый и прочный лед. Это приводит к тому, что практически каждую весну во время ледохода вниз по течению Лены передвигаются серии заторов льда – скопления льдин в русле реки, вызывающего стеснение живого сечения потока.

На образование заторов льда на Лене влияют как краткосрочные, так и долгосрочные факторы: прочность ледяного покрова перед вскрытием, интенсивность снеготаяния и скорость роста расходов воды в период весеннего половодья; большинство из них базируются на общих для всей территории метеосиноптических процессах, которые заключаются в преобладающих типах зимней атмосферной циркуляции (за январь–февраль) и в отклонении мартовских температур воздуха от среднепогодных значений. Всё это делает возможным заблаговременный прогноз вероятности возникновения затора.

Высокий горизонт планирования в этой задаче позволит службам МЧС заранее оптимально распределить ресурсы, чтобы усилить подготовку к противозаторным мероприятиям, предупредить население о надвигающейся опасности и снизить экономический и экологический ущерб.

# ФОРМУЛИРОВКА ЗАДАЧИ

Участникам предлагается решить задачу бинарной классификации: для каждого гидропоста с помощью ежегодных данных до 15 февраля включительно необходимо установить, в какие из дней в период **с 21 апреля по 3 июня происходит затор льда**. Модель должна выдавать не сам лейбл класса («затор» и «отсутствие затора»), а вероятность класса «затор». Под вероятностью мы подразумеваем степень уверенности модели в классе «затор», нормированную до интервала **[0, 1]**.

Прогноз необходим для гидропостов, расположенных рядом с населенными пунктами: город Киренск, посёлок городского типа Витим, посёлок городского типа Пеледуй, село Крестовский Лесоучасток, город Ленск, город Олёкминск, город Покровск, город Якутск, село Батамай, посёлок городского типа Сангар.

**Модель будет оцениваться на тестовой выборке, охватывающей 44-дневный период, во время которого возможны заторы.**

## МЕТРИКА

Усредненное по населенным пунктам (микровзвешенное) значение F1-score.

# ТЕСТОВАЯ ВЫБОРКА

Годы, попавшие в тестовую выборку:

1989

1993

1997

2001

2003

2004

2005

2012

2013

# ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ

## **Краткая аннотация проекта**

описание проекта в 2-4 предложениях;

## **ML-модель, обученная на тренировочной выборке**

Ссылка на я.диск/dropbox/google.drive с ML-моделью, обученной на тренировочной выборке;

## **Код, использованный при подготовке модели**

Ссылка на код, использованный при подготовке модели (ссылка на любой открытый репозиторий с кодом, с помощью которого строилась модель, а также кодом, запускающим модель);

## **Презентация результатов решения**

Ссылка на презентацию для очной защиты

# КРИТЕРИИ ОЦЕНКИ

1. **60 баллов** распределяется в соответствии с **результатом, который ML-модель показала на соревновании**. Будет учитываться не место в турнирной таблице напрямую, а логарифмированное значение ошибки на тестовой выборке относительно величины ошибок других участников. По результатам соревнования на очную защиту перед жюри допускается не более 10 команд.

**40 баллов** распределяется по результатам оценки кода, выложенного в открытый репозиторий, а также выступления на очной защите в соответствии со следующими критериями:

2. обоснованность выбранных методов решения задачи с точки зрения гидрологии и применимости итоговой модели в работе МЧС (**20 баллов**);
3. оригинальность подхода, в частности, использование дополнительных данных из открытых источников для повышения качества модели (**10 баллов**);
4. качество кода: его чистота и понятность (**5 баллов**);
5. качество очной защиты: последовательность и понятность изложения (**5 баллов**).

# ИНСТРУКЦИЯ ПО ПЛАТФОРМЕ

- Перейти по ссылке <https://apply.emergencydatahack.ru/competition> или выбрать раздел “DS Чемпионат”
- Выбрать трек, в котором вы участвуете
- Нажать на кнопку “Скачать датасет”. Пароль для доступа к данным: **dszWwT8x**

**DS ЧЕМПИОНАТ**

Треки:

- Трек 1
- Трек 2
- Трек 3

**ТРЕК #1 ПРЕДСКАЗАНИЕ ЗАТОРОВ НА РЕКЕ ЛЕНА В ПАВОДКООПАСНЫЙ ПЕРИОД**

**ОПИСАНИЕ ЗАДАЧИ**

Построить предиктивную модель, которая позволит за 60 дней рассчитать вероятность возникновения заторов льда на каждые сутки в период ледохода на нескольких участках реки Лена. [Подробнее описание.](#)

[Скачать датасет](#)

**РЕШЕНИЯ ДЛЯ ТЕСТОВОГО ДАТАСЕТА**

Загрузи решения для тестового датасета. Пример, как должен выглядеть файл для загрузки, коды ошибок и метрика: [подробнее](#). У вас есть 35 попыток на время всего соревнования, в зачет идет лучшая. Если возвращается ошибка - попытка не сгорает.

Попыток осталось: 34/35

История попыток:

Код	Дата	Скор	*.CSV	*.ZIP
-2	23.05.21 23:55:11	0	<a href="#">+</a>	<a href="#">+</a>

Codebook доступен  
по [ссылке](#).



# ИНСТРУКЦИЯ ПО ПЛАТФОРМЕ

- Прочитать все, что написано на странице
- Для отправки решения выбрать CSV файл с ответами и архив с кодом, нажать кнопку “отправить” - необходимо отправлять архив и CSV каждый раз, чтобы жюри смогли провалидировать способ, которым был получен скор
- Для просмотра текущего лидерборда - нажмите кнопку “открыть LeaderBoard”

The screenshot shows the EMERGENCY DATANACK platform interface. On the left is a purple sidebar with navigation links: RU, EN, EMERGENCY DATANACK ОНЛАЙН-КАКАТОН, Панель управления, Новости, Анкета, Команда, Решение, DS Чемпионат, Подтверждение, and Выйти. The main content area has a header 'ОПИСАНИЕ ЗАДАЧИ' with a description in Russian about building a predictive model for ice jams on the Lena river. Below this is a 'Скачать датасет' button. The 'РЕШЕНИЯ ДЛЯ ТЕСТОВОГО ДАТАСЕТА' section shows a table of solutions and upload buttons for \*.CSV and \*.ZIP files. Red arrows point to these upload buttons, which are highlighted with a red box. At the bottom, the 'Отправить' button is highlighted with a red box, and the 'Открыть LeaderBoard' button at the very bottom is also highlighted with a red box.

ОПИСАНИЕ ЗАДАЧИ

Построить предиктивную модель, которая позволит за 60 дней рассчитать вероятность возникновения заторов льда на каждые сутки в период ледохода на нескольких участках реки Лена.  
[Подробнее описание.](#)

Скачать датасет

РЕШЕНИЯ ДЛЯ ТЕСТОВОГО ДАТАСЕТА

Загрузи решения для тестового датасета. Пример, как должен выглядеть файл для загрузки, коды ошибок и метрика: [подробнее](#). У вас есть 35 попыток на время всего соревнования, в зачет идет лучшая. Если возвращается ошибка - попытка не сгорает.

Попыток осталось: 34/35

История попыток:

Код	Дата	Скор	*.CSV	*.ZIP
-2	23.05.21 23:55:11	0		
0	23.05.21 23:55:27	0.006783842121492523		

Загрузить (\*.CSV, лимит 3mb)

Загрузить (\*.ZIP, лимит 3mb)

Отправить

Открыть LeaderBoard

# ЗАГРУЗКА РЕШЕНИЙ НА ПЛАТФОРМУ

**30 мая до 11:00**

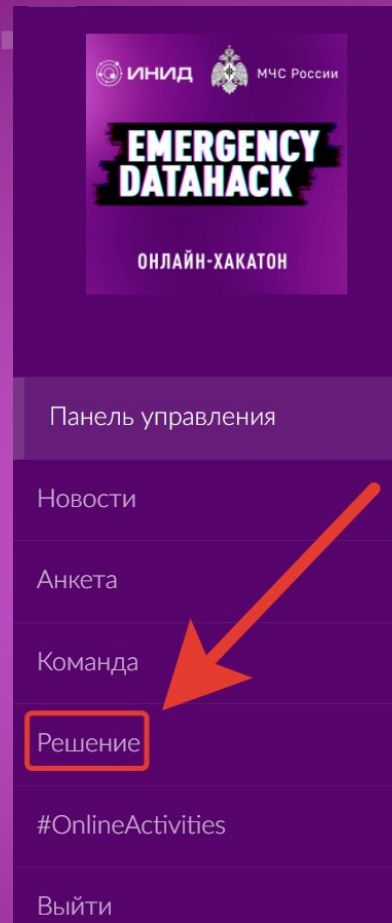
загрузить результаты на платформу во вкладку “Решение”

На защиту проектов будут допущены не более 10 команд согласно результатам лидерборда

**Время защиты проекта:**

5 минут - презентация проекта

3 минуты - ответы на вопросы



# ПРОГРАММА ХАКАТОНА

## 28 мая

**18:00** - Открытие хакатона

**18:20** - Выдача презентаций с задачами и данных

**19:30** - Q&A сессия с экспертами по задачам

## 29 мая

**11:00** - Чек-поинт #1

**15:00** - Мастер-класс от компании Tele2 “Геоаналитика на транспорте”

**18:00** - Чек-поинт #2

## 30 мая

**11:00** - Дедлайн загрузки решений

**11:00 - 15:00** - Проверка кода

**15:00 - 17:00** - Защита проектов

**17:00 - 18:00** - Подведение итогов

**18:00** - Объявление победителей, завершение хакатона

ОНЛАЙН-ХАКАТОН

# EMERGENCY DATAHACK

УСПЕХОВ



МЧС  
РОССИИ



ИНИД

