



## **Задача «Классификация агрокультур на основе изменения показателя вегетационных индексов во временной последовательности»**

### **Введение**

Данные дистанционного зондирования Земли из космоса позволяют решать большое количество производственных задач. Например, определять культуры растений на основе анализа временных рядов значений вегетационных индексов (NDVI), получаемых на протяжении вегетационного периода (периода роста и развития растений).

NDVI – это индекс, определяемый по поглощению и отражению растениями лучей красной и ближней инфракрасной зоны спектра при анализе изображений космоснимков. По значению этого индекса можно судить о развитии зеленой массы растений во время вегетации. Чем больше зеленой биомассы на полях, тем выше значение NDVI.

Общая посевная площадь в России в 2022 году составила более 81 млн га. Для грамотного планирования сельскохозяйственного производства в масштабах страны, применения технологий точного земледелия, проведения мероприятий по защите растений, оценки урожайности критически важным является распознавание культур растений, произрастающих на каждом конкретном поле.

### **Условие задачи**

Перед участниками чемпионата стоит задача классификации агрокультур на основе изменения показателя вегетационных индексов во временной последовательности. Созданная за время чемпионата модель поможет оптимизировать рабочие процессы в сфере сельского хозяйства, обеспечить органы власти актуальными и достоверными сведениями, увеличить социальный, инвестиционный и производственный потенциал отрасли, а также повысить достоверность и удобство использования информации о сельскохозяйственных объектах.

## Описание входных значений

- **train.csv** — файл, содержащий данные о поле и временные ряды
- **test.csv** — файл, содержащий данные для предсказания
- **sample\_solution.csv** — пример файла для отправки

Пояснения к столбцам

- **id** – идентификатор объекта (поля)
- **area** – площадь поля в га
- **nd mean** YYYY-MM-DD – медианное значение вегетационного индекса NDVI для данного поля на указанную дату
- **geo** – координаты границ поля
- **crop** – столбец для предсказания, содержит тип произрастающей культуры по данным сельхозтоваропроизводителей

## Дополнительно расшифруем типы культур:

- 0 - подсолнечник
- 1 - картофель
- 2 - пшеница озимая
- 3 - гречиха
- 4 - кукуруза
- 5 - пшеница яровая
- 6 - сахарная свекла

## На что стоит обратить внимание

Часть значений внутри временных рядов пропущены и заполнены нулями. Для большей точности рекомендуем сделать фильтр который будет заменять пропуски на усредненное значение своих соседей.

## Метрика

В качестве метрики выступает Recall по столбцу **crop**.

Recall считается как:

$$recall = \frac{TP}{TP + FN}$$

TP (True Positive) — количество верно угаданных значений одного класса  
FN (False Negative) — количество не правильно угаданных значений класса

### **Правила чемпионата:**

1. С момента открытия датасета до момента завершения приема решений репозиторий участника, в котором он ведет разработку по задаче текущего чемпионата, должен оставаться закрытым.
2. Участник обязан открыть доступ к репозиторию на чтение по ссылке (которая была прикреплена в ЛК в поле «Ссылка на код (гитхаб)») не позднее чем в течение 12 часов с момента окончания дедлайна отправки решений на чемпионате.
3. Согласно п. 5.8 Положения в процессе верификации решений организаторы и технические эксперты, проверяющие решения участников, вправе назначить интервью с участниками чемпионата. Участник получит приглашение и ссылку на интервью не позднее чем за 12 часов до публикации итогового лидерборда. Пропуск интервью участником является поводом для дисквалификации.
4. Организаторы вправе исключить участника из призовых позиций лидерборда за непредоставление одного из артефактов решения задачи: тизера, скринкаста, презентации, ссылки на репозиторий.
5. Организаторы вправе дисквалифицировать участника в случае выявления плагиата кода или несоблюдения Положения проекта.
6. Участник, получивший 2 дисквалификации за сезон проекта, попадает в чёрный список с дальнейшим отстранением от участия в чемпионатах до конца сезона.