

International  
professional  
development platform

КОНКУРСНОЕ ЗАДАНИЕ  
ДЛЯ ЧЕМПИОНАТА  
BRICS FutureSkills & Tech Challenge 2024  
ПО КОМПЕТЕНЦИИ  
«МАШИННОЕ ОБУЧЕНИЕ И  
БОЛЬШИЕ ДАННЫЕ»



## С2. Модуль С. Построение, обучение и оптимизация модели

### СОДЕРЖАНИЕ

Модуль С данного Конкурсного задания состоит из следующей документации / файлов:

1. Набор данных, полученный в Модуле В.
2. Приложение 2 к Модулю С.
3. Инструкция к Модулю С.

### ИНСТРУКЦИЯ УЧАСТНИКУ

В рамках данного модуля Вам предстоит:

- Предложить подход для определения высоты леса по спутниковым снимкам;
- **Фитомасса верхнего полога древостоя** — это биомасса деревьев, находящихся в верхнем слое лесного полога. Она включает в себя листья, ветви, стволы и другие надземные части деревьев, которые составляют основной объем растительности в лесу. Измерение и оценка фитомассы является важным аспектом для изучения углеродного цикла, экосистемных услуг, лесных ресурсов и мониторинга состояния лесов. Фитомасса верхнего полога древостоя, рассчитанная по разносезонным снимкам ДЗЗ, дает важную информацию о состоянии лесов, помогает оценить углеродные запасы и является полезным инструментом для управления лесными экосистемами и их мониторинга. Использование данных, полученных в разные сезоны, позволяет более точно оценивать сезонные колебания биомассы и получать более детализированную информацию о лесных массивах. Вам необходимо предложить подход для оценки фитомассы верхнего полога древостоя по разносезонным снимкам ДЗЗ.

Дополнительные материалы:

1. <https://cyberleninka.ru/article/n/otsenka-zapasov-fitomassy-lesnyh-porod-po-sputnikovym-izobrazheniyam-vysokogo-prostranstvennogo-razresheniya-na-primere-lesov-hanty/viewer> – оценка фитомассы древостоя по разносезонным снимкам ДЗЗ.

### ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

#### 1. Классификация типов лесов - выделение основных ареалов леса

Необходимо выделить три основных ареала леса:

- Ареал1: не покрыт лесами,
  - Ареал2: преимущественно хвойный лес,
  - Ареал3: преимущественно лиственный лес.
1. Сформируйте целевую переменную. Обоснуйте свое решение.
  2. Подготовьте выборки для обучения, валидации, тестирования. Аргументируйте разделение.
  3. Предложите и обучите модели для не менее 3-х алгоритмов (включая нейронные сети) для многоклассовой классификации (Ареалы 1,2,3).
  4. Для оценки качества моделей используйте следующие метрики: PRECISION, RECALL, F1. Сделайте выводы о точности каждой модели.
  5. Постройте ROC-AUC кривые, сделайте выводы.
  6. По выбранным метрикам определите лучший алгоритм.
  7. Проведите оценку точности на валидационной и тестовой выборках.
  8. Проведите подбор и настройку гиперпараметров лучшей модели.



9. Проведите финальную оценку точности разработанной модели.
10. Реализуйте алгоритм вывода уверенности модели (вероятность правильности определения ареала).
11. Визуализируйте результат разделения на ареалы на карте участка для выбранного алгоритма. Визуализация должна быть снабжена легендой, отражающей информацию по каждому ареалу.

## 2. Оценка фитомассы верхнего полога древостоя по разнотемпературным снимкам ДЗЗ, выявление аномалии или деградации лесов

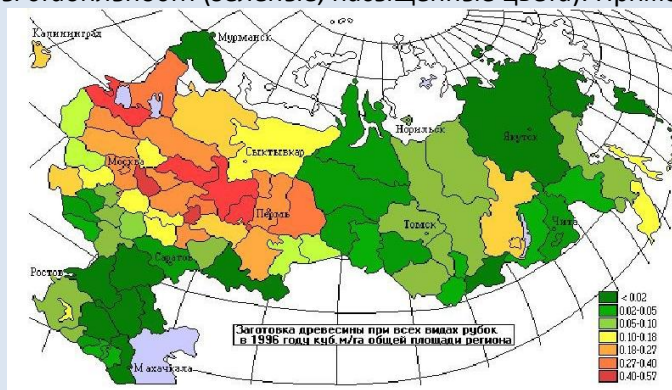
Предложите алгоритм для выявления аномалии или деградации лесов на основе оценки фитомассы верхнего полога древостоя по разнотемпературным снимкам ДЗЗ (используя спектральный индекс **NDVI**).

1. Для каждого пиксела (квадрат 10м x 10м) на карте определить индекс NDVI (Приложение 2, представлена краткая справка по индексу NDVI) в динамике - для каждого сезона и каждого года.
2. Визуализируйте значение индекса NDVI в динамике в границах указанного участка леса.

### Определение аномальных зон (зон деградации леса) на основе динамики индекса NDVI.

#### Пример определения зон деградации:

- **До деградации:** Лес густой, с высоким уровнем фотосинтеза, и NDVI имеет высокие значения (например, от 0.6 до 0.9).
  - **Во время деградации:** В случае пожара или вырубki значительная часть растительности исчезает, и NDVI падает (например, до 0.2–0.3).
  - **После деградации:** Если лес не восстанавливается, а территория остается в деградированном состоянии (например, заброшенные поля, кустарники), NDVI может оставаться на низком уровне (около 0.1–0.3) и не демонстрировать больших сезонных колебаний.
3. Рассчитайте площадь аномалий для каждого года заданного периода (нет растительности), деградации леса (индекс NDVI существенно снижается в динамике).
  4. Визуализируйте на карте (в динамике, для каждого года) аномальные зоны (зоны деградации) как зоны сведения леса (цвет в желтых тонах, полностью был и отсутствует лес – красные)) и зоны стабильности (зеленые, насыщенные цвета). Пример:

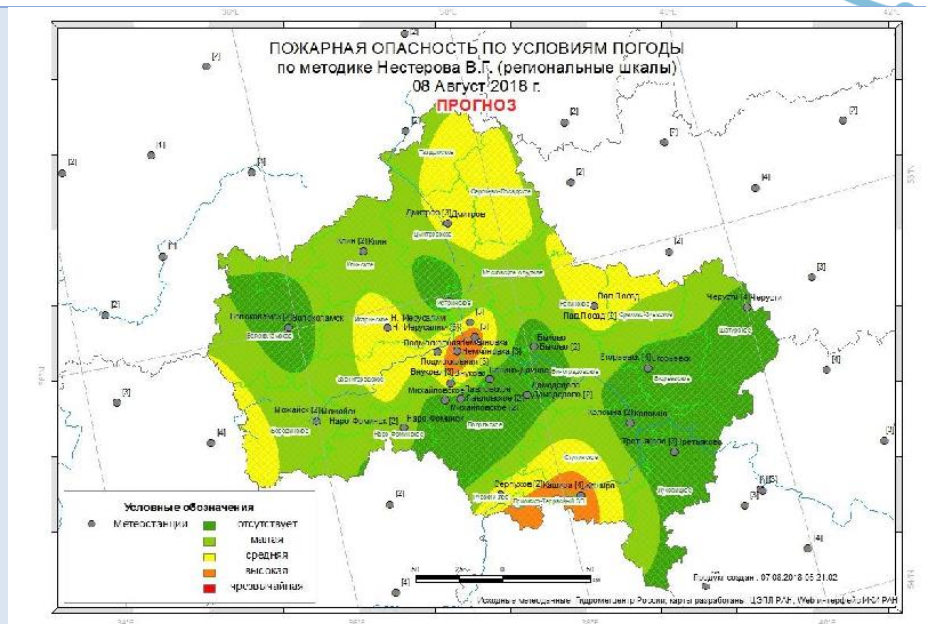


5. Сделайте выводы о деградации леса на заданном в задании участке.

## 3. Определение депрессивных зон леса в летние периоды – пожароопасные зоны

Предложите алгоритм для выявления депрессивных зон леса в **летние периоды** – пожароопасные участки (используя спектральный индекс **NDMI**).

1. Для каждого пиксела (квадрат 10м x 10м) на карте определить индекс NDMI (как комбинацию каналов NIR и SWIR1, Приложение 2) в динамике – в летние периоды каждого года.
2. Определите зоны, где растения испытывают дефицит влаги – зоны повышенной пожароопасности. Рассчитайте площадь этих зон в динамике, за каждый год.
3. Визуализируйте на карте пожароопасные зоны. Пример:



4. Сделайте общие выводы, отражающие качество работы алгоритма.
5. Определите связь между индексами NDVI и NDMI в зонах пожароопасности. Сделайте выводы. Обоснуйте выбранный метод определения связи.

#### 4. Отчет

1. Предоставлен отчет о проделанной работе. Внимание: оценка сессии будет проводиться на основании отчета. Отчет предлагается писать в Jupyter Notebook или аналогичной среде, где участник может последовательно представить, как описание проделанной работы, так и часть программного кода, и результат работы программы.
2. Отчет должен быть предоставлен в папке C2[X]\_MC, где [X] – номер участника, который должен быть размещен на рабочем столе. Папка должна содержать все результаты выполнения модуля, а также все необходимые файлы для запуска и проверки участков кода. Отчет также должен быть в формате \*.pdf или \*.html

## ПРИЛОЖЕНИЕ 1 к Модулю С

**Нормализованный вегетационный индекс NDVI (NDVI)** – это стандартизированный индекс, показывающий наличие и состояние растительности (относительную биомассу). Этот индекс использует контраст характеристик двух каналов из набора мультиспектральных растровых данных – поглощения пигментом хлорофилла в красном канале и высокой отражательной способности растительности в инфракрасном канале (NIR). Один из самых распространенных и используемых индексов для решения задач, использующих количественные оценки растительного покрова.

Индекс рассчитывается по следующей формуле:

$$NDVI = ((NIR - Red)/(NIR + Red))$$

NIR – значения пикселей из ближнего инфракрасного канала.

Red – значения пикселей из красного канала.

Этот индекс принимает значения в диапазоне -1.0 и 1.0. Для растительности значение индекса меняется от 0 до 1.0. Примеры значений спектральных каналов и индекса NDVI приведены в таблице.



Тип объекта	Отражение в красной области спектра	Отражение в инфракрасной области спектра	Значение NDVI
Густая растительность	0.1	0.5	0.7
Разряженная растительность	0.1	0.3	0.5
Открытая почва	0.25	0.3	0.025
Облака	0.25	0.25	0
Снег и лед	0.375	0.35	-0.05
Вода	0.02	0.01	-0.25
Искусственные материалы (бетон, асфальт)	0.3	0.1	-0.5

- Значения NDVI, близкие к 1, указывают на плотную, здоровую растительность.
- Значения около 0 свидетельствуют о слабой или редкой растительности (или грунте).
- Отрицательные значения могут указывать на наличие воды, снега или облаков.

### Интерпретация NDVI в задаче оценки фитомассы верхнего полога древостоя

Для оценки **фитомассы верхнего полога древостоя** по разносезонным снимкам ДЗЗ индекс NDVI помогает оценить количество и состояние растительности в зависимости от времени года.

#### Связь между NDVI и фитомассой:

- **Высокие значения NDVI** (близкие к 1) указывают на плотную и здоровую растительность, что коррелирует с высокой фитомассой. Например, летом, когда деревья находятся в фазе активного роста, их листья более плотные, а верхний полог покрыт густой листвой, что увеличивает отражение в инфракрасном диапазоне и дает высокие значения NDVI.
- **Низкие значения NDVI** (близкие к 0) указывают на малую или слабую растительность, что связано с низкой фитомассой. Например, осенью или зимой, когда листья опадают, или при деградации лесов, значения NDVI снижаются, отражая уменьшение фитомассы.

#### Разносезонные снимки:

- **Лето:** В это время значения NDVI будут самыми высокими, так как деревья в верхнем пологе имеют максимальную листовую массу.
- **Весна и осень:** В этих сезонах значения NDVI будут колебаться. Весной, в начале вегетационного периода, они будут постепенно увеличиваться, а осенью, по мере опадания листвы, уменьшаться.
- **Зима:** Значения NDVI будут низкими, так как большинство деревьев в умеренных и холодных широтах теряют листву. Если деревья лиственные, значения NDVI могут быть около нуля, что отражает низкую фитомассу верхнего полога.

### Мониторинг изменений и динамики фитомассы:

Используя разносезонные данные NDVI, можно:



- **Отслеживать сезонные изменения фитомассы.** Сравнение NDVI в разные сезоны позволяет оценивать, как меняется плотность и состояние растительности в течение года. Это полезно для понимания сезонной динамики лесных экосистем.
- **Выявлять аномалии или деградацию лесов.** Если NDVI аномально снижается в периоды, когда ожидается активный рост растений (например, весной или летом), это может указывать на проблемы в лесных экосистемах, такие как заболевания, вырубка или пожары.
- **Анализировать долгосрочные тренды.** Сравнение NDVI за несколько лет позволяет выявлять тренды в изменении фитомассы, такие как последствия изменения климата или антропогенного воздействия на лесные экосистемы.

---

Индекс **NDMI (Normalized Difference Moisture Index)** – это **нормализованный разностный индекс влажности**. Он используется для определения влажности почвы и растительности.

NDMI рассчитывается по формуле:

$$\text{NDMI} = (\text{NIR} - \text{SWIR}) / (\text{NIR} + \text{SWIR}),$$

где NIR – ближний инфракрасный диапазон, а SWIR – коротковолновый инфракрасный диапазон.

Этот индекс позволяет оценить количество воды в почве и растительности, так как вода поглощает большую часть энергии в ближнем инфракрасном диапазоне, но почти не поглощает энергию в коротковолновом инфракрасном диапазоне.

Таким образом, чем выше значение NDMI, тем больше влаги содержится в почве или растительности. Индекс NDMI может использоваться для оценки заболоченности леса. Заболоченные участки обычно имеют более высокую влажность почвы и растительности, что отражается в высоких значениях NDMI (обычно выше 0,6).

Пожароопасные зоны – NDMI в диапазоне 0 – 0.2. Зоны дефицита влаги – NDMI в диапазоне 0 – 0.4.