



International
professional
development platform

КОНКУРСНОЕ ЗАДАНИЕ

ДЛЯ ЧЕМПИОНАТА BRICS FutureSkills & Tech Challenge 2024

ПО КОМПЕТЕНЦИИ «МАШИННОЕ ОБУЧЕНИЕ И БОЛЬШИЕ ДАННЫЕ»



Конкурсное задание включает в себя следующие разделы:

1.	ФОРМА УЧАСТИЯ В КОНКУРСЕ:	3
2.	ОБЩЕЕ ВРЕМЯ НА ВЫПОЛНЕНИЕ ЗАДАНИЯ:	3
3.	ЗАДАНИЕ ДЛЯ КОНКУРСА	3
4.	МОДУЛИ ЗАДАНИЯ И НЕОБХОДИМОЕ ВРЕМЯ	3
5.	КРИТЕРИИ ОЦЕНКИ	7



1. ФОРМА УЧАСТИЯ В КОНКУРСЕ:

Индивидуальный конкурс

Очная форма

2. ОБЩЕЕ ВРЕМЯ НА ВЫПОЛНЕНИЕ ЗАДАНИЯ: 15 ч.

3. ЗАДАНИЕ ДЛЯ КОНКУРСА

В рамках конкурсного задания на основе большого набора данных, конкурсантам необходимо будет выполнить загрузку, анализ и предварительную обработку больших данных, представленных в «ненормализованном» виде, затем оценить влияние признаков на целевую переменную и выбрать лучшие, с помощью которых построить математическую модель машинного обучения и выполнить оценку точности и качества ее работу на тестовой нормализованной выборке. В качестве завершающего этапа конкурсного задания конкурсанту необходимо разработать графический интерфейс и документацию по созданному решению.

Конкурсное задание является секретным и доступно для ознакомления экспертам в день С-2, участникам – в дни С1-С3 за 15 минут до начала модуля.

4. МОДУЛИ ЗАДАНИЯ И НЕОБХОДИМОЕ ВРЕМЯ

Наименование модуля		Соревновательный день (С1, С2, С3)	Время на задание
A	Парсинг и предобработка данных	С1	3 часа
B	Разведочный анализ данных.	С1	3 часа
C	Построение, обучение и оптимизация модели	С2	3 часа
D	Разработка программного продукта. Разработка средств интеграции и поддержки готового решения.	С2	3 часа
E	Оценка точности модели. Презентация результатов работы	С3	3 часа



Модуль А: Парсинг и предобработка данных.

Данные представлены в одном из форматов: txt, pdf, doc, docx, xls, xlsx, xml, csv, shape-файлы, геоданные.

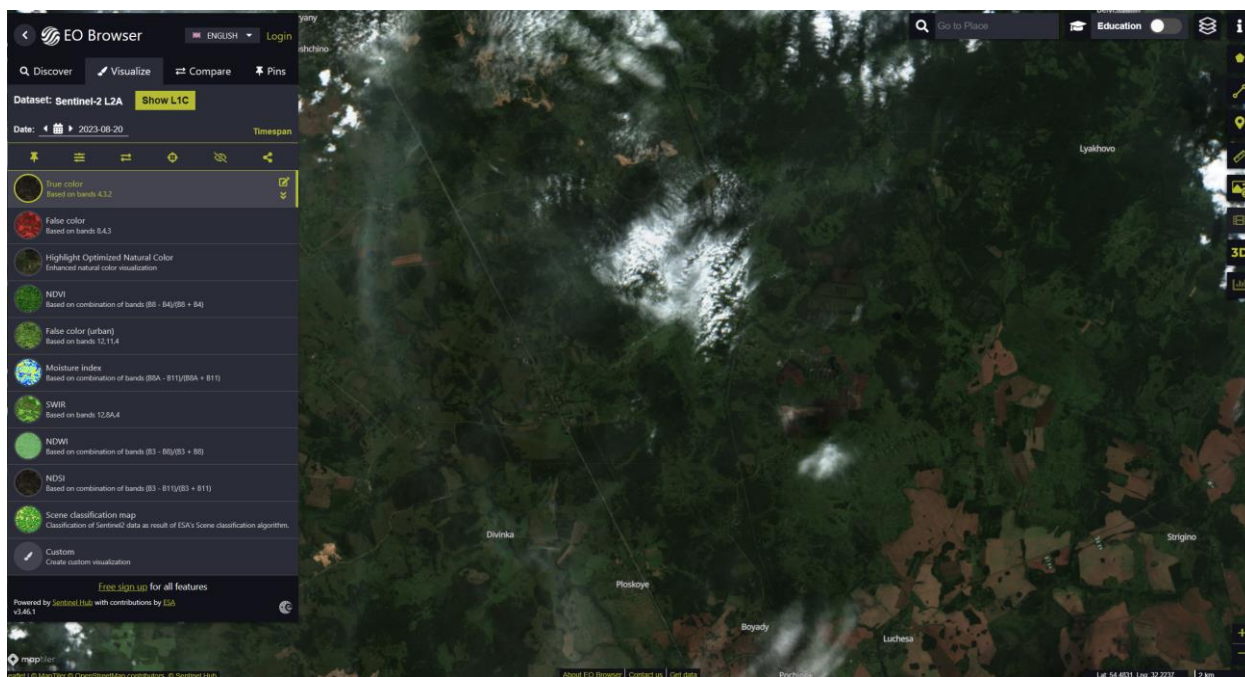
Необходимо преобразовать исходный набор данных к виду, пригодному для загрузки.

Примерные этапы задания в модуле А:

- 1) Загрузка данных и создание дополнительного набора данных;
- 2) Предобработка данных. Анализ и корректировка изображений;
- 3) Очистка данных от служебной информации.

Доп. информация по Модулю А

ЕО Browser обеспечивает открытый доступ к коллекциям полноразмерных спутниковых снимков, включая снимки в исходном разрешении Landsat 5, 7, 8 и Sentinel 2, 3 в различных комбинациях каналов. Доступен по адресу: <https://apps.sentinel-hub.com/eo-browser>.



В рамках конкурсного задания используются только снимки Sentinel-2. Он должен быть выбран на первой странице ЕО Browser и, после нажатия кнопки «Search» появится список доступных снимков за текущую дату. После выбора снимка появляется список доступных для просмотра слоев



(комбинации различных спектральных каналов). При выполнении задания используются слои TrueColor и NDVI.

По умолчанию EO Browser осуществляет поиск снимков за последнюю доступную дату. Для изменения временного диапазона поиска доступны две опции:

- Single Date позволяет выбрать конкретную дату для поиска снимков.
- Timespan позволяет выбрать диапазон дат для поиска снимков.

В правой части экрана находится список инструментов, один из которых позволяет сохранять изображение на экране браузера в файлы формата JPG, PNG. Этот инструмент можно использовать для получения фрагментов спутниковых снимков.

Модуль В: Разведочный анализ данных.

В рамках данного модуля Вам предстоит расширить набор данных, а также сегментировать спутниковые снимки и найти на них поля сельскохозяйственного назначения.

Примерные этапы задания в модуле В:

- 1) Расширение набора данных с помощью методов аугментации;
- 2) Сегментация изображений «без учителя»;
- 3) Формирование итогового набора расширенных и предобработанных данных.

Модуль С: Построение, обучение и оптимизация модели.

В рамках данного модуля Вам предстоит построить, обучить и оптимизировать модель сегментации снимков, а также оценить ее качество.

Примерные этапы задания в модуле С:

- 1) Сегментация изображений «с учителем»;
- 2) Решение задач машинного обучения: классификация, кластеризация, прогнозирование;
- 3) Построение и оптимизация моделей.

Модуль D: Разработка программного продукта. Разработка средств интеграции и поддержки готового решения.



В рамках данного модуля Вам предстоит разработать API, а также программный продукт в виде приложения для заказчика.

Примерные этапы задания в модуле D:

- 1) Разработка API;
- 2) Разработка графического интерфейса пользователя (приложение);
- 3) Подготовка руководства пользователя.

Модуль E: Оценка точности модели. Презентация результатов работы

В рамках данного модуля Вам предстоит оценить точность построенной модели на секретной выборке, оценить результат. Также необходимо будет подготовить продукт для передачи заказчику – оформить инструкцию для пользователя, презентацию и сделать доклад.

Примерные этапы задания в модуле E:

- 1) Оценка точности разработанной модели на «секретной» выборке;
- 2) Разработка инструкции для пользователя (по запуску и эксплуатации приложения);
- 3) Подготовка презентации, доклада, итогового отчета.



5. КРИТЕРИИ ОЦЕНКИ

№ п/п	Раздел Матрицы навыков	Важность раздела в %
1	Организация и управление работой, коммуникация	6
2	Подготовка данных для анализа	18
3	Анализ и обработка данных	22
4	Построение моделей машинного обучения	15
5	Искусственный интеллект (ИИ) и машинное обучение	9
6	Разработка прикладных решений и документирование	30

Критерий оценки (Модули конкурсного задания)							Итого баллов за раздел Матрицы навыков
Разделы Матрицы навыков		A	B	C	D	E	
	1	2	2	0	0	2	6
	2	12	2	0	2	2	18
	3	2	18	0	0	2	22
	4	0	0	15	0	0	15
	5	0	0	9	0	0	9
	6	0	0	0	18	12	30
Итого баллов по Критерию оценки		16	22	24	20	18	100