

International
professional
development platform

КОНКУРСНОЕ ЗАДАНИЕ

ДЛЯ ЧЕМПИОНАТА
BRICS FutureSkills & Tech Challenge 2024

ПО КОМПЕТЕНЦИИ
«МАШИННОЕ ОБУЧЕНИЕ И
БОЛЬШИЕ ДАННЫЕ»



С1. Модуль В. Разведочный анализ данных

СОДЕРЖАНИЕ

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

1. Инструкция к Модулю В.

ИНСТРУКЦИЯ УЧАСТНИКУ

В рамках данного модуля Вам предстоит провести разведочный анализ данных, предложить подход для определения породного состава леса, оценить основные характеристики заданного участка леса, а также классифицировать лесные угодья на заданные ареалы.

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ

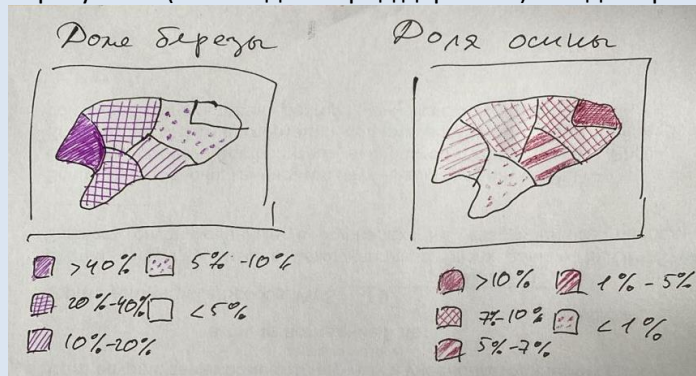
1. Анализ и визуализация данных. Определение породного состава леса

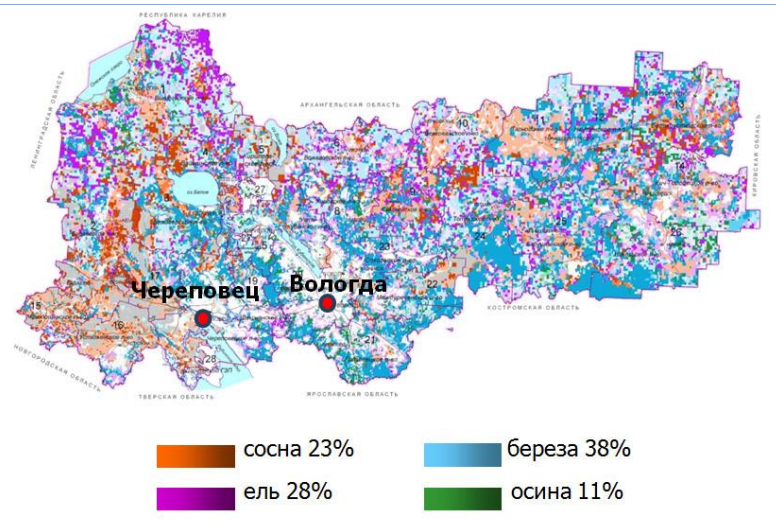
Предложите подход для определения породного состава леса (сосна, ель, береза, осина, лиственница).

1. Для каждой породы дерева (а также для позиции 000000) вывести его распространение в границах заданного участка. Вывести результат в виде таблицы:

Сокращение	Расшифровка	% распространения
С	Сосна обыкновенная	
Е	Ель европейская	
Б	Берёза повислая	
ОС	Осина	
Л	Лиственница сибирская	
000000	Не покрытые древесно-кустарниковой растительностью	

2. Визуализировать результат (только для пород деревьев) в виде карт. Примеры:





3. Рассчитайте площадь породного состава леса (по каждой породе), выведите в таблице.
4. Для каждого ареала рассчитайте: средний возраст леса, средний запас древесины, распределение (в %) зон – хвойные, лиственные, зоны без леса.
5. Выделите TOP-10 участков, с наибольшим запасом леса, выведите координаты этих участков.
6. Визуализируйте эти участки – выделите их границы на карте, закрасьте цветом.

2. Определение ареалов с зонами вырубki леса

1. Определены ареалы с зонами проведения рубок леса за промежуток времени с 2021 по 2024 годы, оценить их площадь.
2. Определите зоны с наибольшей площадью вырубki, отобразите границы этих зон на карте.

3. Сегментация изображений - кластеризации ареалов леса по типам, возрасту, запасам древесины

1. Для кластеризации ареалов леса по типам, возрасту, запасам древесины предложить не менее 3-х вариантов алгоритмов. Обосновать выбор.
2. Выполнить кластеризацию областей – не менее 5 кластеров каждым из выбранных алгоритмов. Сделать предположения о характере каждого кластера.
3. Визуализируйте результат кластеризации на всей карте участка для каждого алгоритма.
4. Предложите метрику для оценки качества кластеризации, рассчитайте ее значение для каждого из трех алгоритмов.
5. Сделайте вывод – о качестве работы каждого из 3-х алгоритмов.
6. Добавьте номер кластера как дополнительный атрибут в исходный массив атрибутов для каждого ареала леса.

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 C1[X]_MB, 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