Задание для хакатона

Цель: Классификация сельскохозяйственных культур, произрастающих на территории Хабаровского края на основе временных рядов значений оптических каналов и вегетационного индекса NDVI.

Для построения модели классификации необходимо использовать сглаженные временные ряды вегетационных индексов NDVI и отдельных спутниковых каналов, рассчитанных с помощью оптических снимков, полученные со спутников Sentinel-2A/B.

Датасет состоит из точек, для которых известна информация о севообороте. Он разделен на обучающую, открытую часть и закрытую часть тестовой выборки. В нем содержатся данные 10500 точек сельскохозяйственных полей Хабаровского края, с указанием произрастающей культуры для обучающей части. Наименование культур приведены в столбце **Culture**. Список культур представлен в таблице.

*Таблица 1. Сельскохозяйственные культуры, включенные в датасет.*

|  |  |
| --- | --- |
| Номер | Название культуры |
| 1 | залежь |
| 2 | соя |
| 3 | многолетние травы |
| 4 | зерновые |
| 5 | кукуруза |
| 6 | овощи |

ВАЖНО! Не меняйте названия культур, именно эти значения будут использоваться для тестирования модели!

Для классификации можно использовать любую модель (разработанную самостоятельно или реализованную в библиотеке, например, scikit-learn (www.scikit-learn.org) языка Python), например, комбинацию деревьев решений Random Forest. По желанию команды в датасет могут быть добавлены дополнительные критерии, рассчитанные на основе имеющихся данных, например, статистические характеристики временных рядов. Некоторый критерии (столбцы) могут быть исключены из модели.

Построенная модель должна поточечно классифицировать сельскохозяйственные культуры на полях, представленных в открытой и закрытой части тестовой выборке. Предварительное тестирование будет проводится путем подсчёта процента правильно распознанных точек на открытом массиве данных. На основе предварительного тестирования формируется предварительный рейтинг команд.

Итоговое решение включает:

1. результаты классификации открытой и закрытой части тестовой выборки: файл classification\_openset.csv для открытой части, classification\_closedset.csv – для закрытой;
2. программу, разработанную на языке Python, которая использовалась для построения модели классификации;
3. файл с расширением pkl (модель классификации, сохраненная с использованием стандартного модуля pickle языка Python);
4. отчет о проделанной работе в формате презентации.

Отправить необходимо архив с указанными файлами.

При оценивании модели на закрытой части выборки будет применятся f1-метрика. Алгоритм расчета и формирования окончательного варианта обучающего массива данных должен быть реализован в файле, ссылка на него включена в презентацию. Презентация должна содержать информацию о модели, используемой для классификации, желательно с обоснование выбора, этапах работы над проектом, ССЫЛКУ на код, по которому строилась модель (можно разместить в облачном хранилище или на GitHub).