

## Применение методов машинного обучения для восстановления значений NDVI в Google Earth Engine

Научный руководитель – Паниди Евгений Александрович

*Воробьева Анна Андреевна*

*Студент (бакалавр)*

Санкт-Петербургский государственный университет, Институт наук о Земле,  
Санкт-Петербург, Россия

*E-mail: st096985@student.spbu.ru*

NDVI (Normalized Difference Vegetation Index – Нормализованный разностный вегетационный индекс) – это показатель, широко используемый для оценки состояния растительного покрова. Он рассчитывается как отношение разности между отражением в ближнем инфракрасном (NIR) и красном (RED) диапазонах к их сумме [1]. При расчете показателей для территорий большой площади или при построении временного ряда в данных могут образовываться разрывы, связанные с облачностью, тенью и другими факторами, создающими помехи при съемке. Неполнота данных может быть устранена различными способами, к примеру методами интерполяции, машинного обучения, нейронных сетей.

В данном исследовании для устранения неполноты данных применяются методы машинного обучения и используются различные модели и алгоритмы (в частности модель линейной регрессии – LinearRegression [4]), а также набор данных для обучения, включающий в себя многолетнюю выборку значений вегетационного индекса [2].

Моделирование данных происходит в Google Earth Engine (GEE) через подключение к Earth Engine Python API. Это платформа для облачных вычислений, которая позволяет анализировать большие объемы данных дистанционного зондирования [3].

Обучение модели для решения задачи производилось на нескольких наборах данных, полученных по снимкам с разных сканирующих систем (Landsat 8, Sentinel - 2, MODIS), снимки отбирались за период с марта 2013 года по март 2020 года. Значения NDVI рассчитывались для каждого набора изображений и объединялись в общую таблицу.

Далее, данные разделялись на тестовую и обучающую выборки (размер тестовой выборки составил 20% от общего количества данных) и производилось обучение модели, а также вычисление средней квадратичной ошибки (СКО) модели для проверки точности. Ошибка моделирования данных на тестовой выборке составила 0,10973 (сравнение полученного и фактического значений), СКО составила 0,01204.

### Источники и литература

- 1) Шовенгердт Р.А., Дистанционное зондирование. Модели и методы обработки изображений. Пер. с англ. Кирюшина А.В., Демьяникова А.И. - М.: Техносфера, 2013. - 589 с.
- 2) Müller A. C., Guido S. Introduction to Machine Learning with Python – Sebastopol: O'Reilly Media, Inc, 2016. – 340 с.
- 3) Google Earth Engine: <https://earthengine.google.com>

### Иллюстрации

NDVI	date
0,07330226	29.03.2013
0,06421872	20.04.2013
0,151822608	15.05.2013
0,04407241	22.05.2013
0,354450912	07.06.2013
...	...

Рис. : 1. Фрагмент набора данных, используемого для обучения модели, составлено автором

```
Предсказанное значение: [0.58892]  
Фактическое значение: 4    0.69865  
Name: NDVI, dtype: float64  
Средняя квадратичная ошибка модели: 0.01204
```

Рис. : 2. Результаты предсказания и оценки точности модели, составлено автором