

Разработка программного обеспечения на языке Python

[Обзорная панель](#)[Мои курсы](#)[Разработка ПО на языке Python](#)[Анализ данных и машинное обучение](#)[Задание 4. Предобработка данных](#)

Задание 4. Предобработка данных

Набор данных `dust` содержит данные значений удельного электрического сопротивления для экспериментальных образцов пыли. Каждая запись содержит следующие поля:

- `resistivity` – УЭС пыли;
- `temperature` – температура образца пыли;
- `humidity` – влажность образца пыли;
- `density` – плотность образца;
- `dust_capacity` – емкость пыли;
- `particle_size` – размер частиц в образце;
- `is_chelyabinsk` – образец был взят на месторождении в Челябинске;
- `conductivity` – электрическая проводимость;
- `dust_dispersiveness` – дисперсность пыли;
- `formation` – способ образования пыли.

Проведите процесс предподготовки данных для возможности дальнейшего применения методов машинного обучения.

1. Выведете на экран сводные данные о датафрейме. Имеются ли в данных пропуски? Определите эти переменные.

2. Постройте гистограммы распределений признаков. Определите численные признаки, которые можно считать дискретными, а какие непрерывными.

3. Постройте корреляционную матрицу численных признаков и определите 3 наиболее сильно линейно связанных признака с целевой переменной `resistivity`.

4. Выполните указанную предобработку для следующих признаков:

- `resistivity` – прологарифмируйте значения переменной;
- `dust_capacity` – заполните пропуски средним значением переменной;
- `particle_size` – проведите стандартизацию переменной;
- `is_chelyabinsk` – удалите переменную;
- `conductivity` – проведите One-Hot Encoding для преобразования категориального признака в численный;
- `dust_dispersiveness` – проведите Label Encoding переменной по размеру частиц;
- `formation` – бинаризируйте переменную.

5. Разделите выборку на две: `train` и `test` в пропорции 70/30. Посмотрите на размерности получившихся датафреймов.

6. Отправьте файл с кодом в формате `.ipynb` в качестве ответа на задание. Название файла должно содержать вашу фамилию, например, `Ivanov-task4.ipynb`.

[dust.xlsx](#)

5 марта 2023, 12:29

Состояние ответа

Номер попытки	Номер этой попытки - 1. (Разрешено попыток - 5)
---------------	---

Состояние ответа на задание	Ни одной попытки
-----------------------------------	------------------

Состояние оценивания	Не оценено
Последнее изменение	-
Комментарии к ответу	▶ Комментарии (0)

Добавить ответ на задание

Вы пока не предоставили ответ на задание

ПРЕДЫДУЩИЙ ЭЛЕМЕНТ КУРСА

[◀ Набор данных Удельное электрическое сопротивление](#)

Перейти на...

СЛЕДУЮЩИЙ ЭЛЕМЕНТ КУРСА

[Лекция 4. ML модель. От данных до внедрения ▶](#)

© 2010-2023 Центр обучающих систем
Сибирского федерального университета, sfu-kras.ru

Разработано на платформе moodle
Beta-version (3.9.1.5.m)

[Политика конфиденциальности](#)

[Соглашение о Персональных данных](#)

[Политика допустимого использования](#)

Контакты +7(391) 206-27-05
info-ms@sfu-kras.ru

[Скачать мобильное приложение](#)

[Инструкции по работе в системе](#)