

УДК 37.013

**Иглин Павел Викторович**

доцент кафедры промышленной  
теплоэнергетики  
Самарский государственный  
технический университет  
Самара, Россия  
p.v.iglin.samgtu@gmail.com

**Pavel V. Iglin**

Associate Professor of the Department of  
Industrial Thermal Power Engineering  
Samara State Technical University  
Samara, Russia

---

**ОБУЧЕНИЕ СТУДЕНТОВ ИСПОЛЬЗОВАНИЮ  
PYTHON ДЛЯ РЕШЕНИЯ ИНЖЕНЕРНЫХ  
ЗАДАЧ В СФЕРЕ ЭНЕРГЕТИКИ**

---

**Аннотация**

В статье рассматриваются возможности использования языка программирования Python для решения различных инженерных задач, которые возникают в ходе трудовой деятельности у студентов-теплоэнергетиков. Обучение языку Python проходит в рамках двух курсов в магистратуре. Для освоения инструментария студенты обучаются основам языка программирования, основам решения алгоритмических задач, понятию Большие данные, а также применяют полученные знания на практике. Например, решают задачи по оптимальному распределению теплоносителя в тепловой сети для обеспечения потребителей необходимой нагрузкой. Обучения проводятся с использованием ресурса Google Colab.

**Ключевые слова:**

большие данные, теплоэнергетика,  
магистратура

---

---

**TEACHING STUDENTS TO USE  
PYTHON FOR SOLVING ENERGY  
ENGINEERING PROBLEMS**

---

**Abstract**

The article discusses the possibilities of using the Python programming language to solve various engineering problems that arise in the course of work for students of thermal power engineering. Python is taught as part of two master's courses. To master the tools, students learn the basics of a programming language, the basics of solving algorithmic problems, the concept of Big Data, and also apply the acquired knowledge in practice. For example, they solve problems for the optimal distribution of the heat carrier in the heating network to provide consumers with the necessary load. Training is conducted using the Google Colab resource.

**Keywords:**

Python, big data, thermal power engineering,  
magistracy

---

Студенты-магистранты кафедры «Промышленная теплоэнергетика» Самарского государственного технического университета обучаются использованию языка Python для решения инженерных задач. Курсы были введены потому, что повсеместная цифровизация приводит к появлению большого количества данных, которые необходимо собирать, хранить и обрабатывать. Этот факт отмечают многие исследователи [1-5]. Обработка больших данных (БД) позволит своевременно находить отклонения в рассматриваемой системе, а также прогнозировать поведение системы в будущем. С помощью регрессионного анализа БД можно получить зависимости, которые лягут в основу цифровых двойников оборудования и систем.

Важно отметить, что знания, полученные магистрантами в рамках учебных курсов по использованию Python применительно к объектам энергетики являются универсальными. Слушатель курса может дальше развивать свои компетенции в

рамках профессии и разрабатывать цифровые двойники систем теплоснабжения (тепловых сетей, тепловых станций, тепловых пунктов и т.д.); может пойти в науку и применять полученные знания для решения фундаментальных задач в рамках научных школ факультета; может использовать навыки в других отраслях сектора реального сектора экономики (согласно сайтам вакансий профессия data scientist на текущий момент времени является востребованной и высокооплачиваемой).

Для обработки больших данных хорошо подходит язык программирования Python. В данном языке программирования существует специальная библиотека для обработки и анализа данных – pandas.

В качестве рабочей среды для написания программ наиболее удобной является - Google Colab. Главными особенностями которого являются возможность работы онлайн без установки ПО на ПК, а также возможность использовать бесплатные мощные графические процессоры GPU и TPU, благодаря которым можно заниматься не только базовой аналитикой данных, но и более сложными задачами. Также Google Colab позволяет работать над одним файлом сразу нескольким пользователям.

Пример (фрагмент кода) обработки массива данных из суточных ведомостей ТЭЦ представлен на рисунке 1.

В качестве примера (рисунок 1) выбран массив эксплуатационных данных паровой турбины ТЭЦ, показания записывались автоматически в течение полутора лет с шагом в один час. Массив насчитывает более десяти тысяч строк.

Обработка включает в себя следующие этапы:

1) Загрузка данных. Pandas поддерживает множество форматов файлов, наиболее распространенные – csv, xlsx, sql.

2) Первичный анализ данных. На данном этапе мы выводим таблицу данных на экран. Далее выводим информацию о количестве строк и столбцов. Это позволяет оценить размеры выборки и сделать выводы о ее репрезентативности.

Также необходимо оценить количество непустых строк, для каждого столбца. Данное действие позволит оценить насколько заполнен массив данных, а также посмотреть тип данных в каждом столбце. Столбцы можно поделить по типу данных на два вида: количественные и категориальные или текстовые данные. Для категориальных данных бывает полезно пронумеровать категории и заменить текстовые названия на номера соответствующих категорий.

Untitled0.ipynb ☆

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка [Изменения сохранены](#)

+ Код + Текст

```
[ ] # импортируем библиотеки
import pandas as pd
import numpy as np
import matplotlib as mpl
import matplotlib.pyplot as plt
```

```
[ ] # подключаем Google Drive и загружаем файл
from google.colab import drive
drive.mount('/content/drive')
data_frame = pd.read_excel('/content/drive/MyDrive/Colab Notebooks/ЗапросБ.xls')
```

Mounted at /content/drive

```
# Посмотрим количество строк и столбцов
data_frame.shape
```

(13438, 13)

Наш массив данных состоит из 13 столбцов и 13438 строк.

```
[ ] data_frame.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13438 entries, 0 to 13437
Data columns (total 13 columns):
# Column Non-Null Count Dtype
---  ---
0 Время 13438 non-null object
1 Активная мощность в ЭЧСР, МВт 13437 non-null float64
2 Температура пара перед ГПЗ, ГРАД 13429 non-null float64
3 Давление свежего пара перед расходомером, МПа 13436 non-null float64
4 Расход свежего пара, т/ч 0 non-null float64
5 Средний вакуум в конденсаторе, кПа 13438 non-null float64
6 Температура циркуляционной воды на подводе в основные пучки конденсатора, ГРАД 13430 non-null float64
7 Температура циркуляционной воды на подводе в основные пучки конденсатора, ГРАД 13432 non-null float64
8 Температура циркуляционной воды на подводе во встроенные пучки конденсатора, ГРАД 13433 non-null float64
9 Температура циркуляционной воды на подводе из встроенных пучков конденсатора, ГРАД 13434 non-null float64
10 Температура циркуляционной воды на подводе из основных пучков конденсатора, ГРАД 13435 non-null float64
11 Температура пара на выходе ЦНД слева, ГРАД 13431 non-null float64
12 Температура пара на выходе ЦНД справа, ГРАД 13431 non-null float64
dtypes: float64(12), object(1)
memory usage: 1.3+ MB
```

Рисунок 1 – Код для обработки массива БД.

3) Далее необходимо обработать пустые строки. Можно либо заполнить их каким-либо значением (средним по столбцу, минимальным, максимальным), либо удалить строки, содержащие пустые значения.

4) Для статистического анализа можно вывести стандартный набор статистик: количество значений, среднее значение, минимальное значение, максимальное значение, медианное значение, квантили. Это позволит оценить массив данных на предмет нормального распределения или смещения.

5) Дальнейшая обработка включает в себя адаптацию массива под заданные цели – удаление лишних столбцов, объединение массивов, сортировка и группировка массивов по ключевым параметрам и так далее. Можно проводить математические операции со столбцами, выделять столбцы в отдельные массивы и много другое. Обработанный массив данных можно подвергнуть регрессионному анализу или подать на вход нейронной сети.

Алгоритм обработки представлен на рисунке 2.

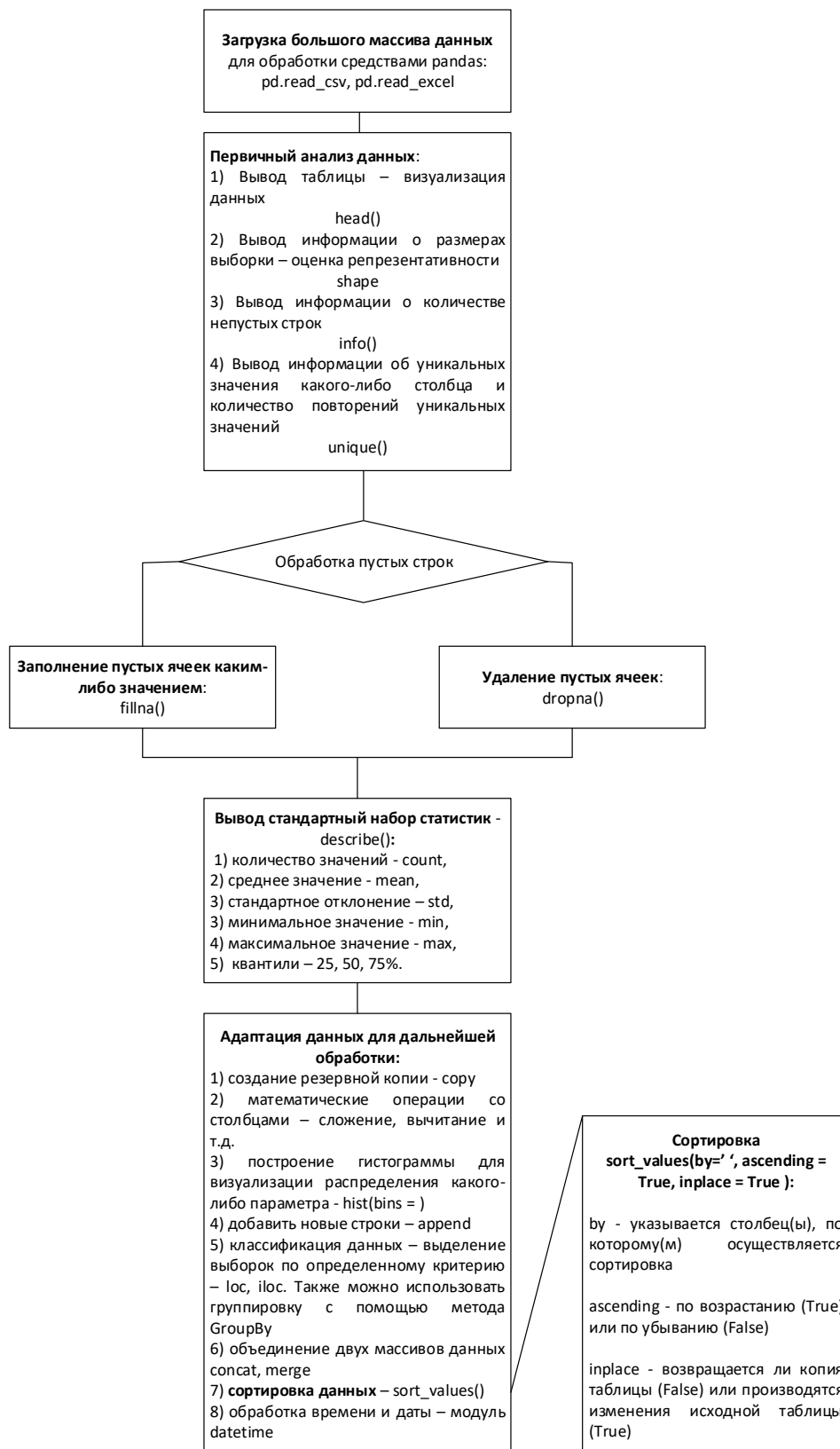


Рисунок 2 – Алгоритм обработки БД в pandas

Таким образом по окончании магистратуры студенты обладают инструментарием, который поможет им в рабочей деятельности или поможет сменить вектор карьеры.

### Список использованных источников

1. Михеев А.В. Анализ больших данных для обоснования решений по научно-технологическому развитию в энергетике // Информационные и математические технологии в науке и управлении. 2020. № 4. С. 158-167.
2. Аникеева А.Е. Применение технологии «большие данные» в энергетике // Современные проблемы телекоммуникаций. Материалы Всероссийской научно-технической конференции с международным участием. Под редакцией А.В. Ефимова, Т.И. Монастырской. Новосибирск. 2023. С. 158-163.
3. Цой А.Д. Перспективы применения больших данных в энергетике железной дороги // Транспорт: наука, образование, производство Современные проблемы телекоммуникаций. Сборник научных трудов международной научно-практической конференции. Том 3. Ростов-на-Дону. 2022. С. 55-58.
4. Гафуров И.А. Создание системы обработки больших данных для прогнозирования отказов оборудования и сокращения времени простоя на объектах топливно-энергетического комплекса // Тинчуринские чтения - 2021 «Энергетика и цифровая трансформация». Материалы Международной молодежной научной конференции. Том 3. Казань, 2021. С. 55-58.
5. Копайгородский А.Н., Хайруллина Е.П. Проектирование и реализация инструментальных средств для семантического анализа больших данных о научных и технологических решениях в области энергетики // Информационные и математические технологии в науке и управлении. 2021. № 4. С. 100-110.