

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

ПРИМЕНЕНИЕ НЕЙРОСЕТИ НА PYTHON ПРИ НАПИСАНИИ МАГИСТЕРСКОЙ ДИССЕРТАЦИИ

Аннотация

В статье рассматривается возможность использования python для проведения аналитического исследования с применением нейросети для решения задач, появляющихся при написании магистерской диссертации по направлению 13.04.01 «Теплоэнергетика и теплотехника». В статье показан пример диссертации, в которой изучается использование водорослей в качестве биотоплива. Программа на языке python используется для аналитического исследования влияния различных параметров на продуктивность роста водорослей. Разработка велась в среде Google Colab. Для создания нейросети использовались библиотеки python tensorflow и sklearn, для предварительной обработки датасета использовались библиотеки numpy и pandas.

Ключевые слова:

большие данные, теплоэнергетика, магистратура, нейросеть

APPLICATION OF A NEURAL NETWORK IN PYTHON WHEN WRITING A MASTER THESIS

Abstract

The article discusses the possibility of using python to conduct an analytical study using a neural network to solve problems that arise when writing a master's dissertation in the direction 13.04.01 "Heat power engineering and heat engineering". The article shows an example of a dissertation that studies the use of algae as a biofuel. A python program is used to analytically study the influence of various parameters on the productivity of algae growth. The development was carried out in the Google Colab environment. The python tensorflow and sklearn libraries were used to create the neural network, and the numpy and pandas libraries were used to preprocess the dataset.

Keywords:

Python, big data, thermal power engineering, magistracy, neural network

Использование нейросетей для решения задач, которые возникают перед студентом-магистрантом, является перспективным направлением [1-5]. Таким образом студенты учатся использовать современный инструментарий в своей профессиональной или исследовательской деятельности. После освоения данного инструментария выпускники становятся более востребованными на рынке труда.

Рассматриваемая магистерская диссертация рассматривает возможность использования водорослей в качестве биотоплива третьего поколения. Помимо подробного анализа литературы [6] и разработки установки для сжигания водорослей работа также включает в себя подробное аналитическое исследование. Аналитические исследования были направлены на определение зависимости

объемов производства (роста) водорослей от ряда параметров (температура, продолжительность роста, насыщением фосфором, аммиаком и т.д.). Данное исследование проводилось на основании набора данных (dataset) по водорослевым фермам в США. Данные были опубликованы в открытом доступе на площадке для машинного обучения kaggle.com и являются частью крупного исследования The Algae Testbed Public-Private Partnership framework; establishment of a national network of testbed sites to support sustainable algae production. Для анализа данных использовался код на языке python с среде Google Colab. Первоначальный набор данных состоял из 21108 записей для 44 параметров, не считая текстовых показателей (дата, название пруда и др.).

Первоначально была проведена подготовка данных:

1) За целевой параметр был взят урожай водорослей (Harvest.Vol..L.), поэтому все пустые строки по этому параметру были, в итоге был получен массив из 2210 записей.

2) Другие пустые ячейки были заполнены значениями (средним, максимальным, минимальным).

3) Была решена проблема разрозненности данных. Некоторые параметры меняются в пределах от сотен до тысяч, другие – от десятых до единиц. Работа аналитических моделей машинного обучения с такими показателями окажется некорректной. Поэтому было принято решение нормализовать данные, использовалась min-max нормализация.

Для установления связи между параметрами была построена матрица корреляций, для лучшей визуализации была построена тепловая карта (рисунок 1).

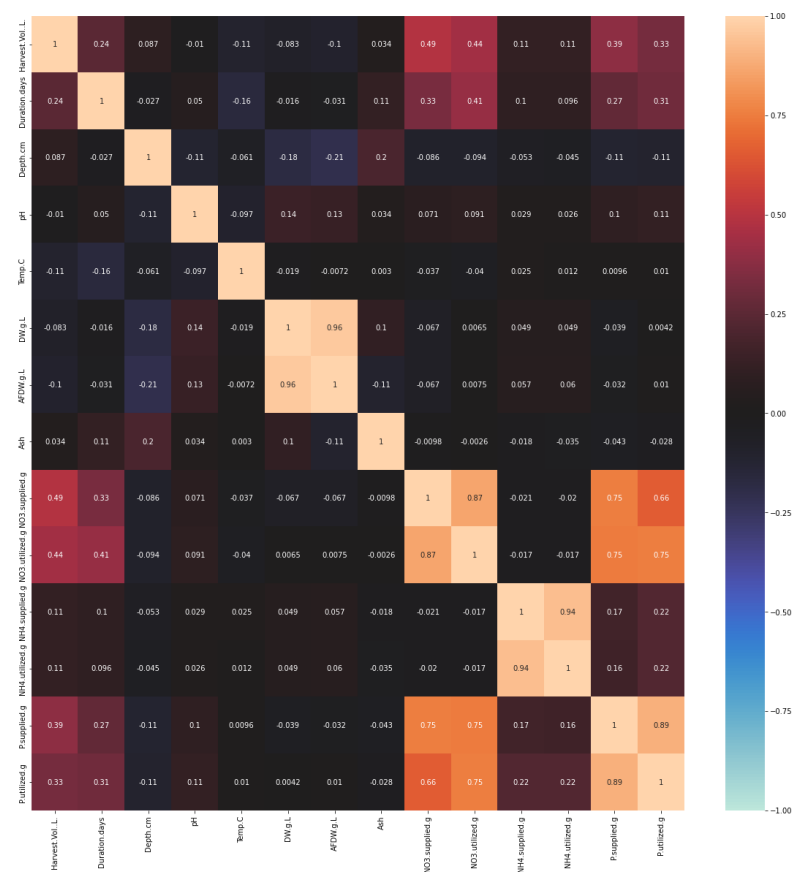


Рисунок 1 – Тепловая карта (матрица корреляций).

Из-за низкой корреляционной связи были удалены следующие параметры – 'pH', 'Ash', 'Temp.C', 'DW.g.L', 'AFDW.g.L', 'NH4.supplied.g', 'NH4.utilized.g'.

Была создана нейросеть, код которой представлен на рисунке 2.

```

# Создаём полносвязную сеть
model = Sequential()
# Добавляем полносвязный слой на 24 нейрона с relu- активацией, указыва
ем входной размер 6, соответствующий размеру входных данных
model.add(Dense(24, input_dim=6, activation="relu"))
# Добавляем полносвязный слой на 48 нейрона с relu-активацией
model.add(Dense(48, activation="relu"))
# Добавляем полносвязный слой на 10 нейрона с softmax-активацией
model.add(Dense(1, activation="softmax"))
# Компилируем модель
model.compile(loss="categorical_crossentropy", optimizer="adam", metrics=["ac
curacy"])

```

```

# Обучаем нейросеть
#fit - функция обучения нейронной сети
# input_data, output_data - обучающая выборка, входные и выходные данные
#batch_size - размер батча, количество примеров, которое обрабатывает ней
ронная сеть перед одним изменением весов
#epochs - количество эпох, когда нейронная
сеть обучается на всех примерах выборки
#verbose - 0 - не визуализировать ход обучения, 1 - визуализировать
model.fit(input_data, output_data, batch_size=32, epochs=15, verbose=1)

```

Рисунок 2 – Код нейросети

В качестве параметра для оценки точности вычислений использовался коэффициент детерминации R^2 , сравнение моделей приведено на рисунке 3.

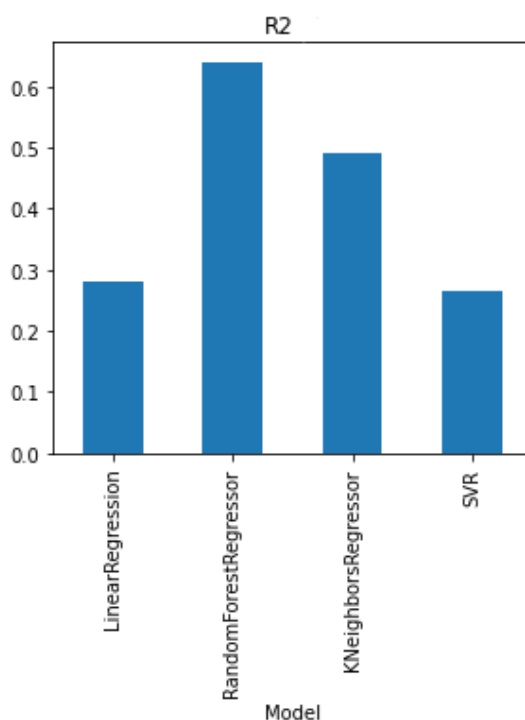


Рисунок 3 – Сравнение моделей для регрессионного анализа

После всестороннего анализа исходных данных был сделан вывод, что данная выборка не может выступать в качестве рабочего материала для создания предиктивной модели. Причинами непригодности выборки являются: малое количество записей, пропуск большого количества значений, несогласованность во времени некоторых замеров (замеры проводились тремя группами исследователей в разное время, но в одни календарный период).

Модель случайного леса показала наилучший результат, однако после всестороннего анализа исходных данных можно сделать вывод, что данная выборка не может выступать в качестве рабочего материала для создания предиктивной модели. Причинами непригодности выборки являются: малое количество записей, пропуск большого количества значений, несогласованность во времени некоторых замеров (замеры проводились тремя группами исследователей в разное время, но в одни календарный период).

Таким образом использование нейросетей, написанных на языке программирования python, является перспективным направлением при проведении диссертационных исследований студентами-магистрантами.

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.

6. Iglina T., Iglin P., Pashchenko D. Industrial CO2 Capture by Algae: A Review and Recent Advances //Sustainability. – 2022. – Т. 14. – №. 7. – С. 3801.