

Разведочный анализ данных о прогнозе погоды

Екимова Яна Сергеевна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Данная научная статья посвящена проведению анализа данных о прогнозе погоды с использованием Google Colaboratory. Для достижения поставленной цели использовался набор данных, содержащий информацию о температуре, влажности, атмосферном давлении и других метеорологических показателях. Используя Python и библиотеки анализа данных, такие как numpy, pandas, matplotlib.pyplot был проведен анализ данных.

Ключевые слова: Google Colaboratory, pandas, numpy, matplotlib.pyplot анализ данных.

Exploratory analysis of weather forecast data

Ekimova Yana Sergeevna

Sholom Aleichem Priamurskiy State University

Student

Abstract

This scientific article is devoted to the analysis of weather forecast data using Google Collaboration. To achieve this goal, a data set containing information on temperature, humidity, atmospheric pressure and other meteorological indicators was used. Using Python and data analysis libraries such as numpy, pandas, matplotlib.pyplot, data analysis was performed.

Key words: Google Coollaboratory, pandas, numpy, matplotlib.pyplot data analysis.

1. Введение

1.1. Актуальность

Google Colab остается актуальным и популярным выбором для проведения исследований, разработки моделей машинного обучения и выполнения вычислений в облаке.

Обзор исследований

А. М. Васильченко описал как проводить анализ данных при помощи библиотек Python [1]. Н.Е. Косых предложил подход к подбору наиболее производительного набора параметров для объекта классификатора текста. Для вычислений использовал облачный сервис Google Colaboratory, выполняющий код на языке Python внутри браузера, используя виртуальные

аппаратные ресурсы [2]. Т.М. Татарникова, Е.Д.Пойманова, П.Ю.Богданов, Е.В.Краева, С.А. Веревкин в статье рассмотрели способы и методы изучения и построения нейронных сетей. Показано, что изучение принципов функционирования нейронных сетей, их применение для решения тех или иных задач возможны только через практику. Проведен анализ различных программных сред, которые могут быть использованы на лабораторных и практических занятиях по изучению и применению нейронных сетей. Выделен современный облачный сервис Google Colaboratory, рекомендуемый для обучения основам нейронных сетей благодаря наличию в нем предустановки библиотеки Tensorflow и библиотеки для работы на языке Python, бесплатного доступа к графическим процессорам, возможности написания и выполнения программного кода в браузере, а также отсутствию необходимости специальной настройки сервиса [3]. Е.А.Григорьев, Н.С.Климов в статье рассматривали разведочный анализ данных. Описали инструменты реализации анализа, библиотеки Python. Представлен пример выполненный на данных обнаружению присутствия людей в помещении [4].

1.2. Цель исследования

Цель исследования – выполнить анализ данных с помощью Google Colaboratory.

2. Материалы и методы

В данном исследовании используется Google Colaboratory для анализа данных.

Набор данных о погоде содержит информацию, о температуре, влажности, атмосферном давлении и других метеорологических показателях. Данные для работы можно скачать по ссылке <https://www.kaggle.com/datasets/thedevastator/weather-prediction/data>.

3. Результаты

Создадим блокнот в Google Colab и загрузим два файла "weather_prediction_dataset.csv", "weather_prediction_bbq_labels.csv". Для этого необходимо открыть вкладку «файл», нажать на «Загрузить», и выбрать два файла "weather_prediction_dataset.csv", "weather_prediction_bbq_labels.csv"(Рис.1).

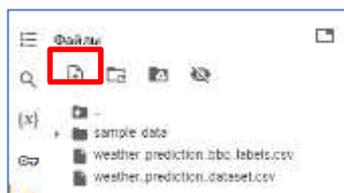


Рис.1-Загрузка данных

В блоке кода импортируем необходимые библиотеки для анализа данных, в данном случае numpy, pandas, matplotlib.pyplot. Загружаем данные

из CSV-файла (Рис.2).

```
[2]: import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
import numpy as np
from sklearn.preprocessing import StandardScaler
import os
for filename, _ in filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(filename, filename))

[1]: df=pd.read_csv("weather_prediction_dataset.csv")
df_idx=pd.read_csv("weather_prediction_idx_labels.csv")
```

Рис.2-Код

Визуализируем первые 5 записей на экран.

```
[ ] df.head()
```

	DATE	MONTH	BASEI_cloud_cover	BASEI_humidity	BASEI_pressure	BASEI_global_radiation	BASEI_precipitation	BASEI_sunshine	BC
0	20000101	1	8	0.89	1.0286	0.20	0.03	0.0	
1	20000102	1	8	0.87	1.0318	0.25	0.00	0.0	
2	20000103	1	5	0.81	1.0314	0.50	0.00	3.7	
3	20000104	1	7	0.79	1.0262	0.83	0.25	6.9	
4	20000105	1	5	0.90	1.0196	0.51	0.07	3.7	

5 rows x 10 columns

Рис.3-Визуализирование первых пяти записей

При помощи команды `columns` выведем то, что есть в столбце 'MONTH'.

```
[ ] df.columns
df['MONTH'].unique()

array([ 1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12])
```

Рис.4-Результат команды `columns`

Теперь построим графики, где видно информацию о погоде за каждый месяц и по разным характеристикам.

```
[ ] cols=df.columns
for label in cols[-1]:
    plt.hist(df[df['MONTH']==1][label],color='blue',label='January',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==2][label],color='red',label='February',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==3][label],color='green',label='March',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==4][label],color='orange',label='April',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==5][label],color='black',label='May',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==6][label],color='cyan',label='June',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==7][label],color='yellow',label='July',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==8][label],color='magenta',label='August',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==9][label],color='pink',label='September',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==10][label],color='olive',label='October',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==11][label],color='grey',label='November',alpha=0.1,density=True)
    plt.hist(df[df['MONTH']==12][label],color='purple',label='December',alpha=0.1,density=True)
plt.xlabel('Month')
plt.ylabel('Month')
plt.legend()
plt.show()
```

Рис.5-Код

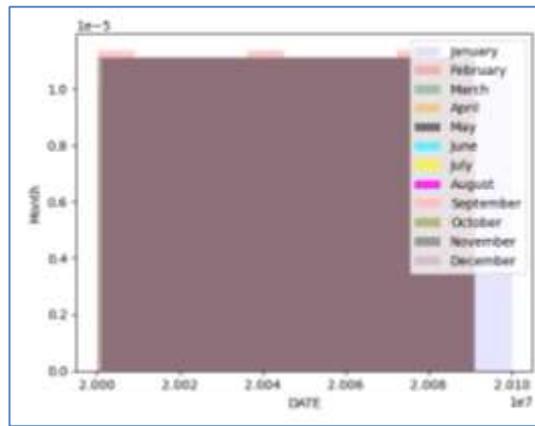


Рис.6-Data

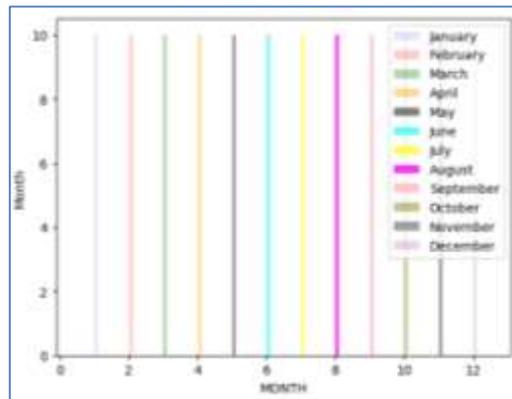


Рис.7-Month

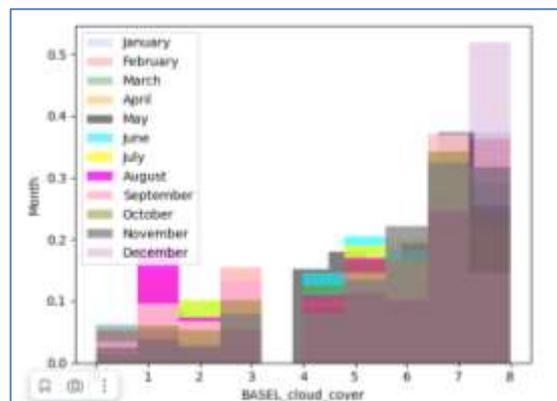


Рис.8- Базовое облачное покрытие в Базели

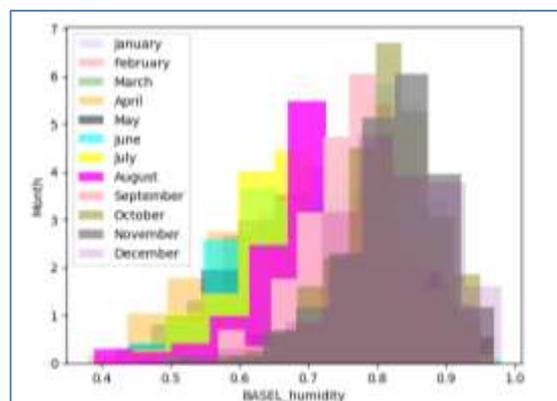


Рис.9- Влажность в Базели

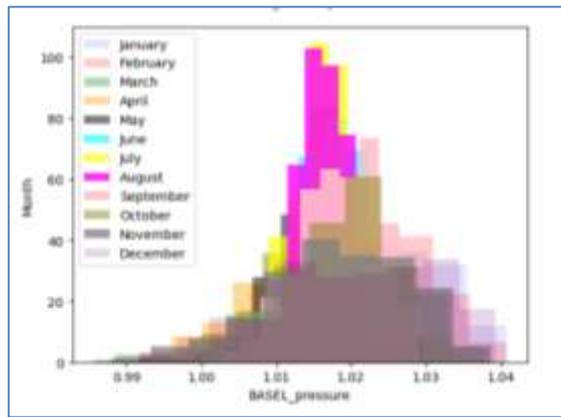


Рис.10- Давление в Базели

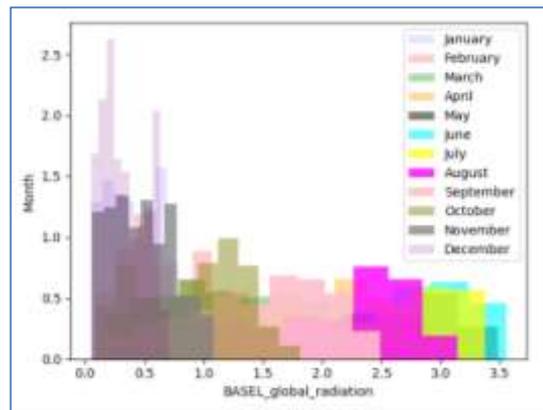


Рис.11- Глобальное излучение в Базели

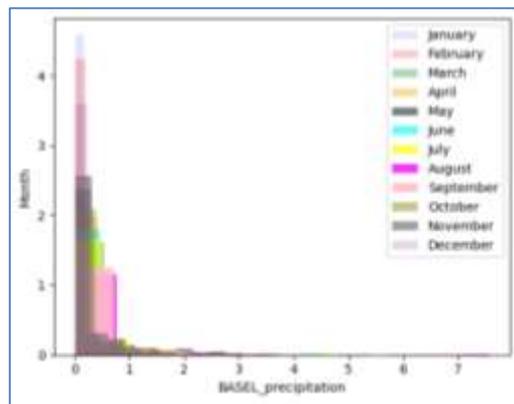


Рис.12- Осадки в Базели

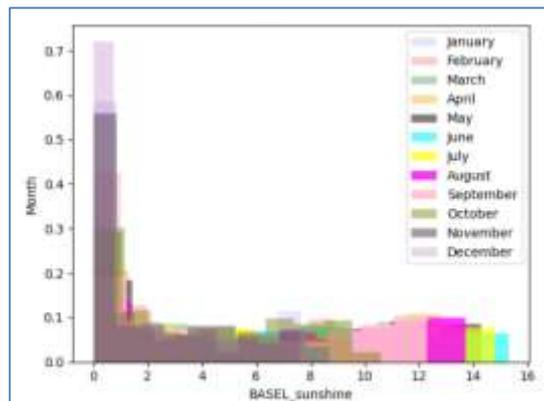


Рис.13- Уровень солнечного света в Базели

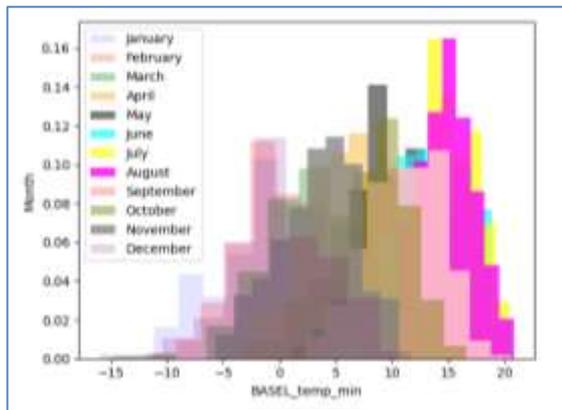


Рис.14- Минимальная температура в Базели

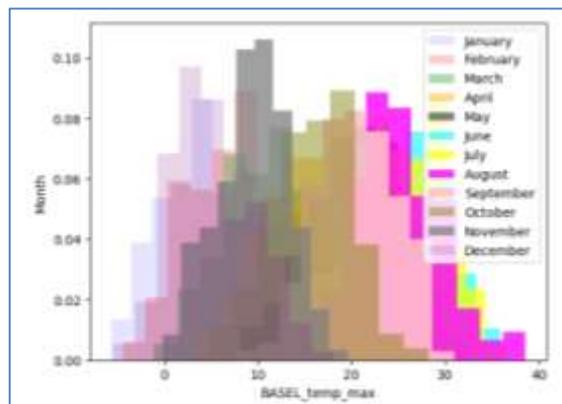


Рис.15- Максимальная температура в Базели

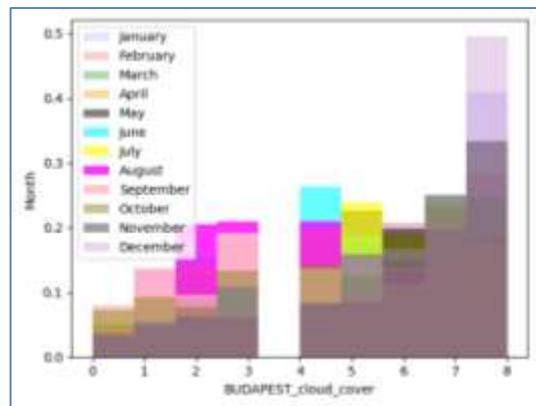


Рис.16- Облачный покров в Будапеште

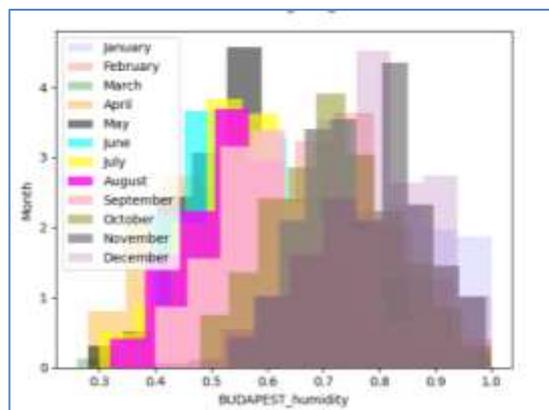


Рис.17- Влажность в Будапеште

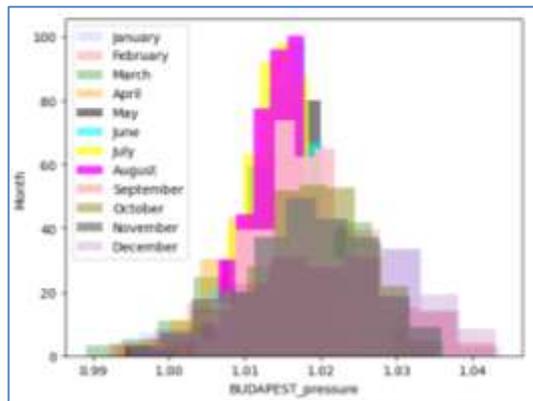


Рис.18- Давление в Будапеште

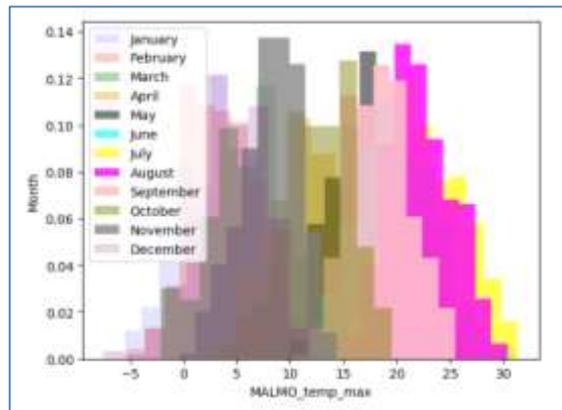


Рис.19- Максимальная температура в Мальмё

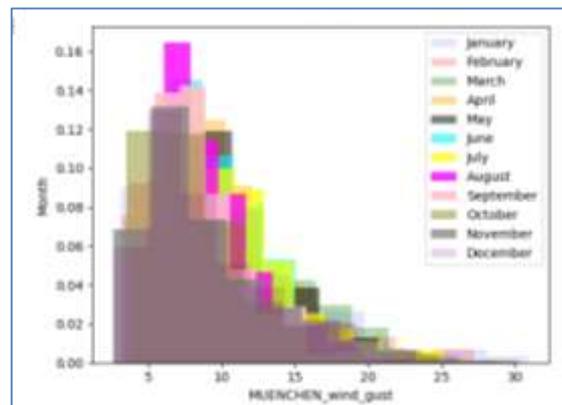


Рис.20-Ветер в Мюнхене

4. Выводы

В ходе исследования были выполнены следующие этапы: загрузка данных, визуализация данных о погоде с использованием графиков. Были выделены регионы с наиболее переменчивой погодой.

Библиографический список

1. Васильченко А. М. Как проводить анализ данных при помощи python? //Иновации и инвестиции. 2023. №. 5. С. 161-165.
2. Косых Н. Е. Оценка гиперпараметров при анализе тональности

-
- русскоязычного корпуса текстов//Интеллектуальные технологии на транспорте. 2020. №. 3 (23). С. 41-44.
3. Богданов П. Ю. и др. Программные среды для изучения основ нейронных сетей //Программные продукты и системы. 2021. №. 1. С. 145-150.
 4. Григорьев Е. А., Климов Н. С. Разведочный анализ данных с помощью python //E-Scio. 2020. №. 2 (41). С. 165-176.