

Bài 10: Giới thiệu Bokeh

Giới thiệu

Bokeh là một thư viện trong Python giúp trực quan hóa dữ liệu có tương tác (interactive data visualization). Đặc biệt Bokeh giúp bạn mang các biểu đồ lên web một cách dễ dàng.

Vài khái niệm cơ bản

Document	Bokeh xem biểu đồ chỉ là một phần của trang web cần trình bày. Vì thế Bokeh tổ chức biểu đồ và nhiều thông tin khác nữa để xuất kết quả, thông dụng nhất thì kết quả là một website, dưới dạng Bokeh Document (tài liệu)
Application	Đây là phần mã nguồn đảm nhận phần nghiệp vụ vẽ biểu đồ và xử lý các sự kiện tương tác (như người dùng thay đổi điều kiện dữ liệu, dữ liệu nguồn có thay đổi, v.v...) để kết ra Bokeh Document.
Server	Gồm 3 thành phần Bokeh Document và Application ở trên để cung cấp dịch vụ hiển thị biểu đồ cho người dùng qua môi trường mạng.
Figure	Là đối tượng được ví như là một bức tranh mà bạn là họa sĩ sẽ vẽ biểu đồ và trang trí lên bức tranh.

Đọc thêm tài liệu tại:

https://docs.bokeh.org/en/latest/docs/user_guide/concepts.html
<https://docs.bokeh.org/en/latest/docs/reference/server.html>

Cài đặt

Tài liệu này dùng bokeh phiên bản 2.2.3 với lệnh cài đặt như sau:

```
pip install bokeh==2.2.3
```

Trải nghiệm ví dụ

Quay lại dữ liệu mẫu GDP của Việt Nam, sử dụng thư viện bokeh để vẽ biểu đồ line như sau:

```
from bokeh.plotting import figure, output_file, show

output_file("output.html")

gdp = [6.78, 5.89, 5.03, 5.42, 5.98, 6.68, 6.21, 6.81, 7.08, 7.02]
```

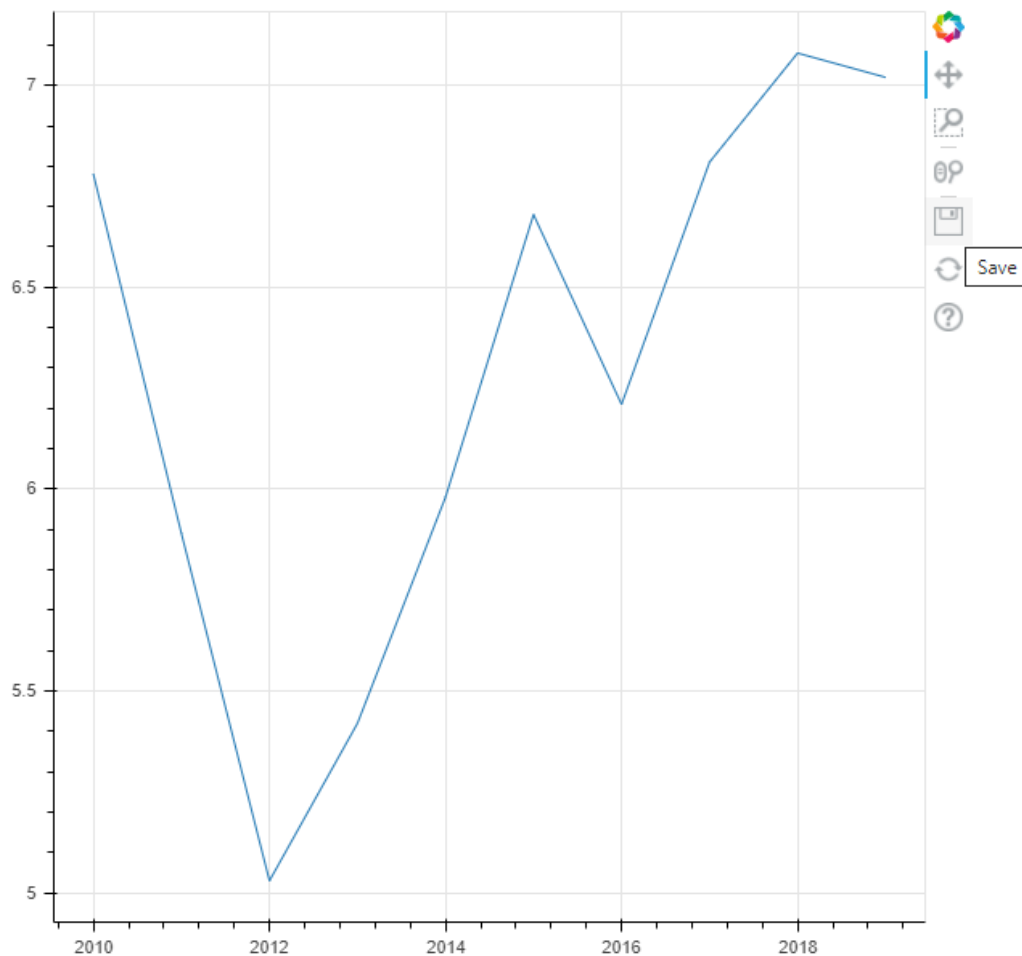
Ứng dụng Phân tích dữ liệu và Trí tuệ nhân tạo với Python

```
year = [2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019]

p = figure()
p.line(x=year, y=gdp)

show(p)
```

Kết quả:



Bạn để ý thấy bên phải có thanh công cụ. Trong đó có biểu tượng cái đĩa (disk) cho phép bạn lưu (Save) cái hình của biểu đồ thành file ảnh.

Phân tích mã nguồn

Dòng đầu tiên để import các hàm sau trong thư viện `bokeh.plotting`:

- Hàm `output_file(file_path)` để thiết lập biểu đồ sẽ được thi thành file trong đường dẫn `file_path`.
- Hàm `figure()` để tạo ra đối tượng để vẽ biểu đồ. Hãy tưởng tượng giống như một họa sĩ sẽ căng khung vải để vẽ bức tranh. Đối tượng được lưu trong biến `p` (tạm hiểu là picture).

- Hàm `p.line(x= array, y = array)` để vẽ một đường thẳng lên picture với thông số trục x và trục y như chỉ định.
- Hàm `show(p)` sẽ gọi trình duyệt mở file trong hàm `output_file(file_path)` ở trên để bạn xem.

Đưa biểu đồ lên Web

Xuất file HTML

Thay vì trong ví dụ trên dùng hàm `output_file(...)` để lưu biểu đồ ra file HTML, chúng ta thử sửa dùng hàm `curdoc()` để thêm “bức tranh” vào tài liệu.

Hãy lưu đoạn code sau vào file. Ví dụ file: `D:\MyPython\bokeh\bokeh_gpd.py`

```
from bokeh.plotting import figure, curdoc

gdp = [6.78, 5.89, 5.03, 5.42, 5.98, 6.68, 6.21, 6.81, 7.08, 7.02]
year = [2010,2011,2012,2013,2014,2015,2016,2017,2018,2019]

p = figure()
p.line(x=year, y=gdp)

curdoc().add_root(p)
```

Đứng trong thư mục `D:\MyPython\bokeh` của cửa sổ lệnh có Python và đã cài thư viện Bokeh thực hiện lệnh sau:

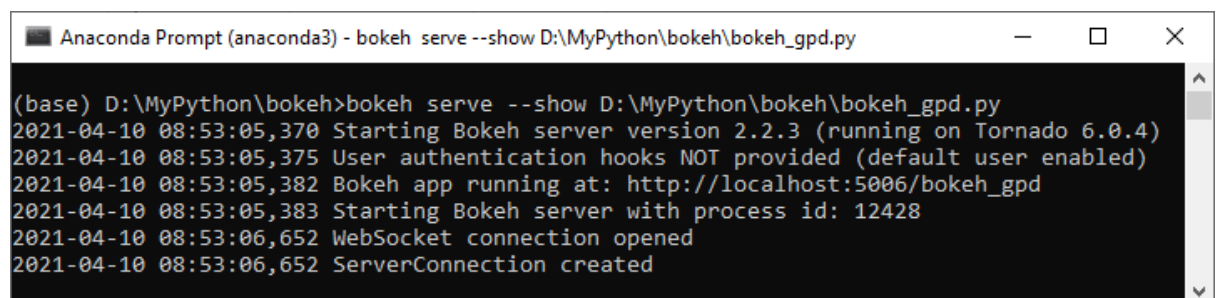
```
bokeh serve --show bokeh_gpd.py
```

Hoặc có thể gõ đường dẫn đầy đủ của file .py như sau:

```
bokeh serve --show D:\MyPython\bokeh\bokeh_gpd.py
```

Cửa sổ lệnh sẽ hiển thị các kết quả sau cho thấy là lệnh bokeh sẽ mở một Bokeh server với port 5006 và phục vụ biểu đồ ở trên tại đường link:

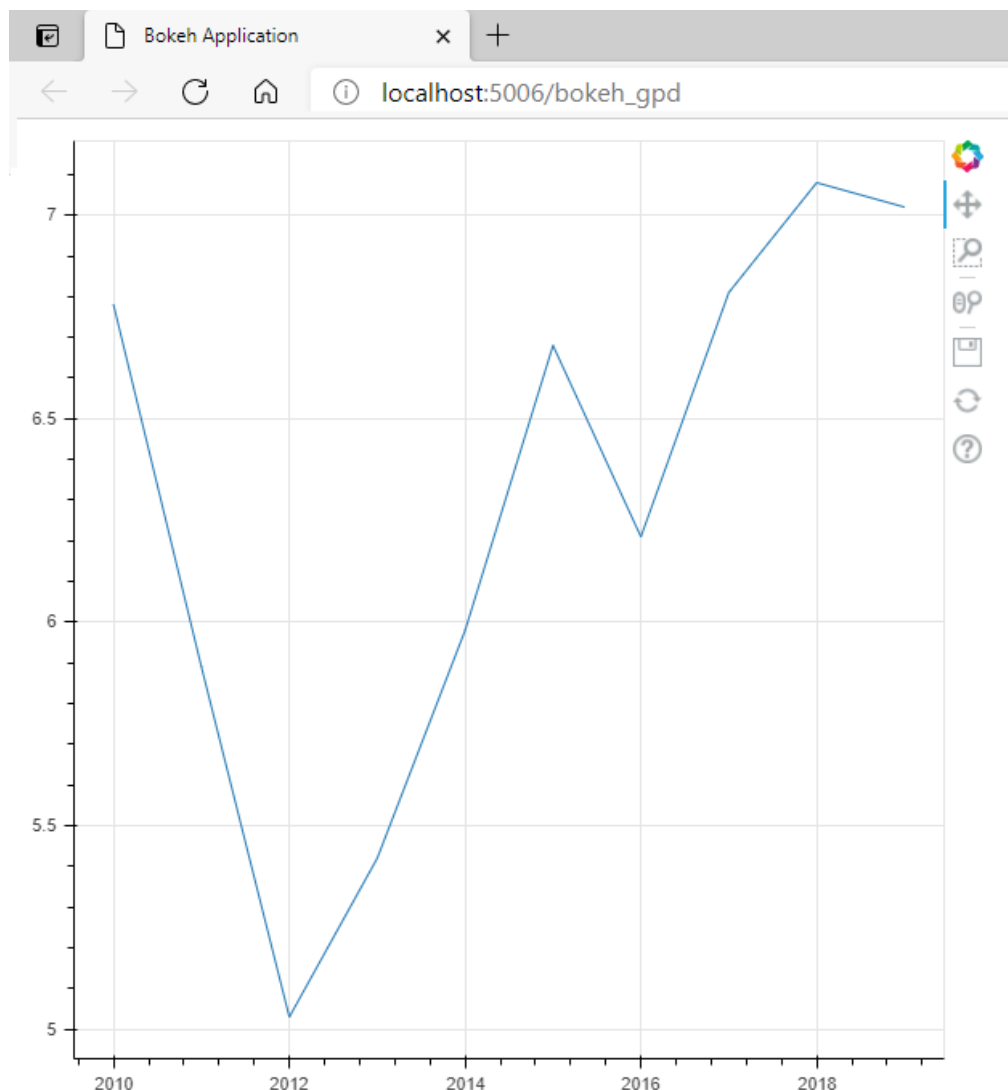
http://localhost:5006/bokeh_gpd



```
Anaconda Prompt (anaconda3) - bokeh serve --show D:\MyPython\bokeh\bokeh_gpd.py

(base) D:\MyPython\bokeh>bokeh serve --show D:\MyPython\bokeh\bokeh_gpd.py
2021-04-10 08:53:05,370 Starting Bokeh server version 2.2.3 (running on Tornado 6.0.4)
2021-04-10 08:53:05,375 User authentication hooks NOT provided (default user enabled)
2021-04-10 08:53:05,382 Bokeh app running at: http://localhost:5006/bokeh_gpd
2021-04-10 08:53:05,383 Starting Bokeh server with process id: 12428
2021-04-10 08:53:06,652 WebSocket connection opened
2021-04-10 08:53:06,652 ServerConnection created
```

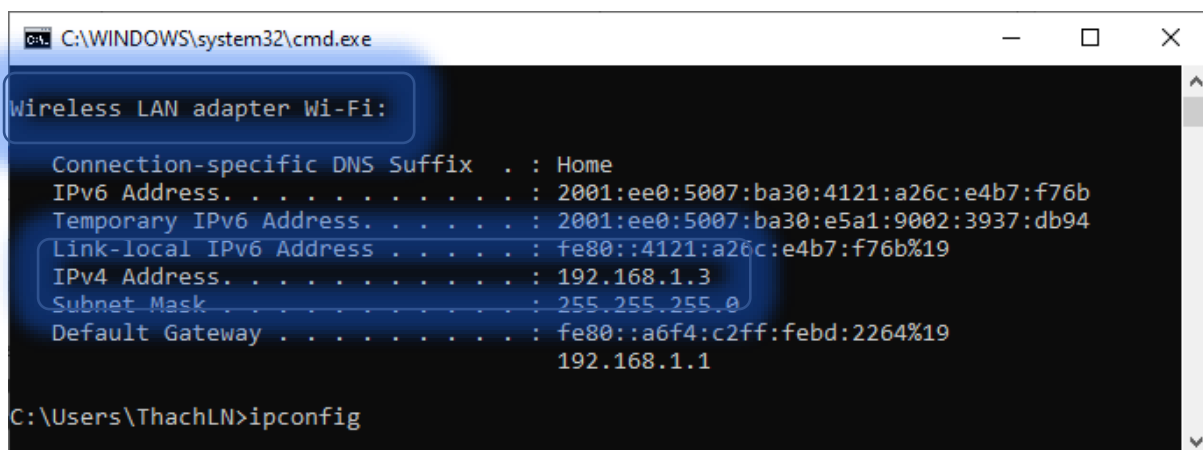
Đồng thời trình duyệt trên máy bạn sẽ tự mở được link ở trên với kết quả như sau:



Nếu bạn muốn bạn bè, đồng nghiệp của mình đang dùng chung Network (ví dụ đang cùng vào chung một WiFi) thì chạy lại lệnh `bokeh` với tham số cho phép truy cập mạng như sau:

```
bokeh serve --show bokeh_gpd.py --allow-websocket-origin=192.168.1.3:5006
```

Tham số `--allow-websocket-origin=192.168.1.3:5006` có nghĩa là cho phép máy tính của bạn có địa chỉ IP là 192.168.1.3 chạy chương trình `bokeh` với cổng 5006. Để kiểm tra chính xác IP mạng của máy bạn thì dùng lệnh `ipconfig`. Sau đó tìm đoạn có chữ “Wireless LAN adapter Wi-Fi” và mục IPv4 Address. Ví dụ:



```
C:\WINDOWS\system32\cmd.exe

Wireless LAN adapter Wi-Fi:

    Connection-specific DNS Suffix  . : Home
    IPv6 Address. . . . . : 2001:ee0:5007:ba30:4121:a26c:e4b7:f76b
    Temporary IPv6 Address. . . . . : 2001:ee0:5007:ba30:e5a1:9002:3937:db94
    Link-local IPv6 Address . . . . . : fe80::4121:a26c:e4b7:f76b%19
    IPv4 Address. . . . . : 192.168.1.3
    Subnet Mask . . . . . : 255.255.255.0
    Default Gateway . . . . . : fe80::a6f4:c2ff:febd:2264%19
                               192.168.1.1

C:\Users\ThachLN>ipconfig
```

Lúc này có thể mời bạn bè, đồng nghiệp vào địa chỉ sau để xem thành quả của bạn:

http://192.168.1.3:5006/bokeh_gpd

Để tắt bokeh server thì đứng trong cửa sổ lệnh đang chạy, nhấn phím Ctrl+C và đợi vài giây, bokeh server sẽ ngưng với thông báo sau:

Interrupted, shutting down

Như vậy bạn đã biết cách vẽ một biểu đồ đơn giản với thư viện Bokeh và Bokeh server để có thể truy cập biểu đồ từ mạng. Tuy nhiên theo như cách chạy lệnh Bokeh ở trên thì có chút bất tiện là mỗi lần code xong phải chạy lại lệnh. Phần tiếp theo sẽ hướng dẫn bạn khởi chạy server bằng code Python luôn.

Sử dụng Bokeh Server

Chạy thử ví dụ sau:

```
from bokeh.plotting import figure
from bokeh.server.server import Server
from bokeh.application import Application
from bokeh.application.handlers.function import FunctionHandler

def make_document(doc):
    gdp = [6.78, 5.89, 5.03, 5.42, 5.98, 6.68, 6.21, 6.81, 7.08, 7.02]
    year = [2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019]

    p = figure()
    p.line(x=year, y=gdp)

    doc.add_root(p)
```

Ứng dụng Phân tích dữ liệu và Trí tuệ nhân tạo với Python

```
doc.title = "Sample chart"
doc.add_root(p)

apps = {'/': Application(FunctionHandler(make_document))}

server = Server(apps, port=5000)
server.start()
```

Sau đó dùng trình duyệt mở địa chỉ sau bạn sẽ thấy kết quả tương tự như phần trước:

<http://localhost:5000>

Cấu trúc mã nguồn:

Để ý 2 lệnh cuối cùng:

```
server = Server(apps, port=5000)
server.start()
```

để khởi động Bokeh server với port 5000. Trong đó Bokeh server cần một tham số là Application (xem lại phần khái niệm cơ bản ở trên). Tên đầy đủ của Application ở đây là “Bokeh Server Tornado application”. Tra cứu tài liệu tại:

<https://docs.bokeh.org/en/latest/docs/reference/server/tornado.html#bokeh.server.tornado.BokehTornado>

bạn sẽ thấy cách khai báo Application này là dạng Dictionary như:

```
{ '/' : applications }
```

Cụ thể trong ví dụ trên thì khai báo Application như sau:

```
apps = {'/': Application(FunctionHandler(make_document))}
```

Dấu ‘/’ có nghĩa là khi người dùng gõ đường dẫn URL trên trình duyệt như <http://localhost:5000/> (nếu không có dấu xuyệt / cuối thì coi như là có) thì Máy tính (trong ngữ cảnh này gọi là máy chủ nhân, chính là cái máy tính mà bạn đang thực hành) sẽ nhận và xử lý URL gồm 2 phần:

- Phần “<http://localhost:5000/>”: sẽ được máy có tên là localhost (tức là máy của bạn đang thực hành) có một phần mềm đang chạy ở port 5000 (trong trường này chính là Bokeh Server trong ví dụ) nhận và xử lý yêu cầu.

- Phần “/” (dấu duyệt phải ở cuối): gọi là yêu cầu (request). Dấu / gọi là root. Có thể URL đầy đủ không có dấu duyệt thì root là rỗng. Khi gặp yêu cầu root này thì đối tượng **applications** sẽ xử lý. Cụ thể applications trong ví dụ trên là:

```
Application(FunctionHandler(make_document))
```

Đây là lệnh khởi tạo đối tượng Application⁶ với tham số là đối tượng Handler⁷. Cụ thể là:

```
FunctionHandler(make_document)
```

Lệnh này sử dụng đối tượng FuncHandler để chỉ định hàm lo xử lý việc tạo ra bokeh document. Bạn xem cách khai báo hàm make_document với tham số là doc như sau:

```
def make_document(doc):  
    p = figure()  
    ...  
    doc.add_root(p)  
  
    ...  
    doc.add_root(p)
```

Hàm make_document cập nhật tham số doc bằng cách gọi hàm add_root(p) với p là đối tượng Figure.

Stop server

Thực hiện lệnh Python sau:

```
server.stop()
```

Mã nguồn đầy đủ có sẵn lệnh server.stop() được chú thích.

```
from bokeh.plotting import figure  
from bokeh.server.server import Server  
from bokeh.application import Application
```

⁶ Tra cứu tại:

<https://docs.bokeh.org/en/latest/docs/reference/application/application.html#bokeh.application.application.Application>

⁷

<https://docs.bokeh.org/en/latest/docs/reference/application/handlers/handler.html#bokeh.application.handlers.handler.Handler>

Ứng dụng Phân tích dữ liệu và Trí tuệ nhân tạo với Python

```
from bokeh.application.handlers.function import FunctionHandler

def make_document(doc):
    gdp = [6.78, 5.89, 5.03, 5.42, 5.98, 6.68, 6.21, 6.81, 7.08, 7.02]
    year = [2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019]

    p = figure()
    p.line(x=year, y=gdp)

    doc.add_root(p)

    doc.title = "Sample chart"
    doc.add_root(p)

apps = {'/': Application(FunctionHandler(make_document))}

print('Start bokeh server with port 5000...')
server = Server(apps, port=5000)
server.start()
print('Try the application with URL: http://localhost:5000')

# server.stop()
```

Nếu bạn chạy đoạn code trên 2 lần thì sẽ bị báo lỗi như sau:

OSError: [winError 10048] Only one usage of each socket address (protocol/network address/port) is normally permitted

Lý do là cái port 5000 đã được sử dụng và bokeh server đang chạy.

Muốn tắt server thì bôi lệnh đoạn `server.stop()` và thực thi nó. Nếu dùng IDE Spyder thì bôi và nhấn F9.