

CS634 Data Mining

Implementation Project

-Alfred Zane Rajan Velladurai

UCID: ar2257

Platforms:

Programming language: Python
IDE: Spyder
Operating System: Windows
Hardware: Laptop

Source Code:

```
import urllib
from bs4 import BeautifulSoup
import urllib.request
import re

#%%
data = "DataMining.txt"

#%%
k = 200

IN = open(data)
base = list()
line = IN.readline()
while line != "":
    a = line.split()
    base.append(a[0])
    line = IN.readline()
IN.close()

nbrhd = [page for page in base]
n = 30
b = 0
adj = {}
for page in base:
    b+=1
    try:
```

```

resp = urllib.request.urlopen(page)
except urllib.error.HTTPError as e:
    continue
except urllib.error.URLError as e:
    continue
adj[b] = set()
c = 0
soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
for link in soup.find_all('a', href=True):
    lnk = link['href']
    if re.search("^http", lnk):
        if lnk not in nbrhd:
            nbrhd.append(lnk)
            n+=1
            adj[b].add(n)
        else:
            t = nbrhd.index(lnk) + 1
            adj[b].add(t)
            c+=1
    if c > k/30:
        break

#%%
n = -1
adj = {}
for page in nbrhd:
    n+=1
    try:
        resp = urllib.request.urlopen(page)
    except urllib.error.HTTPError as e:
        if e.code in (... , 403, ...):
            continue
    except urllib.error.URLError as e:
        continue
    adj[n] = set()
    soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
    for link in soup.find_all('a', href=True):
        lnk = link['href']
        if re.search("^http", lnk):
            if lnk in nbrhd and lnk != page:
                t = nbrhd.index(lnk)
                adj[n].add(t)

```

```

#%%
for edges in adj:
    adj[edges] = list(adj[edges])

#%%
ltr = 10

vert = len(nbrhd)
auth = []
hub = []
for i in range(vert):
    auth.append(1)
    hub.append(1)

for itr in range(ltr):

    auth = [0 for a in auth]
    for u in adj:
        for v in adj[u]:
            auth[v] += hub[u]
    norm = 0
    for a in auth:
        norm += a**2
    norm = norm**0.5
    auth = [a/norm for a in auth]

    hub = [0 for h in hub]
    norm = 0
    for u in adj:
        for v in adj[u]:
            hub[u] += auth[v]
        norm += hub[u]**2
    norm = norm**0.5
    hub = [h/norm for h in hub]

#%%
N = 10

au = {}
for i in range(vert):
    au[i] = auth[i]

print("The top " + str(N) + " Authority weights are for the following websites:")

```

```

print("Authority Weight \t Hub Weight \t\t Website")
for n in range(N):
    m = max(au, key=lambda key: au[key])
    au.pop(m)
    print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])

#%%
hu = {}
for i in range(vert):
    hu[i] = hub[i]

print("The top " + str(N) + " Hub weights are for the following websites:")
print("Hub Weight \t\t Authority Weight \t Website")
for n in range(N):
    m = max(hu, key=lambda key: hu[key])
    hu.pop(m)
    print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])

#%%
for page in nbrhd:
    print(page)
    try:
        resp = urllib.request.urlopen(page)
    except urllib.error.HTTPError as e:
        if e.code in (... , 403, ...):
            print("\n\n")
            continue
    except urllib.error.URLError as e:
        print("\n\n")
        continue
    soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
    try:
        print(soup.find('p').get_text())
    except AttributeError as e:
        print("\n\n")
        continue
    print("\n\n")

```

Description:

The program reads the base set of 30 pages from a txt file written manually. The source code then crawls those pages for links to further pages and looks for connections between these pages. Using these discovered edges an adjacency list is created.

Later from this adjacency list we implement the HITS algorithm to get the auth score and hub score iteratively until they converge.

We then have separate sections to printout the top authority scores, the top hub scores and also the whole neighbourhood graph URLs.

Let us now see the code in action for a maximum neighbourhood size of 300.

k = 300:

We first take the 30 websearch results for “Data Mining” and crawl through those 30 pages to get the links for our neighbourhood graph.

```
Spyder (Python 3.6)
File Edit Search Source Run Debug Consoles Projects Tool View Help
Editor - C:\Users\Alfred Zane Rajan\Documents\Implementation Project\crawlinghits.py [Run current cell (Ctrl+Enter) [Use #%% to create cells] C:\Users\Alfred Zane Rajan\Documents\Implementation Project\crawlinghits.py
Variable explorer
Name Type Value
Ink str http://support.twitter.com/articles/14226-how-to-find-your-twitter-sho ...
Python console
In [261]: k = 300
.....
.... IN = open(data)
.... base = []
.... line = IN.readline()
.... while line != '':
....     a = line.split()
....     base.append(a[0])
....     line = IN.readline()
.... IN.close()
.... nbrhd = [page for page in base]
.... n = 30
.... b = 0
.... adj = {}
.... for page in base:
....     b+=1
....     try:
....         resp = urllib.request.urlopen(page)
....     except urllib.error.HTTPError as e:
....         continue
....     except urllib.error.URLError as e:
....         continue
....     adj[b] = set()
....     c = 0
....     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
....     for link in soup.find_all('a', href=True):
....         link = link['href']
....         if re.search("http", link):
....             if link not in nbrhd:
....                 nbrhd.append(link)
....                 n+=1
....                 adj[b].add(n)
....             else:
....                 t = nbrhd.index(link) + 1
....                 adj[b].add(t)
....         c+=1
....     if c > k/30:
....         break
.... adj
52
```

We get the following list of URLs

```

10 #!/usr/bin/python
11 IN = open(data)
12 base = list()
13 line = IN.readline()
14 while line != '':
15     a = line.split()
16     b = a[0]
17     line = IN.readline()
18     IN.close()
19     IN.readline()
20 IN.close()
21
22 nbrhd = [page for page in base]
23 n = 30
24 k = 0
25 adj = {}
26 for page in base:
27     b+=1
28     try:
29         resp = urllib.request.urlopen(page)
30     except urllib.error.HTTPError:
31         continue
32     except urllib.error.URLError:
33         continue
34     adj[b] = set()
35     c = 0
36     soup = BeautifulSoup(resp, features='lxml')
37     for link in soup.findall('a'):
38         lnk = link['href']
39         if re.search("http", lnk):
40             if lnk not in nbrhd:
41                 nbrhd.append(lnk)
42             n+=1
43             adj[b].add(n)
44         else:
45             t = nbrhd.index(lnk)
46             adj[b].add(t)
47         c+=1
48     if c > k/30:
49         break
50
51 max
52

```

In [4]:

```

In [4]: nbrhd - List (254 elements)
Value
https://en.wikipedia.org/wiki/Data_mining
https://www.sas.com/en_us/insights/analytics/data-mining.html
https://searchsqlserver.techtarget.com/definition/data-mining
http://www.statsoft.com/textbook/data-mining-techniques
https://docs.oracle.com/cd/B28259_01/dam/me.111/b28129/process.htm
https://www.techopedia.com/definition/1181/data-mining
https://theappsolution.com/blog/development/data-mining-guide/
https://www.investopedia.com/terms/d/datamining.asp
https://link.springer.com/journal/10618
https://datafloq.com/read/data-mining-techniques-create-business-value ...
http://dmg.org/
https://www.cs.stonybrook.edu/~cse634/
https://scpd.stanford.edu/public/category/courseCategoryCertificatePro ...
https://onlinelibrary.wiley.com/journal/19321872
https://orange.biolab.si/training/
https://neilpatel.com/blog/data-mining/

```

Next we crawl through all the URLs in the neighbourhood graph inorder to find pages that link to one another and we simultaneously create an adjacency list.

```

49     break
50
51 max
52
53 n = -1
54 adj = {}
55 for page in nbrhd:
56     n+=1
57     try:
58         resp = urllib.request.urlopen(page)
59     except urllib.error.HTTPError as e:
60         if e.code in (... , 403, ...):
61             continue
62     except urllib.error.URLError as e:
63         continue
64     adj[n] = set()
65     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
66     for link in soup.findall('a', href=True):
67         lnk = link['href']
68         if re.search("http", lnk):
69             if lnk in nbrhd and lnk != page:
70                 t = nbrhd.index(lnk)
71                 adj[n].add(t)
72
73 max
74
75 for edges in adj:
76     adj[edges] = list(adj[edges])
77
78 max
79 Itr = 10
80
81 vert = len(nbrhd)
82 auth = []
83 hub = []
84 for i in range(vert):
85     auth.append(1)
86     hub.append(1)
87
88 for itr in range(Itr):
89
90     auth = [0 for a in auth]
91     for u in adj:

```

In [4]:

```

In [4]: n = -1
...: adj = {}
...: for page in nbrhd:
...:     n+=1
...:     try:
...:         resp = urllib.request.urlopen(page)
...:     except urllib.error.HTTPError as e:
...:         if e.code in (... , 403, ...):
...:             continue
...:     except urllib.error.URLError as e:
...:         continue
...:     adj[n] = set()
...:     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
...:     for link in soup.findall('a', href=True):
...:         lnk = link['href']
...:         if re.search("http", lnk):
...:             if lnk in nbrhd and lnk != page:
...:                 t = nbrhd.index(lnk)
...:                 adj[n].add(t)
...:         c+=1
...:     if c > k/30:
...:         break

```

In [4]:

```

In [4]: adj
Value
nbrhd list 254 ['https://en.wikipedia.org/wiki/Data_mining', 'https://www.sas.com/en_ ...']


```

```

49     break
50
51 #OK
52
53 n = -1
54 adj = {}
55 for page in nbrhd:
56     n+=1
57     try:
58         resp = urllib.request.urlopen(page)
59     except urllib.error.HTTPError as e:
60         if e.code in (... , 403, ...):
61             continue
62     except urllib.error.URLError as e:
63         continue
64     adj[n] = set()
65     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
66     for link in soup.find_all('a', href=True):
67         lnk = link['href']
68         if re.search("^http", lnk):
69             if lnk in nbrhd and lnk != page:
70                 t = nbrhd.index(lnk)
71                 adj[n].add(t)
72
73 #OK
74
75 for edges in adj:
76     adj[edges] = list(adj[edges])
77
78 #OK
79 itr = 10
80
81 vert = len(nbrhd)
82 auth = []
83 hub = []
84 for i in range(vert):
85     auth.append(i)
86     hub.append(i)
87
88 for itr in range(itr):
89
90     auth = [0 for a in auth]
91     for u in adj:

```

In [5]:

```

...: except urllib.error.HTTPError as e:
...:     ...
...:     adj[n] = set()
...:     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
...:     for link in soup.find_all('a', href=True):
...:         lnk = link['href']
...:         if re.search("^http", lnk):
...:             if lnk in nbrhd and lnk != page:
...:                 t = nbrhd.index(lnk)
...:                 adj[n].add(t)

```

In [6]:

```

...: adj = Dictionary (240 elements)
...: 
...: Key Type Size Value
...: 0 set 1 set object of builtins module
...: 1 set 1 set object of builtins module
...: 2 set 1 set object of builtins module
...: 3 set 1 set object of builtins module
...: 4 set 1 set object of builtins module
...: 5 set 1 set object of builtins module
...: 6 set 1 set object of builtins module
...: 7 set 1 set object of builtins module
...: R set 1 set object of builtins module

```

We use a dictionary of sets to avoid repetitions of edges. We then convert it to dictionary of lists for further use.

```

49     break
50
51 #OK
52
53 n = -1
54 adj = {}
55 for page in nbrhd:
56     n+=1
57     try:
58         resp = urllib.request.urlopen(page)
59     except urllib.error.HTTPError as e:
60         if e.code in (... , 403, ...):
61             continue
62     except urllib.error.URLError as e:
63         continue
64     adj[n] = set()
65     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
66     for link in soup.find_all('a', href=True):
67         lnk = link['href']
68         if re.search("^http", lnk):
69             if lnk in nbrhd and lnk != page:
70                 t = nbrhd.index(lnk)
71                 adj[n].add(t)
72
73 #OK
74
75 for edges in adj:
76     adj[edges] = list(adj[edges])
77
78 #OK
79 itr = 10
80
81 vert = len(nbrhd)
82 auth = []
83 hub = []
84 for i in range(vert):
85     auth.append(i)
86     hub.append(i)
87
88 for itr in range(itr):
89
90     auth = [0 for a in auth]
91     for u in adj:

```

In [5]:

```

...: except urllib.error.HTTPError as e:
...:     ...
...:     adj[n] = set()
...:     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
...:     for link in soup.find_all('a', href=True):
...:         lnk = link['href']
...:         if re.search("^http", lnk):
...:             if lnk in nbrhd and lnk != page:
...:                 t = nbrhd.index(lnk)
...:                 adj[n].add(t)

```

In [6]:

```

...: adj = Dictionary (240 elements)
...: 
...: Key Type Size Value
...: 0 list 11 [32, 33, 34, 35, 36, 37, 38, 39, 40, 30, ...]
...: 1 list 10 [41, 42, 43, 44, 45, 46, 47, 48, 49, 50]
...: 2 list 9 [51, 52, 53, 54, 55, 56, 57, 58, 59]
...: 3 list 10 [64, 65, 66, 67, 68, 69, 60, 61, 62, 63]
...: 4 list 2 [70, 71]
...: 5 list 7 [72, 73, 74, 75, 76, 77, 78]
...: 6 list 11 [79, 80, 81, 82, 83, 84, 85, 86, 87, 88, ...]
...: 7 list 10 [96, 97, 98, 99, 98, 91, 92, 93, 94, 95]
...: R list 7 [100, 101, 102, 103, 104, 105, 106]

```

This is our adjacency list.

We next calculate the authority weights and hub weights iteratively until they converge.

The screenshot shows the Spyder Python IDE interface. The code editor displays a script named `crawlinghits.py`. A variable explorer window is open, showing the state of variables. A pop-up window titled "auth - List (254 elements)" displays the authority weights for each node. The IPython console at the bottom shows the iterative update of the authority vector.

```

78
79 #itr = 10
80
81 vert = len(nbrhd)
82 auth = []
83 hub = []
84 for i in range(vert):
85     auth.append(1)
86     hub.append(1)
87
88 flag = 1
89 for itr in range(itr):
90 while flag == 1:
91     prevAuth = [i for i in auth]
92     auth = [0 for i in auth]
93     for u in adj:
94         for v in adj[u]:
95             auth[v] += hub[u]
96             norm = 0
97             for a in auth:
98                 norm += a**2
99             norm = norm**0.5
100            auth = [a/norm for a in auth]
101
102    prevhub = [i for i in hub]
103    hub = [0 for h in hub]
104    norm = 0
105    for u in adj:
106        for v in adj[u]:
107            hub[u] += auth[v]
108            norm += hub[u]**2
109    norm = norm**0.5
110    hub = [h/norm for h in hub]
111    flag = 0
112    for i,j in zip(auth, prevAuth):
113        if (i-j) != 0:
114            flag = 1
115    for i,j in zip(hub, prevhub):
116        if (i-j) != 0:
117            flag = 1
118
119
120

```

The screenshot shows the Spyder Python IDE interface again. The code editor displays the same script `crawlinghits.py`. A variable explorer window is open, showing the state of variables. A pop-up window titled "hub - List (254 elements)" displays the hub weights for each node. The IPython console at the bottom shows the iterative update of the hub vector.

```

78
79 #itr = 10
80
81 vert = len(nbrhd)
82 auth = []
83 hub = []
84 for i in range(vert):
85     auth.append(1)
86     hub.append(1)
87
88 flag = 1
89 for itr in range(itr):
90 while flag == 1:
91     prevAuth = [i for i in auth]
92     auth = [0 for i in auth]
93     for u in adj:
94         for v in adj[u]:
95             auth[v] += hub[u]
96             norm = 0
97             for a in auth:
98                 norm += a**2
99             norm = norm**0.5
100            auth = [a/norm for a in auth]
101
102    prevhub = [i for i in hub]
103    hub = [0 for h in hub]
104    norm = 0
105    for u in adj:
106        for v in adj[u]:
107            hub[u] += auth[v]
108            norm += hub[u]**2
109    norm = norm**0.5
110    hub = [h/norm for h in hub]
111    flag = 0
112    for i,j in zip(auth, prevAuth):
113        if (i-j) != 0:
114            flag = 1
115    for i,j in zip(hub, prevhub):
116        if (i-j) != 0:
117            flag = 1
118
119
120

```

Now we have the Authority weights and the Hub weights. So we can proceed to printing out the top weights.

N = 5

Spyder (Python 3.6)

```

120
121 #%%%
122 N = 5
123
124 au = {}
125 for i in range(vert):
126     au[i] = auth[i]
127
128 print("The top " + str(N) + " Authority weights are for the following websites:")
129 print("Authority Weight \t Hub Weight \t\t Website")
130 for n in range(N):
131     m = max(au, key=lambda key: au[key])
132     au.pop(m)
133     print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
134
135 #%%
136
137 hu = {}
138 for i in range(vert):
139     hu[i] = hub[i]
140
141 print("The top " + str(N) + " Hub weights are for the following websites:")
142 print("Hub Weight \t\t Authority Weight \t Website")
143 for n in range(N):
144     m = max(hu, key=lambda key: hu[key])
145     hu.pop(m)
146     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
147
148 #%%
149
150 for page in nbrhd:
151     print(page)
152     try:
153         resp = urllib.request.urlopen(page)
154     except urllib.error.HTTPError as e:
155         if e.code in (... , 403, ...):
156             print("\n\n")
157             continue
158     except urllib.error.URLError as e:
159         print("\n\n")
160         continue
161     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
162     try:

```

Run current cell and go to the next one (Shift+Enter)

Console 1/A

```

... print("The top " + str(N) + " Authority weights are for the following websites:")
... print("Authority Weight \t Hub Weight \t\t Website")
... for n in range(N):
...     m = max(au, key=lambda key: au[key])
...     au.pop(m)
...     print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
The top 5 Authority weights are for the following websites:
Authority Weight   Hub Weight   Website
0.327249411009963  0.0   https://www.wiley.com/privacy
0.327249411009963  0.0   https://hub.wiley.com/community/support/onlinelibrary
0.327249411009963  0.0   https://hub.wiley.com/community/exchanges/
0.2912738121126101  0.32965167902914083  https://onlinelibrary.wiley.com/search/advanced
0.2912738121126101  0.32965167902914083  https://onlinelibrary.wiley.com/search/advanced#citation

```

In [12]: hu = {}

```

... for i in range(vert):
...     hu[i] = hub[i]
...
...
... print("The top " + str(N) + " Hub weights are for the following websites:")
... print("Hub Weight \t\t Authority Weight \t Website")
... for n in range(N):
...     m = max(hu, key=lambda key: hu[key])
...     hu.pop(m)
...     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
The top 5 Hub weights are for the following websites:
Hub Weight   Authority Weight   Website
0.3614390185164664  0.0   https://onlinelibrary.wiley.com/journal/19321872
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/search/advanced
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/search/advanced#citation
0.32965167902914083  0.2912738121126101  https://ordering.onlinelibrary.wiley.com/subs.asp?ref=&doi=
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/terms-and-conditions

```

In [13]:

Permissions: RW End-of-lines: CRLF Encoding: ASCII Line: 148 Column: 1 Memory: 49 %

N = 10

Spyder (Python 3.6)

```

120
121 #%%%
122 N = 10
123
124 au = {}
125 for i in range(vert):
126     au[i] = auth[i]
127
128 print("The top " + str(N) + " Authority weights are for the following websites:")
129 print("Authority Weight \t Hub Weight \t\t Website")
130 for n in range(N):
131     m = max(au, key=lambda key: au[key])
132     au.pop(m)
133     print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
134
135 #%%
136
137 hu = {}
138 for i in range(vert):
139     hu[i] = hub[i]
140
141 print("The top " + str(N) + " Hub weights are for the following websites:")
142 print("Hub Weight \t\t Authority Weight \t Website")
143 for n in range(N):
144     m = max(hu, key=lambda key: hu[key])
145     hu.pop(m)
146     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
147
148 #%%
149
150 for page in nbrhd:
151     print(page)
152     try:
153         resp = urllib.request.urlopen(page)
154     except urllib.error.HTTPError as e:
155         if e.code in (... , 403, ...):
156             print("\n\n")
157             continue
158     except urllib.error.URLError as e:
159         print("\n\n")
160         continue
161     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
162     try:

```

Run current cell and go to the next one (Shift+Enter)

Console 1/A

```

... m = max(au, key=lambda key: au[key])
... au.pop(m)
... print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
The top 10 Authority weights are for the following websites:
Authority Weight   Hub Weight   Website
0.327249411009963  0.0   https://www.wiley.com/privacy
0.327249411009963  0.0   https://hub.wiley.com/community/support/onlinelibrary
0.327249411009963  0.0   https://hub.wiley.com/community/exchanges/
0.2912738121126101  0.32965167902914083  https://onlinelibrary.wiley.com/search/advanced
0.2912738121126101  0.32965167902914083  https://onlinelibrary.wiley.com/search/advanced#citation
0.2912738121126101  0.32965167902914083  https://ordering.onlinelibrary.wiley.com/subs.asp?ref=&doi=
0.2912738121126101  0.32965167902914083  https://onlinelibrary.wiley.com/terms-and-conditions
0.2912738121126101  0.32965167902914083  https://onlinelibrary.wiley.com/cookies
0.2912738121126101  0.32965167902914083  https://onlinelibrary.wiley.com/accessibility
0.2912738121126101  0.32965167902914083  https://onlinelibrary.wiley.com/agents
0.2912738121126101  0.32965167902914083  https://onlinelibrary.wiley.com/agents

```

In [14]: hu = {}

```

... for i in range(vert):
...     hu[i] = hub[i]
...
...
... print("The top " + str(N) + " Hub weights are for the following websites:")
... print("Hub Weight \t\t Authority Weight \t Website")
... for n in range(N):
...     m = max(hu, key=lambda key: hu[key])
...     hu.pop(m)
...     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
The top 10 Hub weights are for the following websites:
Hub Weight   Authority Weight   Website
0.3614390185164664  0.0   https://onlinelibrary.wiley.com/journal/19321872
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/search/advanced
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/search/advanced#citation
0.32965167902914083  0.2912738121126101  https://ordering.onlinelibrary.wiley.com/subs.asp?ref=&doi=
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/terms-and-conditions
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/cookies
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/accessibility
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/agents
0.32965167902914083  0.2912738121126101  https://onlinelibrary.wiley.com/advertisers
1.4e-322  1.4e-322  http://educationaldatamining.org/edm2019/

```

In [15]:

Permissions: RW End-of-lines: CRLF Encoding: ASCII Line: 148 Column: 1 Memory: 50 %

N = 20

```

Spyder (Python 3.6)
File Edit Search Source Run Debug Consoles Projects Tools View Help
Editor - C:\Users\Alfred Zane Rajan\Documents\Data Science\Sem 3\Implementation Project\crawlinghts.py IPython console
hw8.py crawlinghts.py In [15]: N = 20
116     if (i-j) != 0:
117         flag = 1
118
119
120
121 #**
122 N = 20
123
124 au = {}
125 for i in range(vert):
126     au[i] = auth[i]
127
128 print("The top " + str(N) + " Authority weights are for the following websites:")
129 print("Authority Weight \t Hub Weight \t\t Website")
130 for n in range(N):
131     m = max(au, key=lambda key: au[key])
132     au.pop(m)
133     print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
134
135 #**
136
137 hu = {}
138 for i in range(vert):
139     hu[i] = hub[i]
140
141 print("The top " + str(N) + " Hub weights are for the following websites:")
142 print("Hub Weight \t\t Authority Weight \t\t Website")
143 for n in range(N):
144     m = max(hu, key=lambda key: hu[key])
145     hu.pop(m)
146     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
147
148 #**
149
150 for page in nbrhd:
151     print(page)
152     try:
153         resp = urllib.request.urlopen(page)
154     except urllib.error.URLError as e:
155         if e.code in (... , 403, ...):
156             print("\n\n")
157             continue
158     except urllib.error.URLError as e:
In [15]: N = 20
...:
...: au = {}
...: for i in range(vert):
...:     au[i] = auth[i]
...:
...:
...: print("The top " + str(N) + " Authority weights are for the following websites:")
...: print("Authority Weight \t Hub Weight \t\t Website")
...: for n in range(N):
...:     m = max(au, key=lambda key: au[key])
...:     au.pop(m)
...:     print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
The top 20 Authority weights are for the following websites:
Authority Weight Hub Weight Website
0.327249411009963 0.0 https://www.wiley.com/privacy
0.327249411009963 0.0 https://hub.wiley.com/community/support/onlinelibrary
0.327249411009963 0.0 https://hub.wiley.com/community/exchanges/
0.2912738121126101 0.32965167902914083 https://onlinelibrary.wiley.com/search/advanced
0.2912738121126101 0.32965167902914083 https://onlinelibrary.wiley.com/search/advanced#citation
0.2912738121126101 0.32965167902914083 https://ordering.onlinelibrary.wiley.com/subs.asp?ref=&doi=
0.2912738121126101 0.32965167902914083 https://onlinelibrary.wiley.com/terms-and-conditions
0.2912738121126101 0.32965167902914083 https://onlinelibrary.wiley.com/cookies
0.2912738121126101 0.32965167902914083 https://onlinelibrary.wiley.com/accessibility
0.2912738121126101 0.32965167902914083 https://onlinelibrary.wiley.com/agents
0.2912738121126101 0.32965167902914083 https://onlinelibrary.wiley.com/advertisers
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/
1.4e-322 1.4e-322 http://sciences.ucf.edu/statistics/dms/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/registration/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/call-for-papers/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/keynote/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/workshops_tutorials/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/call-for-papers/#important_dates
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/registration/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/sponsors/
In [16]: hu = {}
...: for i in range(vert):
...:     hu[i] = hub[i]
...:
...:
...: print("The top " + str(N) + " Hub weights are for the following websites:")
...: print("Hub Weight \t\t Authority Weight \t\t Website")
...: for n in range(N):
...:     m = max(hu, key=lambda key: hu[key])
...:     hu.pop(m)
...:     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
The top 20 Hub weights are for the following websites:
Hub Weight Authority Weight Website
0.3614390185164664 0.0 https://onlinelibrary.wiley.com/journal/19321872
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/search/advanced
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/search/advanced#citation
0.32965167902914083 0.2912738121126101 https://ordering.onlinelibrary.wiley.com/subs.asp?ref=&doi=
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/terms-and-conditions
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/cookies
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/agents
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/advertisers
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/
1.4e-322 1.4e-322 http://sciences.ucf.edu/statistics/dms/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/registration/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/call-for-papers/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/keynote/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/workshops_tutorials/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/call-for-papers/#important_dates
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/registration/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/sponsors/
In [17]:

```

```

Spyder (Python 3.6)
File Edit Search Source Run Debug Consoles Projects Tools View Help
Editor - C:\Users\Alfred Zane Rajan\Documents\Data Science\Sem 3\Implementation Project\crawlinghts.py IPython console
hw8.py crawlinghts.py In [16]: hu = {}
...: for i in range(vert):
...:     hu[i] = hub[i]
...:
...:
...: print("The top " + str(N) + " Hub weights are for the following websites:")
...: print("Hub Weight \t\t Authority Weight \t\t Website")
...: for n in range(N):
...:     m = max(hu, key=lambda key: hu[key])
...:     hu.pop(m)
...:     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
The top 20 Hub weights are for the following websites:
Hub Weight Authority Weight Website
0.3614390185164664 0.0 https://onlinelibrary.wiley.com/journal/19321872
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/search/advanced
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/search/advanced#citation
0.32965167902914083 0.2912738121126101 https://ordering.onlinelibrary.wiley.com/subs.asp?ref=&doi=
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/terms-and-conditions
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/cookies
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/agents
0.32965167902914083 0.2912738121126101 https://onlinelibrary.wiley.com/advertisers
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/
1.4e-322 1.4e-322 http://sciences.ucf.edu/statistics/dms/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/registration/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/call-for-papers/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/keynote/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/workshops_tutorials/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/call-for-papers/#important_dates
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/registration/
1.4e-322 1.4e-322 http://educationaldatamining.org/edm2019/sponsors/
In [17]:

```

We can now witness the printing out of all the pages in the neighbourhood graph.

The screenshot shows the Spyder Python 3.6 IDE interface. The left pane displays the code for `crawlingNhts.py`, which prints authority and hub weights for various websites. The right pane shows the IPython console output, which includes several URLs and their descriptions related to data mining and web analysis.

```

127 print("The top " + str(N) + " Authority weights are for the following websites:")
128 print("Authority Weight \t Hub Weight \t\t Website")
129 for n in range(N):
130     m = max(au, key=lambda key: au[key])
131     au.pop(m)
132     print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
133
134 #%%%
135
136 hu = {}
137 for i in range(vert):
138     hu[i] = hub[i]
139
140 print("The top " + str(N) + " Hub weights are for the following websites:")
141 print("Hub Weight \t\t Authority Weight \t\t Website")
142 for n in range(N):
143     m = max(hu, key=lambda key: hu[key])
144     hu.pop(m)
145     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
146
147 #%%%
148
149 for page in nbrhd:
150     print(page)
151     try:
152         resp = urllib.request.urlopen(page)
153     except urllib.error.HTTPError as e:
154         if e.code in (... , 403, ...):
155             print("\n\n")
156         continue
157     except urllib.error.URLError as e:
158         print("\n\n")
159         continue
160     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
161     try:
162         print(soup.find('p').get_text())
163     except AttributeError as e:
164         print("\n\n")
165         continue
166     print("\n\n")
167
168

```

Output in IPython console:

- <http://www.statsoft.com/textbook/data-mining-techniques>
- https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm
- <https://www.techopedia.com/definition/1181/data-mining>
- <https://theappssolutions.com/blog/development/data-mining-guide/>
- <https://www.investopedia.com/terms/d/datamining.asp>
- <https://link.springer.com/journal/10618>

The screenshot shows the Spyder Python 3.6 IDE interface. The left pane displays the code for `crawlingNhts.py`. The right pane shows the IPython console output, which includes several URLs and their descriptions related to data mining and web analysis.

```

127 print("The top " + str(N) + " Authority weights are for the following websites:")
128 print("Authority Weight \t Hub Weight \t\t Website")
129 for n in range(N):
130     m = max(au, key=lambda key: au[key])
131     au.pop(m)
132     print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
133
134 #%%%
135
136 hu = {}
137 for i in range(vert):
138     hu[i] = hub[i]
139
140 print("The top " + str(N) + " Hub weights are for the following websites:")
141 print("Hub Weight \t\t Authority Weight \t\t Website")
142 for n in range(N):
143     m = max(hu, key=lambda key: hu[key])
144     hu.pop(m)
145     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
146
147 #%%%
148
149 for page in nbrhd:
150     print(page)
151     try:
152         resp = urllib.request.urlopen(page)
153     except urllib.error.HTTPError as e:
154         if e.code in (... , 403, ...):
155             print("\n\n")
156         continue
157     except urllib.error.URLError as e:
158         print("\n\n")
159         continue
160     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
161     try:
162         print(soup.find('p').get_text())
163     except AttributeError as e:
164         print("\n\n")
165         continue
166     print("\n\n")
167
168

```

Output in IPython console:

- <https://medium.com/data-mining-the-city>
- <https://extension.uscd.edu/courses-and-programs/data-mining-for-advanced-analytics>
- <https://www1.villanova.edu/villanova/business/centers/businessanalytics/studentresources/datamining.html>
- <https://support.crossref.org/hc/en-us/articles/215750183-Crossref-Text-and-Data-Mining-Services>
- <http://sciences.ucf.edu/statistics/dms/>
- <http://dml.cs.uiuc.edu/>
- <http://persistencedatamining.com/>

Above are just a couple samples. I will be dumping the whole output at the end of the file if size permits.

Now in similar fashion we find the top 10 authority and hub weights for other web search results:

Machine Learning:

K = 300

```

1 import urllib
2
3 from bs4 import BeautifulSoup
4 import urllib.request
5 import re
6
7 #%%%
8 #data = "DataMining.txt"
9 data = "MachineLearning.txt"
10
11 #%%%
12 k = 300
13
14 IN = open(data)
15 base = list()
16 line = IN.readline()
17 while line != '':
18     a = line.split()
19     base.append(a[0])
20     line = IN.readline()
21 IN.close()
22
23 nbrhd = [page for page in base]
24 n=0
25 b = 0
26 adj = {}
27 for page in base:
28     b+=1
29     try:
30         resp = urllib.request.urlopen(page)
31     except urllib.error.HTTPError as e:
32         continue
33     except urllib.error.URLError as e:
34         continue
35     adj[b] = set()
36     c = 0
37     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
38     for link in soup.find_all('a', href=True):
39         lnk = link['href']
40         if re.search("http", lnk):
41             if lnk not in nbrhd:
42                 nbrhd.append(lnk)
43                 n+=1

```

Variable explorer:

Name	Type	Size	Value
a	list	1	[https://www.geeksforgeeks.org/machine-learning/]
adj	dict	131	{0: set, 1: set, 2: set, 3: set, 4: set, 5: set, 6: set, 9: set, 10: set, 11: set, ...}
b	int	1	30
base	list	30	[https://arxiv.org, https://developers.google.com/machine-learning/ ...]
c	int	1	11
data	str	1	MachineLearning.txt
k	int	1	300
line	str	1	
lnk	str	1	http://www.netmediaeurope.com/
n	int	1	141
nbrhd	list	241	[https://arxiv.org, https://developers.google.com/machine-learning/ ...]
page	str	1	http://www.zdnet.co.kr/
t	int	1	143

IPython console:

```

In [1]: hu = {}
...: for i in range(vert):
...:     hu[i] = hub[i]
...
...
...: print("The top " + str(N) + " Hub weights are for the following websites:")
...: print("Hub Weight \t\t Authority Weight \t Website")
...: for n in range(N):
...:     m = max(hu, key=lambda key: hu[key])
...:     hu.pop(m)
...:     print(str(hub[m]) + '\t' + str(auth[m]) + '\t' + nbrhd[m])
...:
...:

```

N = 10

```

102 norm = norm**0.5
103 auth = [a/norm for a in hub]
104
105 prevHub = [i for i in hub]
106 hub = [0 for h in hub]
107 norm = 0
108 for u in adj:
109     for v in adj[u]:
110         hub[u] += auth[v]
111         norm += hub[u]**2
112 norm = norm**0.5
113 hub = [h/norm for h in hub]
114 flag = 0
115 for i,j in zip(auth, prevAuth):
116     if (i-j) != 0:
117         flag = 1
118 for i,j in zip(hub, prevHub):
119     if (i-j) != 0:
120         flag = 1
121
122
123
124 #%%%
125 N = 20
126
127 au = {}
128 for i in range(vert):
129     au[i] = auth[i]
130
131 print("The top " + str(N) + " Authority weights are for the following websites:")
132 print("Authority Weight \t Hub Weight \t\t Website")
133 for i in range(N):
134     m = max(au, key=lambda key: au[key])
135     au.pop(m)
136     print(str(auth[m]) + '\t' + str(hub[m]) + '\t' + nbrhd[m])
137
138 #%%%
139
140 hu = {}
141 for i in range(vert):
142     hu[i] = hub[i]
143
144 print("The top " + str(N) + " Hub weights are for the following websites:")

```

Variable explorer:

Name	Type	Size	Value
N	int	1	10

IPython console:

```

In [1]: m = max(hu, key=lambda key: hu[key])
...: hu.pop(m)
...: print(str(hub[m]) + '\t' + str(auth[m]) + '\t' + nbrhd[m])
...:
...:
The top 10 Authority weights are for the following websites:
Authority Weight    Hub Weight    Website
0.33241088415545814  0.29287037273601324  https://www.sas.com/en_us/home.html
0.33241088415545814  0.29287037273601324  https://www.sas.com/en_si/home.html
0.33241088415545814  0.29287037273601324  https://www.sas.com/es_ar/home.html
0.33241088415545814  0.29287037273601324  https://www.sas.com/en_au/home.html
0.33241088415545814  0.29287037273601324  https://www.sas.com/de_de/home.html
0.33241088415545814  0.29287037273601324  https://www.sas.com/en_bo/home.html
0.33241088415545814  0.29287037273601324  https://www.sas.com/pt_br/home.html
0.33241088415545814  0.29287037273601324  https://www.sas.com/en_ca/home.html
0.33241088415545814  0.29287037273601324  https://www.sas.com/es_cl/home.html
0.074376229842939  0.0  https://www.sas.com/profile/user/contact.htm?
locale=en_us&returnURL=https://www.sas.com/en_us/insights/analytics/machine-learning.html
The top 10 Hub weights are for the following websites:
Hub Weight    Authority Weight    Website
0.33767032958287746  0.0  https://www.sas.com/en_us/insights/analytics/machine-learning.html
0.33767032958287746  0.0  https://www.sas.com/en_us/insights/analytics/machine-learning.html
0.29287037273601324  0.33241088415545814  https://www.sas.com/us/home.html
0.29287037273601324  0.33241088415545814  https://www.sas.com/en_si/home.html
0.29287037273601324  0.33241088415545814  https://www.sas.com/en_au/home.html
0.29287037273601324  0.33241088415545814  https://www.sas.com/de_at/home.html
0.29287037273601324  0.33241088415545814  https://www.sas.com/en_be/home.html
0.29287037273601324  0.33241088415545814  https://www.sas.com/pt_br/home.html
0.29287037273601324  0.33241088415545814  https://www.sas.com/en_ca/home.html

```

Deep Learning:

This dataset unfortunately took too long to converge to zero change. So a tolerance of 10^{-5} was allowed.

```

for i,j in zip(auth, prevAuth):
    if abs(i-j) > 10**-5:
        flag = 1
for i,j in zip(hub, prevHub):
    if abs(i-j) > 10**-5:
        flag = 1

```

The screenshot shows the Spyder Python 3.6 IDE interface. The code editor displays the `crawlingHits.py` file with the above code. The variable explorer on the right shows several variables and their values:

Name	Type	Size	Value
line	str	1	
lnk	str	1	http://ai.stanford.edu/~amaas/data/data.zip
n	int	1	258
nbrhd	list	259	[https://en.wikipedia.org/wiki/Deep_learning , http://deeplearning.net ...]
norm	float	1	10.04263416512626
page	str	1	http://ufldl.stanford.edu/tutorial/StarterCode
prevAuth	list	259	[0.0, 8.921909008684167e-09, 1.0024474650393866e-09, 0.0, 0.0, 0.00194 ...]
prevHub	list	259	[3.10884843374937e-78, 7.9943832515889e-09, 4.439674762444535e-53, 2. ...]
t	int	1	257
u	int	1	258
v	int	1	257
vert	int	1	259

The IPython console shows the full script code again. The status bar at the bottom indicates permissions: RW, end-of-lines: CRLF, encoding: ASCII, line: 117, column: 29, memory: 49 %.

N = 20

The screenshot shows the Spyder Python 3.6 IDE interface with the `crawlingHits.py` file open. The variable explorer on the right shows the variable `j` with a value of `1.4955256124500531e-123`. The IPython console shows the top authority weights for various websites:

```

Authority Weight \t Hub Weight \t\t Website
...: print("Authority Weight \t Hub Weight \t\t Website")
...: for n in range(N):
...:     m = max(au, key=lambda key: au[key])
...:     au.pop(m)
...:     print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
The top 20 Authority weights are for the following websites:
Authority Weight   Hub Weight   Website
0.30043031348202903  0.2988735314229906  https://www.deeplearning.ai/
0.30043031348202903  0.2988735314229906  https://www.deeplearning.ai/ai-for-everyone/
0.30043031348202903  0.2988735314229906  https://www.deeplearning.ai/tensorflow-from-basics-to-
mastery/
0.30043031348202903  0.2988735314229906  https://www.deeplearning.ai/bootcamp/
0.30043031348202903  0.2988735314229906  https://www.deeplearning.ai/events/
0.30043031348202903  0.2988735314229906  https://www.deeplearning.ai/forum/
0.30043031348202903  0.2988735314229906  https://www.deeplearning.ai/blog/
0.30043031348202903  0.2988735314229906  https://www.deeplearning.ai/about-us/
0.30043031348202903  0.2988735314229906  https://www.deeplearning.ai/deep-learning-specialization/
0.2984884477857092  0.3083323441124202  https://www.deeplearning.ai/ai-career-program-for-
university-graduates/
0.2984884477857092  0.3083323441124202  https://www.deeplearning.ai/ai-career-program-for-
experienced-engineers/
0.001040495297804197  9.389043091494474e-08  https://www.coursera.org/specializations/deep-learning
0.001963783499158758e-08  9.479576039017513e-08  https://developer.nvidia.com/deep-learning-software
0.001942517773539380e-08  0.0002135390420113796  https://developer.nvidia.com/deep-learning
8.649904217435463e-05  7.750909258296402e-05  https://www.scientificamerican.com/store/subscribe/
scientific-american-magazine/
8.649904217435463e-05  7.750909258296402e-05  https://www.scientificamerican.com/
8.649904217435463e-05  7.750909258296402e-05  https://www.scientificamerican.com/store/shopping-cart/
8.649904217435463e-05  7.750909258296402e-05  https://www.scientificamerican.com/my-account/forgot-
password/
8.649904217435463e-05  7.750909258296402e-05  https://www.scientificamerican.com/my-account/register/
8.649904217435463e-05  7.750909258296402e-05  https://www.scientificamerican.com/page/newsletter-sign-up/

```

The IPython console shows the top authority weights for the first 20 websites. The status bar at the bottom indicates permissions: RW, end-of-lines: CRLF, encoding: ASCII, line: 152, column: 1, memory: 51 %.

```

110     for u in adj[u]:
111         hub[u] += auth[v]
112         norm += hub[u]**2
113         norm = norm**0.5
114         hub = [norm/norm for h in hub]
115         flag = 0
116         for i,j in zip(auth, prevAuth):
117             if abs(i-j) > 10**-5:
118                 flag = 1
119         for i,j in zip(hub, prevHub):
120             if abs(i-j) > 10**-5:
121                 flag = 1
122
123
124
125 #%%
126 N = 20
127
128 au = {}
129 for i in range(vert):
130     au[i] = auth[i]
131
132 print("The top " + str(N) + " Authority weights are for the following websites:")
133 print("Authority Weight \t Hub Weight \t Website")
134 for n in range(N):
135     m = max(au, key=lambda key: au[key])
136     au.pop(m)
137     print(str(auth[m]) + '\t' + str(hub[m]) + '\t' + nbrhd[m])
138
139 #%%
140
141 hu = {}
142 for i in range(vert):
143     hu[i] = hub[i]
144
145 print("The top " + str(N) + " Hub weights are for the following websites:")
146 print("Hub Weight \t Authority Weight \t Website")
147 for n in range(N):
148     m = max(hu, key=lambda key: hu[key])
149     hu.pop(m)
150     print(str(hub[m]) + '\t' + str(auth[m]) + '\t' + nbrhd[m])
151
152 #%%

```

Console 1/A

```

...: m = max(au, key=lambda key: au[key])
...: hu.pop(m)
...: print(str(hub[m]) + '\t' + str(auth[m]) + '\t' + nbrhd[m])
The top 20 Hub weights are for the following websites:
Hub Weight Authority Weight Website
0.3083323441124202 0.2994884477857092 https://www.deeplearning.ai/deep-learning-specialization/
0.3083323441124202 0.2994884477857092 https://www.deeplearning.ai/ai-career-program-for-
0.3083323441124202 0.2994884477857092 https://www.deeplearning.ai/ai-career-program-for-
0.3083323441124202 0.2994884477857092 https://www.deeplearning.ai/tensorflow-from-basics-to-
0.3083323441124202 0.2994884477857092 https://www.deeplearning.ai/experienced-engineers/
0.2988735314229906 0.30043031348202993 https://www.deeplearning.ai/
0.2988735314229906 0.30043031348202993 https://www.deeplearning.ai/ai-for-everyone/
0.2988735314229906 0.30043031348202993 https://www.deeplearning.ai/tensorflow-from-basics-to-
mastery/
0.2988735314229906 0.30043031348202993 https://www.deeplearning.ai/bootcamp/
0.2988735314229906 0.30043031348202993 https://www.deeplearning.ai/events/
0.2988735314229906 0.30043031348202993 https://www.deeplearning.ai/forums/
0.2988735314229906 0.30043031348202993 https://www.deeplearning.ai/blog/
0.2988735314229906 0.30043031348202993 https://www.deeplearning.ai/about-us/
0.09975399775023967 2.1265725619450834e-05 https://www.nvidia.com/en-us/deep-learning-ai/education/?nvidias-dis-dlclass=7258&id=7258#multi
0.00021356390420113706 2.1265725619450834e-05 https://www.nvidia.com/en-us/deep-learning-ai/education/?nvidias-dis-dlclass=7258&id=7258#multi
0.612227521530715e-05 8.641270614431484e-06 https://developer.nvidia.com/deep-learning-ive-into-deep-learning/
7.83791022592603e-05 8.640312388959062e-05 https://www.scientificamerican.com/observations/a-deep-
scientific-american-magazine/
7.750909258296402e-05 8.649904217435463e-05 https://www.scientificamerican.com/tech/
7.750909258296402e-05 8.649904217435463e-05 https://www.scientificamerican.com/store/subscribe/
password/

```

In [11]:

IPython console History log

Here we see that due to our tolerance, the weights did not get as low as in previous cases.

Big Data Analytics:

```

9 file = "MachineLearning.txt"
10 urlData = "Deeplearning.txt"
11 data = "BigDataAnalytics.txt"
12
13 #%%
14 k = 300
15
16 IN = open(data)
17 base = []
18 line = IN.readline()
19 while line != '':
20     a = line.split()
21     base.append(a[0])
22     line = IN.readline()
23 IN.close()
24
25 nbrhd = [page for page in base]
26 n = 30
27 b = 0
28 adj = {}
29 for page in base:
30     b+=1
31     try:
32         resp = urllib.request.urlopen(page)
33     except urllib.error.HTTPError as e:
34         continue
35     except urllib.error.URLError as e:
36         continue
37     adj[b] = set()
38     c = 0
39     soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))
40     for link in soup.findAll('a', href=True):
41         lnk = link['href']
42         if re.search("http", lnk):
43             if lnk in nbrhd:
44                 nbrhd.append(lnk)
45                 n+=1
46                 adj[b].add(n)
47             else:
48                 t = nbrhd.index(lnk) + 1
49                 adj[b].add(t)
50                 c+=1
51             if c > k/30:

```

Variable explorer

Name	Type	Size	Value
k	int	1	20
a	list	0	[]
adj	dict	249	{0:[32, 33, 34, 35, 36, ...], 1:[40, 41, 42, 43, 44, ...], 2:[49, 50, ...]}
au	dict	239	{0:0.0, 1:1.7174463736857553e-09, 2:8.061088049656194e-10, 3:0.0, 4:0.0 ...}
auth	list	259	[0.0, 8.714463736857553e-09, 8.061088049656194e-10, 0.0, 0.0, 0.01942 ...]
b	int	1	30
base	list	1	['https://searchbusinessanalytics.techtarget.com/definition/big-data-a ...']
c	int	1	2
data	str	1	BigDataAnalytics.txt
edges	int	1	258
flag	int	1	0

IPython console

```

In [14]: k = 300
...:
...: IN = open(data)
...: base = []
...: line = IN.readline()
...: while line != '':
...:     a = line.split()
...:     base.append(a[0])
...:     line = IN.readline()
...: IN.close()
...:
...: nbrhd = [page for page in base]
...: n = 30
...: b = 0
...: adj = {}
...: for page in base:

```

Permissions: RW End-of-lines: CRLF Encoding: ASCII Line: 54 Column: 1 Memory: 51 %

N = 10

Spyder (Python 3.6)

File Edit Search Source Run Debug Consoles Projects Tools View Help

Editor - C:\Users\Alfred Zane Rajan\Documents\Dataset\Sem 3\Implementation Project\crawlinghits.py

hw.py crawlinghits.py

```
113     norm += hub[u]**2
114     norm = norm**0.5
115     hub = [h/norm for h in hub]
116     flag = 0
117     for i,j in zip(auth, prevAuth):
118         if abs(i-j) > 10**-5:
119             if (i-j) != 0:
120                 flag = 1
121     for i,j in zip(hub, prevHub):
122         if abs(i-j) > 10**-5:
123             if (i-j) != 0:
124                 flag = 1
125
126
127
128 #END
129 N = 20
130
131 au = {}
132 for i in range(vert):
133     au[i] = auth[i]
134
135 print("The top " + str(N) + " Authority weights are for the following websites:")
136 print("Authority Weight \t Hub Weight \t Website")
137 for n in range(N):
138     m = max(au, key=lambda key: au[key])
139     au.pop(m)
140     print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
141
142 #END
143
144 hu = {}
145 for i in range(vert):
146     hu[i] = hub[i]
147
148 print("The top " + str(N) + " Hub weights are for the following websites:")
149 print("Hub Weight \t Authority Weight \t Website")
150 for n in range(N):
151     m = max(hu, key=lambda key: hu[key])
152     hu.pop(m)
153     print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
154
155 #END
```

...: print(str(auth[m]) +'\t'+ str(hub[m]) +'\t'+ nbrhd[m])
The top 10 Authority weights are for the following websites:
Authority Weight \t Website
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/category/machine-learning-ai/
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/category/postgraduate-
program-in-data-science-machine-learning/
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/category/internet-of-things/
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/category/careers-in-
analyticals/
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/category/hr-analytics/
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/category/tools_techniques/
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/category/big-data-analytics/
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/category/student-speak/
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/all-time-videos/

In [19]: hu = {}
...: for i in range(vert):
...: hu[i] = hub[i]
...
...
...: print("The top " + str(N) + " Hub weights are for the following websites:")
...: print("Hub Weight \t Authority Weight \t Website")
...: for n in range(N):
...: m = max(hu, key=lambda key: hu[key])
...: hu.pop(m)
...: print(str(hub[m]) +'\t'+ str(auth[m]) +'\t'+ nbrhd[m])
The top 10 Hub weights are for the following websites:
Hub Weight \t Authority Weight \t Website
0.30151134457776363 0.28618190351607126 https://analyticstraining.com/5-skills-need-know-become-big-data-analyst/
0.28618190351607126 0.30151134457776363 https://analyticstraining.com/category/machine-learning-ai/
0.28618190351607126 0.30151134457776363 https://analyticstraining.com/category/postgraduate-
program-in-data-science-machine-learning/
0.28618190351607126 0.30151134457776363 https://analyticstraining.com/category/internet-of-things/
0.28618190351607126 0.30151134457776363 https://analyticstraining.com/category/careers-in-
analyticals/
0.28618190351607126 0.30151134457776363 https://analyticstraining.com/category/hr-analytics/
0.28618190351607126 0.30151134457776363 https://analyticstraining.com/category/tools_techniques/
0.28618190351607126 0.30151134457776363 https://analyticstraining.com/category/big-data-analytics/
0.28618190351607126 0.30151134457776363 https://analyticstraining.com/category/student-speak/

IPython console History log Permissions: RW End-of-lines: CRLF Encoding: ASCII Line: 155 Column: 1 Memory: 48 %

Type here to search

Windows Taskbar icons: File Explorer, Edge, Google Chrome, Mail, Task View, Start, Task Manager.

System status bar: Permissions: RW End-of-lines: CRLF Encoding: ASCII Line: 155 Column: 1 Memory: 48 %

27-04-2019 17:45

Here I will dump the data too big and shabby to allow a clear report

“Data Mining” root set:

https://en.wikipedia.org/wiki/Data_mining
https://www.sas.com/en_us/insights/analytics/data-mining.html
<https://searchsqlserver.techtarget.com/definition/data-mining>
<http://www.statsoft.com/textbook/data-mining-techniques>
https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm
<https://www.techopedia.com/definition/1181/data-mining>
<https://theappsolutions.com/blog/development/data-mining-guide/>
<https://www.investopedia.com/terms/d/datamining.asp>
<https://link.springer.com/journal/10618>
<https://datafloq.com/read/data-mining-techniques-create-business-value/121>
<http://dmg.org/>
<https://www3.cs.stonybrook.edu/~cse634/>
<https://scpd.stanford.edu/public/category/courseCategoryCertificateProfile.do%3Fmethod%3Dload%26certificateId%3D1209602>
<https://onlinelibrary.wiley.com/journal/19321872>
<https://orange.biolab.si/training/>
<https://neilpatel.com/blog/data-mining/>
<https://connect.informs.org/data-mining/home>
<http://educationaldatamining.org/edm2019/>
<https://rapidminer.com/data-mining-tools-try-rapidminer/>
https://store.steampowered.com/app/993630/Data_mining_4/
<http://guidetodatamining.com/>
<https://inhomelandsecurity.com/fake-news-data-mining/>
<https://thenextweb.com/gaming/2019/03/15/gamers-epic-games-rift-data-mining-tencent/>
<https://medium.com/data-mining-the-city>
<https://extension.ucsd.edu/courses-and-programs/data-mining-for-advanced-analytics>
<https://www1.villanova.edu/villanova/business/centers/businessanalytics/studentresources/datamining.html>
<https://support.crossref.org/hc/en-us/articles/215750183-Crossref-Text-and-Data-Mining-Services>
<http://sciences.ucf.edu/statistics/dms/>
<http://dm1.cs.uiuc.edu/>
<http://persistencedatamining.com/>

“Machine Learning” root set:

<https://arxiv.org>
<https://developers.google.com/machine-learning/crash-course/>
https://en.wikipedia.org/wiki/Machine_learning
<https://cloud.google.com/products/ai/>
<https://xkcd.com/1838/>
<https://aws.amazon.com/machine-learning/>
<https://www.wired.com/tag/machine-learning/>

<https://developer.apple.com/machine-learning/>
<https://towardsdatascience.com/machine-learning/home>
<https://searchenterpriseai.techtarget.com/definition/machine-learning-ML>
<https://azure.microsoft.com/en-us/services/machine-learning-service/>
<https://www.brookings.edu/research/what-is-machine-learning/>
https://www.sas.com/en_us/insights/analytics/machine-learning.html
<https://www.zdnet.com/article/what-is-machine-learning-everything-you-need-to-know/>
<https://www.forbes.com/.../20-examples-of-machine-learning-used-in-customer-experi>
<https://emerj.com>
<https://www.edx.org/learn/machine-learning>
<https://www.mathworks.com/discovery/machine-learning.html>
<https://www.youtube.com/watch?v=ukzFI9rgwfU>
<https://www.ml.cmu.edu/>
<https://royalsociety.org/topics-policy/projects/machine-learning/>
<https://www.sap.com/products/leonardo/machine-learning.html>
<https://www.elastic.co/products/stack/machine-learning>
<https://www.udemy.com/topic/machine-learning/>
<https://medium.com/machine-learning.../why-machine-learning-matters-6164faf1df12>
https://www.sas.com/en_us/insights/analytics/machine-learning.html
<https://www.expertsystem.com/machine-learning-definition/>
<https://www.coursera.org/learn/machine-learning>
<https://www.udacity.com/course/intro-to-machine-learning--ud120>
<https://www.geeksforgeeks.org/machine-learning/>

“Deep Learning” root set:

[['https://en.wikipedia.org/wiki/Deep_learning'](https://en.wikipedia.org/wiki/Deep_learning),
['http://deeplearning.net/'](http://deeplearning.net/),
['https://www.mathworks.com/discovery/deep-learning.html'](https://www.mathworks.com/discovery/deep-learning.html),
['https://www.udacity.com/course/intro-to-tensorflow-for-deep-learning--ud187'](https://www.udacity.com/course/intro-to-tensorflow-for-deep-learning--ud187),
['https://machinelearningmastery.com/what-is-deep-learning/'](https://machinelearningmastery.com/what-is-deep-learning/),
['https://developer.nvidia.com/deep-learning'](https://developer.nvidia.com/deep-learning),
['https://www.forbes.com/sites/bernardmarr/2018/10/01/what-is-deep-learning-ai-a-simple-guide-with-8-practical-examples/'](https://www.forbes.com/sites/bernardmarr/2018/10/01/what-is-deep-learning-ai-a-simple-guide-with-8-practical-examples/),
['https://skymind.ai/wiki/neural-network'](https://skymind.ai/wiki/neural-network),
['http://neuralnetworksanddeeplearning.com/chap6.html'](http://neuralnetworksanddeeplearning.com/chap6.html),
['https://www.investopedia.com/terms/d/deep-learning.asp'](https://www.investopedia.com/terms/d/deep-learning.asp),
['https://www.technologyreview.com/s/513696/deep-learning/'](https://www.technologyreview.com/s/513696/deep-learning/),
['https://www.coursera.org/specializations/deep-learning'](https://www.coursera.org/specializations/deep-learning),
['https://www.deeplearningbook.org/'](https://www.deeplearningbook.org/),
['https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network'](https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network),
['https://www.sas.com/en_us/insights/analytics/deep-learning.html'](https://www.sas.com/en_us/insights/analytics/deep-learning.html),
['https://medium.com/tensorflow/mit-introduction-to-deep-learning-4a6f8dde1f0c'](https://medium.com/tensorflow/mit-introduction-to-deep-learning-4a6f8dde1f0c),
['https://www.edx.org/professional-certificate/ibm-deep-learning'](https://www.edx.org/professional-certificate/ibm-deep-learning),

'<https://www.udemy.com/deeplearning/>',
'<https://deeplearning.mit.edu/>',
'<https://aws.amazon.com/deep-learning/>',
'<https://www.ibm.com/cloud/deep-learning>',
'<https://blogs.scientificamerican.com/observations/a-deep-dive-into-deep-learning/>',
'<https://www.datacamp.com/courses/deep-learning-in-python>',
'<https://www.deeplearning.ai/>',
'<https://software.intel.com/en-us/ai/courses/deep-learning>',
'<https://www.nature.com/articles/s41576-019-0122-6>',
'<http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science深深学习.html>',
'<https://towardsdatascience.com/detecting-malaria-with-deep-learning-9e45c1e34b60>',
'<https://course.fast.ai/>',
'<http://deeplearning.stanford.edu/tutorial/>',
'<https://arxiv.org/list/cs.LG/recent>',
'<http://adsabs.harvard.edu/abs/2015Natur.521..436L>',
'<http://ieeexplore.ieee.org/document/6248110/>',
'https://www.cs.toronto.edu/~kriz/imagenet_classification_with_deep_convolutional.pdf',
'<https://techcrunch.com/2017/05/24/alphago-beats-planets-best-human-go-player-ke-jie/amp/>',
'<http://adsabs.harvard.edu/abs/1996Natur.381..607O>',
'<http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf>',

'<http://sanghv.com/download/soft/machine%20learning,%20artificial%20intelligence,%20mathematics%20ebooks/ML/learning%20deep%20architectures%20for%20AI%20%282009%29.pdf>',
'http://www.scholarpedia.org/article/Deep_Learning',
'<http://adsabs.harvard.edu/abs/2009SchpJ...4.5947H>',
'<http://deeplearning.net/reading-list/>',
'<http://deeplearning.net/reading-list/tutorials/>',
'http://deeplearning.net/software_links/',
'<http://deeplearning.net/blog/>',
'<http://deeplearning.net/demos/>',
'<http://deeplearning.net/datasets/>',
'<http://deeplearning.net/events/>',
'<http://deeplearning.net/bibliography/>',
'<http://deeplearning.net/deep-learning-research-groups-and-labs/>',
'https://www.mathworks.com/matlabcentral/?s_tid=gn_mlc',
'<https://www.mathworks.com/matlabcentral/fileexchange/60659-deep-learning-in-11-lines-of-matlab-code>',
'<https://www.mathworks.com/help/vision/examples/object-detection-using-deep-learning.html>',
'https://www.mathworks.com/matlabcentral/fileexchange/?s_tid=hp_ff_p_fx',
'https://www.mathworks.com/downloads/web_downloads/?s_iid=hp_ff_t_downloads',
'https://www.mathworks.com/store?s_iid=hp_ff_t_buy',
'https://www.mathworks.com/help/?s_tid=hp_ff_l_doc',
'https://www.mathworks.com/support/learn-with-matlab-tutorials.html?s_tid=hp_ff_l_tutorials',
'https://www.mathworks.com/help/examples.html?s_tid=hp_ff_l_examples',

'https://www.mathworks.com/support/install-matlab.html?s_tid=hp_ff_s_install',
'https://auth.udacity.com/sign-in?next=https://classroom.udacity.com/authenticated',
'https://auth.udacity.com/sign-up?next=https://classroom.udacity.com/authenticated',
'https://www.udacity.com/tech-requirements',
'https://www.udacity.com/school-of-ai',
'https://www.udacity.com/school-of-autonomous-systems',
'https://www.udacity.com/school-of-business',
'https://www.udacity.com/school-of-data-science',
'https://www.udacity.com/school-of-programming',
'https://career-resource-center.udacity.com',
'https://news.developer.nvidia.com/',
'https://devblogs.nvidia.com/',
'https://devtalk.nvidia.com/',
'https://developer.nvidia.com/deep-learning-software',
'https://developer.nvidia.com/tensorrt',
'https://www.nvidia.com/en-us/deep-learning-ai/education/?ncid=so-dis-dldlwsd1-72342',
'https://courses.nvidia.com/courses/course-v1:DLI+C-FX-01+V2/about?ncid=so-dis-dldlwsd1-72346',
'https://www.nvidia.com/en-us/deep-learning-ai/education/?ncid=so-dis-dldlwsd1-72528#dlti',
'http://www.nvidia.com/object/legal_info.html',
'https://www.forbes.com/worlds-billionaires/',
'https://www.forbes.com/billionaires/',
'https://www.forbes.com/forbes-400/',
'https://www.forbes.com/self-made-women/',
'https://www.forbes.com/china-billionaires/',
'https://www.forbes.com/india-billionaires/',
'https://www.forbes.com/indonesia-billionaires/',
'https://www.forbes.com/korea-billionaires/list/',
'https://www.forbes.com/thailand-billionaires/list/ ',
'https://www.forbes.com/japan-billionaires/list/',
'https://www.exxactcorp.com/Deep-Learning-NVIDIA-GPU-Solutions?utm_source=neuralnetworksanddeeplearning&utm_medium=banner&utm_campaign=sponsors',
'http://gsquaredcapital.com/',
'http://www.tineye.com',
'http://www.visionsmarts.com',
'https://twitter.com/michael_nielsen',
'https://github.com/mnielsen/neural-networks-and-deep-learning',
'http://eepurl.com/0Xxjb',
'http://www.deeplearningbook.org/',
'http://cognitivemedium.com',
'http://michaelnielsen.org',
'https://www.investopedia.com/dictionary/',
'https://www.investopedia.com/university/beginner/',
'https://www.investopedia.com/articles/markets/101415/4-best-sp-500-index-funds.asp',

'https://www.investopedia.com/insights/worlds-top-economies/','
'https://www.investopedia.com/university/stocks/','
'https://www.investopedia.com/university/options/','
'https://www.investopedia.com/university/economics/','
'https://www.investopedia.com/markets/','
'https://www.investopedia.com/markets/watchlist/','
'https://www.investopedia.com/markets/stocks/aapl/','

'https://subscribe.technologyreview.com/ecom/mtr/app/live/subcustserv?pagemode=start&org=MTR&publ=TR&php=Y','
'https://go.technologyreview.com/newsletters/the-download/','
'https://go.technologyreview.com/newsletters/chain-letter/','
'https://go.technologyreview.com/newsletters/the-algorithm/','
'https://go.technologyreview.com/newsletters/fwd-economy','
'https://go.technologyreview.com/newsletters/the-airlock','
'https://go.technologyreview.com/newsletters','
'https://events.technologyreview.com/emtech/','
'https://events.technologyreview.com/emtech/digital/','
'https://events.technologyreview.com/emtech/next/','
'https://events.technologyreview.com/presents/business-of-blockchain/2019/','
'https://learner.coursera.help/hc/','
'https://itunes.apple.com/app/apple-store/id736535961?pt=2334150&ct=Coursera%20Web%20Promo%20Banner&mt=8','
'http://play.google.com/store/apps/details?id=org.coursera.android','
'https://about.coursera.org/','
'https://about.coursera.org/leadership','
'https://www.coursera.community','
'https://www.coursera.org/about/partners','
'https://building.coursera.org/developer-program/','
'https://www.coursera.community/beta-tester-news-announcements-20/join-the-coursera-beta-tester-community-47','
'https://www.coursera.community/gtc-news-announcements-17/join-the-coursera-global-translator-community-45','
'https://blog.coursera.org','
'https://www.amazon.com/Deep-Learning-Adaptive-Computation-Machine/dp/0262035618/ref=sr_1_1?ie=UTF8&qid=1472485235&sr=8-1&keywords=deep+learning+book','
'https://groups.google.com/forum/#!forum/deeplearningbook','
'https://github.com/goodfeli/dlbook_notation',

'https://docs.google.com/document/d/1ABlp7FluwZ0B82_fjNOFVQ2uOZkfuF8elbofhZmNXag/edit?usp=sharing','
'https://searchenterpriseai.techtarget.com','
'https://users.techtarget.com/registration/searchEnterpriseAI/LoginRegister.page',

'https://users.techtarget.com/registration/searchEnterpriseAI/Register.page',
'https://www.techtarget.com/network',
'https://searchenterpriseai.techtarget.com/info/news',
'https://searchenterpriseai.techtarget.com/features',
'https://searchenterpriseai.techtarget.com/tips',
'https://searchenterpriseai.techtarget.com/answers',
'https://whatis.techtarget.com/resources/Buyers-Guides',
'https://www.sas.com/en_us/home.html',

'https://www.sas.com/profile/user/contact.htm?locale=en_us&returnURL=https://www.sas.com/en_us/insights/analytics/deep-learning.html',
'https://www.sas.com/en_si/home.html',
'https://www.sas.com/es_ar/home.html',
'https://www.sas.com/en_au/home.html',
'https://www.sas.com/de_at/home.html',
'https://www.sas.com/en_be/home.html',
'https://www.sas.com/pt_br/home.html',
'https://www.sas.com/en_ca/home.html',
'https://www.sas.com/es_cl/home.html',
'https://deeplearning.mit.edu',
'https://github.com/lexfridman/mit-deep-learning',
'https://selfdrivingcars.mit.edu/deeptraffic/',
'https://www.youtube.com/watch?v=1L0TKZQcUtA&list=PLrAXtmErZgOeiKm4sgNOknGvNjby9efdf',
'https://twitter.com/lexfridman',
'https://www.youtube.com/channel/UCSHZKyawb77ixDdsGog4iWA',
'https://lexfridman.us2.list-manage.com/subscribe?u=2bd2649a06183adee599340a1&id=0c5b4f2d63',
'https://lexfridman.com',
'https://aws.amazon.com/?nc2=h_lg',
'https://console.aws.amazon.com/support/home?nc2=h_ql_cu',
'https://portal.aws.amazon.com/gp/aws/developer/registration/index.html?nc2=h_ct&src=default',
'https://docs.aws.amazon.com/index.html?nc2=h_ql_doc',
'https://aws.amazon.com/marketplace/?nc2=h_ql_mp',
'https://aws.amazon.com/ar/?nc1=h_ls',
'https://aws.amazon.com/id/?nc1=h_ls',
'https://aws.amazon.com/de/?nc1=h_ls',
'https://aws.amazon.com/?nc1=h_ls',
'https://aws.amazon.com/es/?nc1=h_ls',
'https://aws.amazon.com/fr/?nc1=h_ls',
'https://www.ibm.com/cloud/why-ibm',
'https://www.ibm.com/cloud/products',
'https://www.ibm.com/cloud/solutions',
'https://www.ibm.com/cloud/garage',
'https://www.ibm.com/cloud/pricing',
'https://www.ibm.com/cloud/partners',

'https://www.ibm.com/blogs/bluemix/','
'https://cloud.ibm.com/docs','
'https://www.ibm.com/cloud/support','
'https://cloud.ibm.com/registration','

'https://dataplatform.cloud.ibm.com/registration/stepone?apps=data_science_experience&context=wdp',
'https://www.scientificamerican.com/store/subscribe/scientific-american-magazine/','
'https://www.scientificamerican.com/','
'https://www.scientificamerican.com/store/shopping-cart/','
'https://www.scientificamerican.com/my-account/forgot-password/','
'https://www.scientificamerican.com/my-account/register/','
'https://www.scientificamerican.com/page/newsletter-sign-up/','
'https://www.scientificamerican.com/the-sciences/','
'https://www.scientificamerican.com/mind/','
'https://www.scientificamerican.com/health/','
'https://www.scientificamerican.com/tech/','
'https://www.datacamp.com/community/podcast',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=1',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=2',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=3',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=4',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=5',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=6',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=7',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=8',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=9',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=10',
'https://www.deeplearning.ai/deep-learning-specialization/','
'https://www.deeplearning.ai/ai-for-everyone/','
'https://www.deeplearning.ai/tensorflow-from-basics-to-mastery/','
'https://www.deeplearning.ai/bootcamp/','
'https://www.deeplearning.ai/ai-career-program-for-university-graduates/','
'https://www.deeplearning.ai/ai-career-program-for-experienced-engineers/','

'https://www.deeplearning.ai/events/','
'https://www.deeplearning.ai/forums/','
'https://www.deeplearning.ai/blog/','
'https://www.deeplearning.ai/about-us/','
'https://software.intel.com/en-us/user/login?destination=node/740582','
'https://software.intel.com/registration?lang=en-us','
'https://www.intel.ai/investing-in-the-pytorch-developer-community/','
'http://en.paddlepaddle.org/en','
'https://www.intel.ai/ngraph/','
'https://01.org/mkl-dnn/overview','
'https://www.intel.ai/plaidML/','
'https://www.intel.ai/hardware/','
'https://software.seek.intel.com/AIWebinarSeries?registration_source=IDZ','
'https://software.intel.com/en-us/ai/docs?search_api_views_fulltext=¤t_page=0&value=20780','
'https://www.intel.ai/publications/','
'https://www.nature.com/my-account/alerts/subscribe-journal?list-id=16','
'https://mts-nrg.nature.com/cgi-bin/main.plex',

'https://idp.nature.com/authorize/natureuser?client_id=grover&redirect_uri=https%3A%2F%2Fwww.nature.com%2Farticles%2Fs41576-019-0122-6',
'http://orcid.org/0000-0001-9579-2909',
'http://orcid.org/0000-0002-2419-1943',

'https://idp.nature.com/login/natureuser?client_id=grover&redirect_uri=http://www.nature.com/articles/s41576-019-0122-6',
'https://idp-saml-nature-federated-login.live.cf.public.nature.com/saml/login?idp=https://idp.eduserv.org.uk/openathens&targetUrl=http://www.nature.com/articles/s41576-019-0122-6',
'https://idp.nature.com/login/federated?redirect_uri=http://www.nature.com/articles/s41576-019-0122-6',
'https://kundajelab.github.io/dragonn/tutorials.html',
'https://www.kaggle.com/sudalairajkumar/submitting-solutions-to-kaggle-competitions',
'https://keras.io/',
'https://en.wikipedia.org/wiki/Feedforward_neural_network',
'http://h2o.ai/resources',
'https://htmlpreview.github.io/?https://github.com/ledell/sldm4-h2o/blob/master/sldm4-deeplearning-h2o.html',
'https://www.h2o.ai/blog/h2o-tensorflow-on-aws-gpu/',
'https://github.com/h2oai/sparkling-water/blob/master/py/examples/notebooks/TensorFlowDeepLearning.ipynb',
'https://www.youtube.com/watch?v=zGdXaRug7LI/','
'https://www.youtube.com/watch?v=LM255qs8Zsk/','
'https://www.youtube.com/watch?v=KWdkVoKJG3U/','
'https://en.wikipedia.org/wiki/Modulo_operation',

'<http://www.idg.pl/mirrors/CRAN/web/packages/gbm/vignettes/gbm.pdf>',
'http://premolab.ru/pub_files/pub88/qhkDNEyp8.pdf',
'<https://github.com/fastai/course-v3>',
'<http://www.fast.ai>',
'<http://docs.fast.ai>',
'<https://github.com/fastai/fastai/blob/master/README.md>',
'<http://course18.fast.ai>',
'<https://forums.fast.ai/t/recommended-python-learning-resources/26888>',
'<http://www.fast.ai/2017/11/16/what-you-need/>',
'<https://www.fast.ai/2019/01/24/course-v3/>',
'<https://docs.microsoft.com/azure/virtual-machine-scale-sets/virtual-machine-scale-sets-use-low-priority>',
'<https://forums.fast.ai/>',
'<https://www.discourse.org/about>',
'<http://ufldl.stanford.edu/tutorial/>',
'<http://ufldl.stanford.edu/tutorial/StarterCode>']

“Big Data Analytics” root set:

<https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>
<https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>
https://www.sas.com/en_us/insights/analytics/big-data-analytics.html
<https://www.ibm.com/analytics/hadoop/big-data-analytics>
<https://www.qubole.com/big-data-analytics/>
<https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>
<https://www.ngdata.com/what-is-big-data-analytics/>
<https://analyticstraining.com/5-skills-need-know-become-big-data-analyst/>
https://en.wikipedia.org/wiki/Big_data
<https://mixpanel.com/blog/2018/12/11/big-data-analytics/>
<https://www.techopedia.com/definition/28659/big-data-analytics>
https://www.webopedia.com/TERM/B/big_data_analytics.html
<https://www.marutitech.com/big-data-analytics-need-business/>
<https://www.sisense.com/product/big-data-analytics/>
<https://databricks.com/glossary/big-data-analytics>
<https://www.teradata.com/Outcomes/Big-Data>
<https://www.edx.org/course/big-data-analytics-adelaidex-analyticsx>
<https://bigdataanalytics.mit.edu/>
<https://www.datamation.com/big-data/big-data-analytics.html>
<https://www.tibco.com/products/tibco-spotfire/big-data-analytics>
<https://www.informationbuilders.com/big-data-analytics>
<https://bdataanalytics.biomedcentral.com/>
<https://www.eweek.com/big-data-and-analytics>
<https://www.edureka.co/big-data-and-analytics>
<https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/>

<https://www.alteryx.com/solutions/analytics-need/advanced-analytics/database-analytics>
<https://www.extension.harvard.edu/course-catalog/courses/crn/15499>
https://www.tutorialspoint.com/big_data_analytics/
<https://www.linkedin.com/jobs/big-data-analytics-jobs>
<https://www.zdnet.com/topic/big-data/>
<https://www.informatica.com/products/big-data.html>
<https://www.ibmbigdatahub.com/>
<https://www.informationweek.com/big-data-analytics.asp>
https://www.splunk.com/en_us/solutions/solution-areas/big-data.html
<https://cloud.google.com/products/big-data/>
<https://cloud.google.com/solutions/big-data/>
<https://www.talend.com/solutions/information-technology/big-data-analytics/>
<https://pivot3.com/solution/big-data-analytics/>

"Data Mining" neighbourhood graph (truncated by the output console):

<https://inhomelandsecurity.com/category/border-wall/>

In Homeland Security's coverage of President Donald Trump's signature campaign promise: The Border Wall.

<https://inhomelandsecurity.com/category/homeland-security/>

From natural disasters to international terrorism threats, the United States has become increasingly vulnerable to issues regarding national security. Recently, strategic military and cyber-hacking moves by countries around the world have highlighted the importance of staying up-to-date with the most recent homeland security news coverage. Accessing the most recent updates has never been more vital, as threats made to the United States are a serious concern to our country's citizens.

<https://inhomelandsecurity.com/category/terrorism-threats/>

From nuclear threats made by North Korea to domestic terrorist attacks, terrorism news has become a primary point of interest in United States affairs. Recently, strategic moves by foreign governments and terrorist groups around the world have highlighted the importance of staying up-to-date with the most recent updates. Accessing the most current terrorism news continues to be vital to concerned citizens, as domestic and foreign attacks on the United States and other allied nations remain a serious concern and threat.

<https://inhomelandsecurity.com/category/politics-government/>

From international scandals to changes in foreign policy, the United States government and other domestic affairs have become a primary point of interest to American constituents. Recent shifts in political positions have highlighted the importance of staying up-to-date with the most recent news coverage. Accessing the most current U.S. politics news has never been more vital, as important events occurring within all branches of United States government are a serious concern to Americans and the rest of the developed world.

<https://inhomelandsecurity.com/>

<https://index.co/>

Using our super flexible search.

<https://tq.co/>

We build curated spaces for fast-growing tech companies and talent.

https://deals.thenextweb.com/?utm_source=thenextweb.com&utm_medium=referral&utm_campaign=navbar

Get exclusive coverage to the world's top publisher sites through the StackCommerce network.

<https://answers.thenextweb.com/>

Live AMAs with the world's most influential thought leaders in business, design and technology.

https://thenextweb.com/conference/?utm_source=thenextweb.com&utm_medium=referral&utm_campaign=nav-link

We're at a crossroads, and it's time to choose a direction.

<https://www.facebook.com/thenextweb>

These processors are bringing laptops ever closer to desktop level performance

<https://myextension.ucsd.edu/>

The portal for managing your UC San Diego Extension account

<https://ucsdxextension.blackboard.com>

<https://instructorlink.ucsd.edu/>

In order to login, you will use the email address we have on record for you.

<https://extcart.ucsd.edu/index.cfm?vAction=view>

<http://osha.ucsd.edu/>

UCSD's Occupational Safety and Health (OSHA) Department is focused on providing high-quality training services for health and safety professionals. UCSD's OSHA Department currently offers:

<http://academicconnections.ucsd.edu/>

UC San Diego Academic Connections 2018

<http://gels.ucsd.edu>

UC San Diego - A World Class Research University

<https://apply.ucsd.edu/extension/startapp.php?subjectcode=CERT&coursenumber=DMAA>

<https://extension.ucsd.edu/courses-and-programs/statistics-for-data-analytics>

<http://www.forthegreatergreat.com/>

<https://mynova.villanova.edu/>

<http://outlook.villanova.edu/>

<http://elearning.villanova.edu/>

https://secure.touchnet.com/C21740_ustores/web/store_main.jsp?STOREID=4&SINGLESTORE=true

<http://webmail.villanova.edu/>

Welcome. You have reached Villanova University's Webmail Login Form. This login is for current VU students and employees. If you have forgotten your UserID or password, you may use the Self Service Reset Password Form.

<http://library.villanova.edu>

Due to the University closure, the Library service desk will be closed on Friday, March 1st.

<http://www.villanova.com>

<https://www1.villanova.edu/villanova/business/centers/businessanalytics/studentresources/datamining/registration.html>

Developing business leaders for a better world.®

The Villanova School of Business is a premier educational institution that offers undergraduate programs, graduate programs and executive education. Founded in 1922, VSB is known for academic rigor; creativity and innovation; hands-on and service learning opportunities; a firm grounding in ethics and an applied education that prepares students to become outstanding leaders and global citizens within the ever-changing, complex, and fast-paced world of business.

<https://www1.villanova.edu/content/dam/villanova/VSB/centers/cba/WestCap1-1.xlsx>

PK ! A7,ln [Content_Types].xml ¢(-TÉnÂ0%Wê?D¾V‰j‡¤¤[§è˜x'X\$¶å(ü)'fQU±Á%Ql¡[

<http://www.crossref.org>

Crossref makes research outputs easy to find, cite, link, and assess. We're a not-for-profit membership organization that exists to make scholarly communications better.

<http://api.crossref.org/>

GitHub is home to over 36 million developers working together to host and review code, manage projects, and build software together.

<https://youtu.be/LBYgq6jPoyk>

Loading...

http://www.crossref.org/08downloads/webinars/tdm_walkthrough_032415.mp4

<http://support.crossref.org/hc/en-us/articles/215750263>

Q: Do publishers have to pay to participate in Crossref text and data mining?

<http://support.crossref.org/hc/en-us/articles/215750203>

Q: Do I have to pay to use the CrossRef REST API?

<https://www.facebook.com/share.php?title=Crossref+Text+and+Data+Mining+Services&u=https%3A%2F%2Fsupport.crossref.org%2Fhc%2Fen-us%2Farticles%2F215750183-Crossref-Text-and-Data-Mining-Services>

Forgot account?

<https://twitter.com/share?lang=en&text=Crossref+Text+and+Data+Mining+Services&url=https%3A%2F%2Fsupport.crossref.org%2Fhc%2Fen-us%2Farticles%2F215750183-Crossref-Text-and-Data-Mining-Services>

Remember me

[Forgot password?](#)

<https://www.linkedin.com/shareArticle?mini=true&source=Crossref&title=Crossref+Text+and+Data+Mining+Services&url=https%3A%2F%2Fsupport.crossref.org%2Fhc%2Fen-us%2Farticles%2F215750183-Crossref-Text-and-Data-Mining-Services>

<http://sciences.ucf.edu/statistics/dms/programs-2/>
DATA MINING PROGRAM

<http://sciences.ucf.edu/statistics/dms/news-events-2/>
DATA MINING PROGRAM

<http://sciences.ucf.edu/statistics/dms/people/>
DATA MINING PROGRAM

<http://sciences.ucf.edu/statistics/dms/2018-cfe-lending-analytics-competition/>
DATA MINING PROGRAM

<http://sciences.ucf.edu/statistics/dms/bdas2018/>
DATA MINING PROGRAM

<http://sciences.ucf.edu/statistics/dms/data-mining-lab/>
DATA MINING PROGRAM

<http://sciences.ucf.edu/statistics/dms/test-page-as-job-opportunities/>
DATA MINING PROGRAM

<http://sciences.ucf.edu/statistics/dms/useful-links/>

DATA MINING PROGRAM

<http://sciences.ucf.edu/statistics/dms/sponsors/>

DATA MINING PROGRAM

<https://wiki.illinois.edu/wiki/display/cs591han/Research+Publications+-+Data+Mining+Research+Group+at+CS,+UIUC>

Chao Zhang and Jiawei Han, Multidimensional Mining of Massive Text Data, Morgan & Claypool Publishers, 2019

<http://illimine.cs.uiuc.edu/>

<http://hanj.cs.illinois.edu/>

<http://dais.cs.illinois.edu/>

The Data and Information Systems Laboratory (DAIS) conducts fundamental and applied cutting-edge research in many areas related to building intelligent data and information systems, especially Databases, Data Mining, Information Retrieval, and Web Information Systems.

http://hanj.cs.illinois.edu/projs/social_media.htm

NSF III: Small: Multi-Dimensional
Structuring, Summarizing and Mining of Social Media Data

<https://knoweng.org/>

The KnowEnG Team at 2016 Annual meeting

<http://hzhuang3.web.engr.illinois.edu/>

PhD Candidate
Department of Computer Science
University of Illinois at Urbana-Champaign
Email: hzhuang3 [AT] illinois [DOT] edu
Address:
Siebel Center for Computer Science
201 N. Goodwin Avenue
Urbana, IL 61801, USA

<https://cs.illinois.edu/news/professor-jiawei-han-named-aiken-endowed-chair>
Illinois Computer Science Professor Jiawei Han has been named a Michael Aiken Chair, one of the most distinguished honors on campus.

<https://www.cse.gatech.edu/news/610222/assistant-professor-chao-zhang-joins-cse>

<http://www-bcf.usc.edu/~xiangren/>

Assistant Professor, USC Computer Science
USC Information Sciences Institute (joint appt.)
Director, USC INK Research Lab
Information Director, SIGKDD
Forbes' Asia 30 Under 30

<https://www.linkedin.com/in/fangbo-tao-96456623/>

<http://persistencedatamining.com/how-it-works/>
Laboratory chemical analysis is time consuming and requires transporting samples to a lab.

<http://persistencedatamining.com/why-it-matters/>

We spend billions of dollars to grow more food yet use imprecise methods to make decisions that impact those input dollars.

<http://persistencedatamining.com/our-team/>

Brian has extensive experience in fundraising, natural resources, technology funding and development, joint ventures, debt, equity financing and financial brokerage.

<http://persistencedatamining.com/partners/>

Comments are closed.

<http://persistencedatamining.com/whats-new/>

There are ongoing economic pressures in production agriculture to increase crop yields. However, high grain yield production comes at a cost of applying significant quantities of various agricultural inputs, i.e.... read more →

<http://persistencedatamining.com/contact-us/>

San Diego Office

9404 Genesee Ave

Suite 340

La Jolla, California 92037

858-454-1062

Rural Office

3103 N. 1850 East Road

Mount Auburn, IL 62547

<https://www.linkedin.com/company/persistence-mining-inc-/>

<https://twitter.com/PersistenceData>

We've detected that JavaScript is disabled in your browser. Would you like to proceed to legacy Twitter?

<https://www.facebook.com/PersistenceData>
Another amazing show!

```
for page in nbrhd:  
    print(page)  
    try:  
        resp = urllib.request.urlopen(page)  
    except urllib.error.HTTPError as e:  
        if e.code in (... , 403, ...):  
            print("\n\n")  
            continue  
    except urllib.error.URLError as e:  
        print("\n\n")  
        continue  
    soup = BeautifulSoup(resp, from_encoding=resp.info().get_param('charset'))  
    try:  
        print(soup.find('p').get_text())  
    except AttributeError as e:  
        print("\n\n")  
        continue  
    print("\n\n")
```

https://en.wikipedia.org/wiki/Data_mining

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.[1] Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use.[1][2][3][4] Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.[5] Aside from the raw analysis step, it also involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.[1] The difference between data analysis and data mining is that data analysis is used to test models and hypotheses on the dataset, e.g., analyzing the effectiveness of a marketing campaign, regardless of the amount of data; in contrast, data mining uses machine-learning and statistical models to uncover clandestine or hidden patterns in a large volume of data.[6]

https://www.sas.com/en_us/insights/analytics/data-mining.html

Sign In

<https://searchsqlserver.techtarget.com/definition/data-mining>

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends.

<http://www.statsoft.com/textbook/data-mining-techniques>

https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm

This chapter provides a high-level orientation to data mining technology.

<https://www.techopedia.com/definition/1181/data-mining>

Data mining is the process of analyzing hidden patterns of data according to different perspectives for categorization into useful information, which is collected and assembled in common areas, such as data warehouses, for efficient analysis, data mining algorithms, facilitating business decision making and other information requirements to ultimately cut costs and increase revenue.

<https://theappsolutions.com/blog/development/data-mining-guide/>

We live in the age of massive data production. If you think about it - pretty much every gadget or service we are using creates a lot of information (for example, Facebook processes around 500+ terabytes of data each day). All this data goes straight back to the product owners, which they can use it to make a better product. This process of gathering data and making sense of it is called Data Mining.

<https://www.investopedia.com/terms/d/datamining.asp>

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing, and computer processing.

<https://link.springer.com/journal/10618>

[Skip to main content](#)

<https://datafloq.com/read/data-mining-techniques-create-business-value/121>

Datafloq is the one-stop source for big data, blockchain and artificial intelligence. We offer information, insights and opportunities to drive innovation with emerging technologies.

<http://dmg.org/>

PMML is the leading standard for statistical and data mining models and supported by over 20 vendors and organizations. With PMML, it is easy to develop a model on one system using one application and deploy the model on another system using another application, simply by transmitting an XML configuration file.

Neighbourhood graph for “Data Mining” (URLs only):

[https://en.wikipedia.org/wiki/Data_mining,
https://www.sas.com/en_us/insights/analytics/data-mining.html,
<https://searchsqlserver.techtarget.com/definition/data-mining>,
<http://www.statsoft.com/textbook/data-mining-techniques>,
https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm,
<https://www.techopedia.com/definition/1181/data-mining>,
<https://theappsolutions.com/blog/development/data-mining-guide/>,
<https://www.investopedia.com/terms/d/datamining.asp>,
<https://link.springer.com/journal/10618>,
<https://datafloq.com/read/data-mining-techniques-create-business-value/121>,
<http://dmg.org/>,
<https://www3.cs.stonybrook.edu/~cse634/>,

<https://scpd.stanford.edu/public/category/courseCategoryCertificateProfile.do%3Fmethod%3Dload%26certificateId%3D1209602>,
<https://onlinelibrary.wiley.com/journal/19321872>,
<https://orange.biolab.si/training/>,
<https://neilpatel.com/blog/data-mining/>,
<https://connect.informs.org/data-mining/home>,
<http://educationaldatamining.org/edm2019/>,
<https://rapidminer.com/data-mining-tools-try-rapidminer/>,

'https://store.steampowered.com/app/993630/Data_mining_4/',
'http://guidetodatamining.com/',
'https://inhomelandsecurity.com/fake-news-data-mining/',
'https://thenextweb.com/gaming/2019/03/15/gamers-epic-games-rift-data-mining-tencent/',
'https://medium.com/data-mining-the-city',
'https://extension.ucsd.edu/courses-and-programs/data-mining-for-advanced-analytics',

'https://www1.villanova.edu/villanova/business/centers/businessanalytics/studentresources/dataminin
g.html',
'https://support.crossref.org/hc/en-us/articles/215750183-Crossref-Text-and-Data-Mining-Services',
'http://sciences.ucf.edu/statistics/dms/',
'http://dm1.cs.uiuc.edu/',
'http://persistencedatamining.com/',
'https://arxiv.org/list/cs.LG/recent',
'http://www.kdnuggets.com/meetings/kdd89/',
'http://hurwitz.com/recent-research/item/advanced-analytics-the-hurwitz-victory-index',
'http://www.kdd.org/curriculum/index.html',
'http://www.britannica.com/EBchecked/topic/1056150/data-mining',
'https://web.archive.org/web/2009110212529/http://www-stat.stanford.edu/~tibs/ElemStatLearn/',
'http://www-stat.stanford.edu/~tibs/ElemStatLearn/',
'http://proquest.safaribooksonline.com.proxy.library.carleton.ca/book/databases/data-
warehouses/9780123814791',
'http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf',
'http://www.okairp.org/documents/2005%20Fall/F05_ROMEDataQualityETC.pdf',

'https://web.archive.org/web/20140201170452/http://www.okairp.org/documents/2005%20Fall/F05_
ROMEDataQualityETC.pdf',
'https://www.sas.com/en_us/home.html',

'https://www.sas.com/profile/user/contact.htm?locale=en_us&returnURL=https://www.sas.com/en_us
/insights/analytics/data-mining.html',
'https://www.sas.com/en_si/home.html',
'https://www.sas.com/es_ar/home.html',
'https://www.sas.com/en_au/home.html',
'https://www.sas.com/de_at/home.html',
'https://www.sas.com/en_be/home.html',
'https://www.sas.com/pt_br/home.html',
'https://www.sas.com/en_ca/home.html',
'https://www.sas.com/es_cl/home.html',
'https://searchsqlserver.techtarget.com',
'https://users.techtarget.com/registration/searchSQLServer/LoginRegister.page',
'https://users.techtarget.com/registration/searchSQLServer/Register.page',
'https://www.techtarget.com/network',
'https://searchsqlserver.techtarget.com/info/news',

'https://searchsqlserver.techtarget.com/features',
'https://searchsqlserver.techtarget.com/tips',
'https://searchsqlserver.techtarget.com/answers',
'https://whatis.techtarget.com/resources/Buyers-Guides',
'http://www.statistica.io/',
'http://statistica.io/products/',
'http://statistica.io/uses/',
'http://statistica.io/contact-us/sales/',
'http://statistica.io/resources/trial-download/',
'https://support.tibco.com/s/',
'http://www.statistica.io',
'http://www.statsoft.com/Textbook',
'http://www.statsoft.com/Textbook/Data-Mining-Techniques',
'http://www.statsoft.com/textbook/fraud-detection/',
'http://www.kdnuggets.com/',
'http://www.twocrows.com/',
'https://www.facebook.com/techopedia',
'https://www.linkedin.com/company/techopedia',
'https://twitter.com/techopedia',

'https://www.datadoghq.com/lpg/?utm_source=Advertisement&utm_medium=Advertisement&utm_campaign=Techopedia-Textlink',
'https://www.techopedia.com/reg/managing-multiple-database-roles-how-many-hats-do-you-wear/33762?utm_source=techopedia&utm_medium=hellobar',
'https://www.techopedia.com/reg/how-to-continuously-monitor-and-analyze-mysql-and-mariadb-with-ideras-sql-diagnostic-manager/33743?utm_source=techopedia&utm_medium=hellobar',
'https://www.techopedia.com/resources?utm_source=techopedia&utm_medium=hellobar',
'https://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/',
'https://www.salesforce.com/products/marketing-cloud/best-practices/data-mining-predictive-analytics/',
'https://aws.amazon.com/ru/pinpoint/customer-engagement/customer-segmentation/',
'https://www.statista.com/statistics/254266/global-big-data-market-forecast/',
'https://appcost.theappsolutions.com/web/#calcAppForm',
'https://appcost.theappsolutions.com/#calcAppForm',
'https://theappsolutions.com/blog/development/machine-learning-algorithm-types/',
'https://theappsolutions.com/blog/development/what-why-how-adtech/',
'https://theappsolutions.com/blog/development/how-to-develop-social-media-app/',
'https://theappsolutions.com/blog/development/ar-vr-in-education/',
'https://theappsolutions.com/blog/development/develop-app-with-geolocation/',
'https://www.investopedia.com/dictionary/',
'https://www.investopedia.com/university/beginner/',
'https://www.investopedia.com/articles/markets/101415/4-best-sp-500-index-funds.asp',
'https://www.investopedia.com/insights/worlds-top-economies/'

'https://www.investopedia.com/university/stocks/','
'https://www.investopedia.com/university/options/','
'https://www.investopedia.com/university/economics/','
'https://www.investopedia.com/markets/','
'https://www.investopedia.com/markets/watchlist/','
'https://www.investopedia.com/markets/stocks/aapl/','
'http://www.springer.com/journal/10618/about',
'http://www.springer.com/journal/10618/edboard',
'http://www.springer.com/journal/10618/submit',
'http://www.springer.com/','
'http://www.springerprotocols.com/','
'http://materials.springer.com/','
'http://adis.springer.com/','
'https://twitter.com/datafloq',
'https://www.linkedin.com/company/datafloq',
'https://www.facebook.com/datafloq',
'https://imagjn.datafloq.com',
'https://imagjn.datafloq.com/#Documents',
'http://twitter.com/DMG_Standards',
'http://dmg.org/documents/dmg-pmml-license-2016.pdf',
'http://dmg.org/documents/dmg-pfa-license-2016.pdf',
'http://bit.ly/StandardsSIGKDD2017',
'http://www.kdd.org/kdd2016/program',
'http://bit.ly/DMGRoundtable',
'http://www.kdnuggets.com/2016/01/portable-format-analytics-models-production.html',
'http://www.cs.sunysb.edu/%7Eanita',
'http://www.cs.waikato.ac.nz/%7Eml/weka/index.html',
'http://www.kdnuggets.com/datasets/index.html',
'http://www.kdnuggets.com/datasets/competitions.html',
'http://kdd.ics.uci.edu/','

'http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:20388241%7Em
enuPK:665266%7EpagePK:64165401%7EpiPK:64165026%7EtheSitePK:469382,00.html',
'http://www.stonybrook.edu/uaa/academicjudiciary/','
'http://http://studentaffairs.stonybrook.edu/dss/','
'http://www.sunysb.edu/ehs/fire/disabilities.shtml/','
'http://seminars.stanford.edu',
'https://mvideos.stanford.edu/Previews',
'https://scpd.stanford.edu/portal/student/studentHome.do?method=load',
'https://scpd.stanford.edu/portal/student/studentProfile.do?method=edit',
'https://scpd.stanford.edu/portal/student/studentCredential.do?method=loadCredential',
'https://scpd.stanford.edu/portal/student/studentAcademicHistory.do?method=load',
'https://scpd.stanford.edu/portal/student/applicationInstanceSearch.do?method=load',
'https://scpd.stanford.edu/portal/student/studentCertificates.do?method=load',

'<https://scpd.stanford.edu/admissions/additional-resources/forms-documents>',
'<https://scpd.stanford.edu/admissions/additional-resources/registration-policies>',
'<http://scpd.stanford.edu/free-stuff/free-stuff>',
'<https://onlinelibrary.wiley.com/search/advanced>',
'<https://onlinelibrary.wiley.com/search/advanced#citation>',
'<https://ordering.onlinelibrary.wiley.com/subs.asp?ref=&doi=>',
'<https://www.wiley.com/privacy>',
'<https://onlinelibrary.wiley.com/terms-and-conditions>',
'<https://onlinelibrary.wiley.com/cookies>',
'<https://onlinelibrary.wiley.com/accessibility>',
'<https://hub.wiley.com/community/support/onlinelibrary>',
'<https://onlinelibrary.wiley.com/agents>',
'<https://onlinelibrary.wiley.com/advertisers>',
'<https://hub.wiley.com/community/exchanges/>',
'<http://www.informs.org/>',
'<https://www.certifiedanalytics.org/>',
'<http://pubsonline.informs.org/>',
'<http://careercenter.informs.org/>',
'<http://meetings2.informs.org/wordpress/2019international/>',
'https://www.informs.org/user/login?redirect_uri=https%3a%2f%2fconnect.informs.org%2fdata-mining%2fhome',
'<https://www.informs.org/Discover>',
'<https://www.informs.org/Explore>',
'<https://www.informs.org/Get-Involved>',
'<https://www.informs.org/Impact>',
'<http://educationaldatamining.org/edm2019>',
'<http://educationaldatamining.org/edm2019/call-for-papers>',
'<http://educationaldatamining.org/edm2019/keynote>',
'http://educationaldatamining.org/edm2019/workshops_tutorials',
'http://educationaldatamining.org/edm2019/call-for-papers/#important_dates',
'<http://educationaldatamining.org/edm2019/registration>',
'<http://educationaldatamining.org/edm2019/sponsors>',
'<http://educationaldatamining.org/edm2019/committee>',
'<http://educationaldatamining.org/edm2019/location>',
'https://store.steampowered.com/login/?redir=app%2F993630%2FData_mining_4%2F&redir_ssl=1',
'https://store.steampowered.com',
'<https://store.steampowered.com/explore>',
'<https://store.steampowered.com/curators>',
'<https://steamcommunity.com/my/wishlist>',
'<https://store.steampowered.com/news>',
'<https://store.steampowered.com/stats>',
'<https://steamcommunity.com/>',
'<https://steamcommunity.com/discussions>',
'<http://zacharski.org>',

'<http://creativecommons.org/licenses/by-nc/4.0/>',
'<https://github.com/zacharski/pg2dm-python>',

'https://docs.google.com/spreadsheets/d/1pkIU_T3lo1089hSKv_LhRI5QWWYBk5oOh4mASImIDVI/edit#gid=0',
'<http://github.com/zacharski/pg2dm-python>',
'<https://inhomelandsecurity.com>',
'<https://inhomelandsecurity.com/about-the-blog/>',
'<https://inhomelandsecurity.com/category/columnists/>',
'<https://inhomelandsecurity.com/category/border-wall/>',
'<https://inhomelandsecurity.com/category/homeland-security/>',
'<https://inhomelandsecurity.com/category/terrorism-threats/>',
'<https://inhomelandsecurity.com/category/politics-government/>',
'<https://inhomelandsecurity.com/>',
'<https://index.co/>',
'<https://tq.co/>',

'https://deals.thenextweb.com/?utm_source=thenextweb.com&utm_medium=referral&utm_campaign=navbar',
'<https://answers.thenextweb.com/>',

'https://thenextweb.com/conference/?utm_source=thenextweb.com&utm_medium=referral&utm_campaign=nav-link',
'<https://www.facebook.com/thenextweb>',
'<https://myextension.ucsd.edu/>',
'<https://ucsddextension.blackboard.com>',
'<https://instructorlink.ucsd.edu/>',
'<https://extcart.ucsd.edu/index.cfm?vAction=view>',
'<http://osha.ucsd.edu/>',
'<http://academicconnections.ucsd.edu/>',
'<http://gels.ucsd.edu>',
'<https://apply.ucsd.edu/extension/startapp.php?subjectcode=CERT&coursenumber=DMAA>',
'<https://extension.ucsd.edu/courses-and-programs/statistics-for-data-analytics>',
'<http://www.thegreatergreat.com/>',
'<https://mynova.villanova.edu/>',
'<http://outlook.villanova.edu/>',
'<http://elearning.villanova.edu/>',
'https://secure.touchnet.com/C21740_ustores/web/store_main.jsp?STOREID=4&SINGLESTORE=true',
'<http://webmail.villanova.edu/>',
'<http://library.villanova.edu>',
'<http://www.villanova.com>',

'<https://www1.villanova.edu/villanova/business/centers/businessanalytics/studentresources/datamining/registration.html>',

'https://www1.villanova.edu/content/dam/villanova/VSB/centers/cba/WestCap1-1.xlsx',
'http://www.crossref.org',
'http://api.crossref.org/',
'https://youtu.be/LBYgq6jPoyk',
'http://www.crossref.org/08downloads/webinars/tdm_walkthrough_032415.mp4',
'http://support.crossref.org/hc/en-us/articles/215750263',
'http://support.crossref.org/hc/en-us/articles/215750203',

'https://www.facebook.com/share.php?title=Crossref+Text+and+Data+Mining+Services&u=https%3A%2F%2Fsupport.crossref.org%2Fhc%2Fen-us%2Farticles%2F215750183-Crossref-Text-and-Data-Mining-Services',

'https://twitter.com/share?lang=en&text=Crossref+Text+and+Data+Mining+Services&url=https%3A%2F%2Fsupport.crossref.org%2Fhc%2Fen-us%2Farticles%2F215750183-Crossref-Text-and-Data-Mining-Services',

'https://www.linkedin.com/shareArticle?mini=true&source=Crossref&title=Crossref+Text+and+Data+Mining+Services&url=https%3A%2F%2Fsupport.crossref.org%2Fhc%2Fen-us%2Farticles%2F215750183-Crossref-Text-and-Data-Mining-Services',
'http://sciences.ucf.edu/statistics/dms/programs-2/',
'http://sciences.ucf.edu/statistics/dms/news-events-2/',
'http://sciences.ucf.edu/statistics/dms/people/',
'http://sciences.ucf.edu/statistics/dms/2018-cfe-lending-analytics-competition/',
'http://sciences.ucf.edu/statistics/dms/bdas2018/',
'http://sciences.ucf.edu/statistics/dms/data-mining-lab/',
'http://sciences.ucf.edu/statistics/dms/test-page-as-job-opportunities/',
'http://sciences.ucf.edu/statistics/dms/useful-links/',
'http://sciences.ucf.edu/statistics/dms/sponsors/',
'https://wiki.illinois.edu//wiki/display/cs591han/Research+Publications+-+Data+Mining+Research+Group+at+CS,+UIUC',
'http://illimine.cs.uiuc.edu/',
'http://hanj.cs.illinois.edu/',
'http://dais.cs.illinois.edu/',
'http://hanj.cs.illinois.edu/projs/social_media.htm',
'https://knoweng.org/',
'http://hzhuang3.web.engr.illinois.edu/',
'https://cs.illinois.edu/news/professor-jiawei-han-named-aiken-endowed-chair',
'https://www.cse.gatech.edu/news/610222/assistant-professor-chao-zhang-joins-cse',
'http://www-bcf.usc.edu/~xiangren/',
'https://www.linkedin.com/in/fangbo-tao-96456623/',
'http://persistencedatamining.com/how-it-works/',
'http://persistencedatamining.com/why-it-matters/',
'http://persistencedatamining.com/our-team/',
'http://persistencedatamining.com/partners/'

```
'http://persistencedatamining.com/whats-new/',
'http://persistencedatamining.com/contact-us/',
'https://www.linkedin.com/company/persistence-mining-inc-/',
'https://twitter.com/PersistenceData',
'https://www.facebook.com/PersistenceData']
```

Neighbourhood graph for “Machine Learning” (URLs only):

nbrhd

Out[11]:

```
['https://arxiv.org',
'https://developers.google.com/machine-learning/crash-course/',
'https://en.wikipedia.org/wiki/Machine_learning',
'https://cloud.google.com/products/ai/',
'https://xkcd.com/1838/',
'https://aws.amazon.com/machine-learning/',
'https://www.wired.com/tag/machine-learning/',
'https://developer.apple.com/machine-learning/',
'https://towardsdatascience.com/machine-learning/home',
'https://searchenterpriseai.techtarget.com/definition/machine-learning-ML',
'https://azure.microsoft.com/en-us/services/machine-learning-service/',
'https://www.brookings.edu/research/what-is-machine-learning/',
'https://www.sas.com/en_us/insights/analytics/machine-learning.html',
'https://www.zdnet.com/article/what-is-machine-learning-everything-you-need-to-know/',
'https://www.forbes.com/.../20-examples-of-machine-learning-used-in-customer-experi',
'https://emerj.com',
'https://www.edx.org/learn/machine-learning',
'https://www.mathworks.com/discovery/machine-learning.html',
'https://www.youtube.com/watch?v=ukzFl9rgwfU',
'https://www.ml.cmu.edu/',
'https://royalsociety.org/topics-policy/projects/machine-learning/',
'https://www.sap.com/products/leonardo/machine-learning.html',
'https://www.elastic.co/products/stack/machine-learning',
'https://www.udemy.com/topic/machine-learning/',
'https://medium.com/machine-learning.../why-machine-learning-matters-6164faf1df12',
'https://www.sas.com/en_us/insights/analytics/machine-learning.html',
'https://www.expertsystem.com/machine-learning-definition/',
'https://www.coursera.org/learn/machine-learning',
'https://www.udacity.com/course/intro-to-machine-learning--ud120',
'https://www.geeksforgeeks.org/machine-learning/',
'https://simonsfoundation.org',
'https://confluence.cornell.edu/x/ALIRF',
'https://bit.ly/arXivDONATE1',
```

'https://www.cornell.edu/','
'https://arxiv.org/login','
'https://arxiv.org/help','
'https://arxiv.org/search/advanced','
'https://confluence.cornell.edu/x/rIzLFQ','
'https://arxiv.org/search/astro-ph','
'https://arxiv.org/search/cond-mat','
'https://developers.google.com/machine-learning/','
'https://developers.google.com/machine-learning/practica','
'https://developers.google.com/machine-learning/guides','
'https://developers.google.com/machine-learning/glossary','
'https://developers.google.com/products/','
'https://accounts.google.com/ServiceLogin?passive=true&continue=https%3A%2F%2Fdevelopers.google.com%2Fmachine-learning%2Fcrash-course%2F%3Frefresh%3D1&service=ahsid','
'https://developers.google.com/machine-learning/problem-framing/index','
'https://developers.google.com/machine-learning/data-prep/index','
'https://developers.google.com/machine-learning/crash-course/?nav=true','
'https://arxiv.org/list/cs.LG/recent','
'http://eprints.ecs.soton.ac.uk/12954/','
'https://www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan/ckelmtt?context=3','
'http://projecteuclid.org/download/pdf_1/euclid.ss/1009213726','
'http://www-bcf.usc.edu/~gareth/ISL/','
'https://mitpress.mit.edu/books/introduction-machine-learning','

'https://books.google.com/books?id=7f5bBAAAQBAJ&printsec=frontcover#v=onepage&q=classification&f=false','
'http://www.athenasc.com/dpbook.html','
'http://www.athenasc.com/ndpbook.html','
'http://machinelearning.wustl.edu/mlpapers/paper_files/AISTATS2011_CoatesNL11.pdf','
'http://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/csurka-eccv-04.pdf','
'https://cloud.google.com/','
'https://cloud.google.com/why-google-cloud/','
'https://cloud.google.com',
'https://cloud.google.com/security/','
'https://cloud.google.com/open-cloud/','
'https://cloud.google.com/customers/','
'https://cloud.google.com/analyst-reports/','
'http://what-if.xkcd.com',
'http://blag.xkcd.com',
'http://store.xkcd.com/','
'https://xkcd.com/newsletter/','
'http://threewordphrase.com/','
'http://www.smbc-comics.com/','

'http://www.qwantz.com',
'http://oglaf.com/',
'http://www.asofterworld.com',
'http://buttersafe.com/',
'http://pbfcomics.com/',
'https://aws.amazon.com/?nc2=h_lg',
'https://console.aws.amazon.com/support/home?nc2=h_ql_cu',
'https://portal.aws.amazon.com/gp/aws/developer/registration/index.html?nc2=h_ct&src=default',
'https://docs.aws.amazon.com/index.html?nc2=h_ql_doc',
'https://aws.amazon.com/marketplace/?nc2=h_ql_mp',
'https://aws.amazon.com/ar/?nc1=h_ls',
'https://aws.amazon.com/id/?nc1=h_ls',
'https://aws.amazon.com/de/?nc1=h_ls',
'https://aws.amazon.com/?nc1=h_ls',
'https://aws.amazon.com/es/?nc1=h_ls',
'https://aws.amazon.com/fr/?nc1=h_ls',
'https://subscribe.wired.com/subscribe/splits/wired/WIR_FAILSAFE?source=WIR_FAILSAFE_MOBILE_GLOBAL_NAVBAR',
'https://subscribe.wired.com/subscribe/splits/wired/WIR_FAILSAFE?source=WIR_FAILSAFE_GLOBAL_NAVBAR',
'https://subscribe.wired.com/subscribe/splits/wired/WIR_FAILSAFE?source=WIR_FAILSAFE_MPUS',
'https://www.facebook.com/wired',
'http://www.twitter.com/wired',
'https://pinterest.com/wired',
'http://www.youtube.com/wired',
'https://instagram.com/wired',
'https://www.wired.com/subscribe/',
'https://www.wired.com/wired-advertising/',
'https://www.wired.com/sitemap/',
'https://searchenterpriseai.techtarget.com',
'https://users.techtarget.com/registration/searchEnterpriseAI/LoginRegister.page',
'https://users.techtarget.com/registration/searchEnterpriseAI/Register.page',
'https://www.techtarget.com/network',
'https://searchenterpriseai.techtarget.com/info/news',
'https://searchenterpriseai.techtarget.com/features',
'https://searchenterpriseai.techtarget.com/tips',
'https://searchenterpriseai.techtarget.com/answers',
'https://whatis.techtarget.com/resources/Buyers-Guides',
'https://azure.microsoft.com/en-us/',
'https://azure.microsoft.com/en-us/overview/sales-number/',
'https://azure.microsoft.com/en-us/search/',
'https://azure.microsoft.com/en-us/account/',
'https://portal.azure.com/',
'https://azure.microsoft.com/en-us/free/>,

'https://azure.microsoft.com/en-us/solutions/',
'https://azure.microsoft.com/en-us/services/virtual-machines/',
'https://www.brookings.edu/about-us/',
'https://www.brookings.edu/executive-education/',
'https://brookings.foxcart.com/cart?cart-view',
'https://www.brookings.edu/',
'https://www.brookings.edu/research/what-is-machine-learning/?share=facebook',
'https://www.brookings.edu/research/what-is-machine-learning/?share=twitter',
'https://www.brookings.edu/research/what-is-machine-learning/?share=linkedin',
'https://www.brookings.edu/research/what-is-machine-learning/#print',
'https://www.brookings.edu/research/what-is-machine-learning/?share=custom-1464743455',
'https://www.brookings.edu/research/what-is-machine-learning/?share=email',
'https://www.brookings.edu/research/what-is-machine-learning/?share=reddit',
'https://www.sas.com/en_us/home.html',
'https://www.sas.com/profile/user/contact.htm?locale=en_us&returnURL=https://www.sas.com/en_us/insights/analytics/machine-learning.html',
'https://www.sas.com/en_si/home.html',
'https://www.sas.com/es_ar/home.html',
'https://www.sas.com/en_au/home.html',
'https://www.sas.com/de_at/home.html',
'https://www.sas.com/en_be/home.html',
'https://www.sas.com/pt_br/home.html',
'https://www.sas.com/en_ca/home.html',
'https://www.sas.com/es_cl/home.html',
'http://www.zdnet.com.cn/',
'http://www.zdnet.fr/',
'http://www.zdnet.de/',
'http://www.zdnet.co.kr/',
'http://japan.zdnet.com/',
'https://www.zdnet.com/',
'http://www.techproresearch.com/',
'http://www.zdnet.com/article/vpn-services-2018-the-ultimate-guide-to-protecting-your-data-on-the-internet/',
'https://www.techrepublic.com/resource-library/whitepapers/',
'https://www.techrepublic.com/forums/',
'https://www.zdnet.com/user/login/?appId=1101',
'https://www.mathworks.com/matlabcentral/?s_tid=gn_mlc',
'https://www.mathworks.com/help/stats/support-vector-machine-classification.html',
'https://www.mathworks.com/help/stats/classification-ensembles.html',
'https://www.mathworks.com/help/stats/classification-trees.html',
'https://www.mathworks.com/help/stats/classification-nearest-neighbors.html',
'https://www.mathworks.com/help/stats/classification-naive-bayes.html',
'https://www.mathworks.com/help/stats/classification-discriminant-analysis.html',
'https://www.mathworks.com/help/stats/generalized-linear-regression-1.html',

'<https://www.mathworks.com/help/deeplearning/pattern-recognition-and-classification.html>',
'<https://www.google.com/get/videoqualityreport/?v=ukzFl9rgwfU>',
'https://accounts.google.com/ServiceLogin?passive=true&continue=https%3A%2F%2Fwww.youtube.com%2Fsignin%3Faction_handle_signin%3Dtrue%26app%3Ddesktop%26next%3D%252Fwatch%253Fv%253DukzFl9rgwfU%26hl%3Den%26feature%3D_FEATURE__&service=youtube&uilel=3&hl=en',
'https://www.youtube.com/user/Simplilearn?sub_confirmation=1',
'<https://www.google.com/intl/en/policies/privacy/>',
'https://accounts.google.com/ServiceLogin?passive=true&continue=https%3A%2F%2Fwww.youtube.com%2Fsignin%3Faction_handle_signin%3Dtrue%26app%3Ddesktop%26next%3D%252Fwatch%253Fv%253DukzFl9rgwfU%26hl%3Den%26feature%3Dplaylist&service=youtube&uilel=3&hl=en',
'<https://www.daibuilds.com/>',
'<https://www.daibuilds.com/pittsburgh-web-design/>',
'<https://www.daibuilds.com/pittsburgh-seo/>',
'<https://blog.daibuilds.com>',
'<https://www.robertoiriondo.com>',
'<https://www.ml.cmu.edu>',
'<https://cs.cmu.edu>',
'<https://cmu.edu>',
'https://www.ml.cmu.edu/Faculty_Hiring.html',
'<https://blog.ml.cmu.edu>',
'<https://www.cs.cmu.edu/bs-in-artificial-intelligence/>',
'<https://grants.royalsociety.org/>',
'<https://royalsociety.org/fellows-dashboard/journal-access/>',
'<https://prints.royalsociety.org/>',
'<http://www.enable-javascript.com/>',
'<https://royalsociety.org/stay-in-touch/email-newsletters/>',
'<https://blogs.royalsociety.org/in-verba/2019/02/21/ai-for-society-creating-ai-that-supports-equality-transparency-and-democracy/>',
'<https://blogs.royalsociety.org/in-verba/2018/12/12/synthetic-biology-where-are-we-now-and-where-are-we-going/>',
'<https://blogs.royalsociety.org/in-verba/2018/12/10/close-encounters-machine-learning-meets-public-policy-at-neurips-2018/>',
'<https://developers.sap.com/index.html>',
'<https://cloudplatform.sap.com/index.html>',
'<https://support.sap.com/en/index.html>',
'<https://support.sap.com/en/my-support.html>',
'<https://support.sap.com/en/my-support/knowledge-base.html>',
'<https://support.sap.com/en/my-support/software-downloads.html>',
'<https://support.sap.com/en/my-support/incidents.html>',
'<https://support.sap.com/en/alm/sap-cloud-alm.html>',
'<https://support.sap.com/en/solution-manager.html>',
'https://support.sap.com/en/my-support.html#section_1969201630',
'<https://www.elastic.co/products/>',
'<https://www.elastic.co/>',

'https://www.elastic.co/contact',
'https://www.elastic.co/de/products/stack/machine-learning',
'https://www.elastic.co/es/products/stack/machine-learning',
'https://www.elastic.co/fr/products/stack/machine-learning',
'https://www.elastic.co/jp/products/stack/machine-learning',
'https://www.elastic.co/kr/products/stack/machine-learning',
'https://www.elastic.co/cn/products/stack/machine-learning',
'https://www.elastic.co/pt/products/stack/machine-learning',
'https://www.expertsystem.com/category/blog/',
'https://www.expertsystem.com/contacts/',
'https://twitter.com/Expert_System',
'http://www.linkedin.com/company/expert-system',
'https://www.expertsystem.com',
'https://www.expertsystem.com/industries/',
'https://www.expertsystem.com/industries/banking-insurance/',
'https://www.expertsystem.com/customers/media-publishing/',
'https://www.expertsystem.com/solutions/defense-and-intelligence/',
'https://learner.coursera.help/hc/',
'https://itunes.apple.com/app/apple-store/id736535961?pt=2334150&ct=Coursera%20Web%20Promo%20Banner&mt=8',
'http://play.google.com/store/apps/details?id=org.coursera.android',
'https://about.coursera.org/',
'https://about.coursera.org/leadership',
'https://www.coursera.community',
'https://www.coursera.org/about/partners',
'https://building.coursera.org/developer-program/',
'https://www.coursera.community/beta-tester-news-announcements-20/join-the-coursera-beta-tester-community-47',
'https://www.coursera.community/gtc-news-announcements-17/join-the-coursera-global-translator-community-45',
'https://blog.coursera.org',
'https://auth.udacity.com/sign-in?next=https://classroom.udacity.com/authenticated',
'https://auth.udacity.com/sign-up?next=https://classroom.udacity.com/authenticated',
'https://www.udacity.com/course/data-analyst-nanodegree--nd002',
'https://www.udacity.com/school-of-ai',
'https://www.udacity.com/school-of-autonomous-systems',
'https://www.udacity.com/school-of-business',
'https://www.udacity.com/school-of-data-science',
'https://www.udacity.com/school-of-programming',
'https://career-resource-center.udacity.com',
'https://www.geeksforgeeks.org/',
'https://practice.geeksforgeeks.org/courses/',
'https://contribute.geeksforgeeks.org/wp-admin/post-new.php',
'http://www.geeksforgeeks.org/fundamentals-of-algorithms/'

'<http://www.geeksforgeeks.org/fundamentals-of-algorithms/#AnalysisofAlgorithms>',
'<http://www.geeksforgeeks.org/searching-algorithms/>',
'<http://www.geeksforgeeks.org/sorting-algorithms/>',
'<http://www.geeksforgeeks.org/graph-data-structure-and-algorithms/>',
'<http://www.geeksforgeeks.org/bitwise-algorithms/>',
'<https://www.geeksforgeeks.org/algorithms-gq/pattern-searching/>']

Neighborhood set for “Deep Learning”:

[https://en.wikipedia.org/wiki/Deep_learning',
'<http://deeplearning.net/>',
'<https://www.mathworks.com/discovery/deep-learning.html>',
'<https://www.udacity.com/course/intro-to-tensorflow-for-deep-learning--ud187>',
'<https://machinelearningmastery.com/what-is-deep-learning/>',
'<https://developer.nvidia.com/deep-learning>',
'<https://www.forbes.com/sites/bernardmarr/2018/10/01/what-is-deep-learning-ai-a-simple-guide-with-8-practical-examples/>',
'<https://skymind.ai/wiki/neural-network>',
'<http://neuralnetworksanddeeplearning.com/chap6.html>',
'<https://www.investopedia.com/terms/d/deep-learning.asp>',
'<https://www.technologyreview.com/s/513696/deep-learning/>',
'<https://www.coursera.org/specializations/deep-learning>',
'<https://www.deeplearningbook.org/>',
'<https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network>',
'https://www.sas.com/en_us/insights/analytics/deep-learning.html',
'<https://medium.com/tensorflow/mit-introduction-to-deep-learning-4a6f8dde1f0c>',
'<https://www.edx.org/professional-certificate/ibm-deep-learning>',
'<https://www.udemy.com/deeplearning/>',
'<https://deeplearning.mit.edu/>',
'<https://aws.amazon.com/deep-learning/>',
'<https://www.ibm.com/cloud/deep-learning>',
'<https://blogs.scientificamerican.com/observations/a-deep-dive-into-deep-learning/>',
'<https://www.datacamp.com/courses深深学习-in-python>',
'<https://www.deeplearning.ai/>',
'<https://software.intel.com/en-us/ai/courses深深学习>',
'<https://www.nature.com/articles/s41576-019-0122-6>',
'<http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science深深学习.html>',
'<https://towardsdatascience.com/detecting-malaria-with-deep-learning-9e45c1e34b60>',
'<https://course.fast.ai/>',
'<http://deeplearning.stanford.edu/tutorial/>',
'<https://arxiv.org/list/cs.LG/recent>',
'<http://adsabs.harvard.edu/abs/2015Natur.521..436L>',
'<http://ieeexplore.ieee.org/document/6248110/>',
'https://www.cs.toronto.edu/~kriz/imagenet_classification_with_deep_convolutional.pdf',

'<https://techcrunch.com/2017/05/24/alphago-beats-planets-best-human-go-player-ke-jie/amp/>',
'<http://adsabs.harvard.edu/abs/1996Natur.381..607O>',
'<http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf>',

'<http://sanghv.com/download/soft/machine%20learning,%20artificial%20intelligence,%20mathematics%20ebooks/ML/learning%20deep%20architectures%20for%20AI%20%282009%29.pdf>',
'http://www.scholarpedia.org/article/Deep_Learning',
'<http://adsabs.harvard.edu/abs/2009SchpJ...4.5947H>',
'<http://deeplearning.net/reading-list/>',
'<http://deeplearning.net/reading-list/tutorials/>',
'http://deeplearning.net/software_links/',
'<http://deeplearning.net/blog/>',
'<http://deeplearning.net/demos/>',
'<http://deeplearning.net/datasets/>',
'<http://deeplearning.net/events/>',
'<http://deeplearning.net/bibliography/>',
'<http://deeplearning.net/deep-learning-research-groups-and-labs/>',
'https://www.mathworks.com/matlabcentral/?s_tid=gn_mlc',
'<https://www.mathworks.com/matlabcentral/fileexchange/60659-deep-learning-in-11-lines-of-matlab-code>',
'<https://www.mathworks.com/help/vision/examples/object-detection-using-deep-learning.html>',
'https://www.mathworks.com/matlabcentral/fileexchange/?s_tid=hp_ff_p_fx',
'https://www.mathworks.com/downloads/web_downloads/?s_iid=hp_ff_t_downloads',
'https://www.mathworks.com/store?s_iid=hp_ff_t_buy',
'https://www.mathworks.com/help/?s_tid=hp_ff_l_doc',
'https://www.mathworks.com/support/learn-with-matlab-tutorials.html?s_tid=hp_ff_l_tutorials',
'https://www.mathworks.com/help/examples.html?s_tid=hp_ff_l_examples',
'https://www.mathworks.com/support/install-matlab.html?s_tid=hp_ff_s_install',
'<https://auth.udacity.com/sign-in?next=https://classroom.udacity.com/authenticated>',
'<https://auth.udacity.com/sign-up?next=https://classroom.udacity.com/authenticated>',
'<https://www.udacity.com/tech-requirements>',
'<https://www.udacity.com/school-of-ai>',
'<https://www.udacity.com/school-of-autonomous-systems>',
'<https://www.udacity.com/school-of-business>',
'<https://www.udacity.com/school-of-data-science>',
'<https://www.udacity.com/school-of-programming>',
'<https://career-resource-center.udacity.com>',
'<https://news.developer.nvidia.com/>',
'<https://devblogs.nvidia.com/>',
'<https://devtalk.nvidia.com/>',
'<https://developer.nvidia.com/deep-learning-software>',
'<https://developer.nvidia.com/tensorrt>',
'<https://www.nvidia.com/en-us/deep-learning-ai/education/?ncid=so-dis-dldlwsd1-72342>',
'<https://courses.nvidia.com/courses/course-v1:DLI+C-FX-01+V2/about?ncid=so-dis-dldlwsd1-72346>',

'<https://www.nvidia.com/en-us/deep-learning-ai/education/?ncid=so-dis-dldlwsd1-72528#dlti>',
'http://www.nvidia.com/object/legal_info.html',
'<https://www.forbes.com/worlds-billionaires/>',
'<https://www.forbes.com/billionaires/>',
'<https://www.forbes.com/forbes-400/>',
'<https://www.forbes.com/self-made-women/>',
'<https://www.forbes.com/china-billionaires/>',
'<https://www.forbes.com/india-billionaires/>',
'<https://www.forbes.com/indonesia-billionaires/>',
'<https://www.forbes.com/korea-billionaires/list/>',
'<https://www.forbes.com/thailand-billionaires/list/>',
'<https://www.forbes.com/japan-billionaires/list/>',
'https://www.exxactcorp.com/Deep-Learning-NVIDIA-GPU-Solutions?utm_source=neuralnetworksanddeeplearning&utm_medium=banner&utm_campaign=sponsors',
'<http://gsquaredcapital.com/>',
'<http://www.tineye.com>',
'<http://www.visionsmarts.com>',
'https://twitter.com/michael_nielsen',
'<https://github.com/mnielsen/neural-networks-and-deep-learning>',
'<http://eepurl.com/0Xxjb>',
'<http://www.deeplearningbook.org/>',
'<http://cognitivemedium.com>',
'<http://michaelnielsen.org>',
'<https://www.investopedia.com/dictionary/>',
'<https://www.investopedia.com/university/beginner/>',
'<https://www.investopedia.com/articles/markets/101415/4-best-sp-500-index-funds.asp>',
'<https://www.investopedia.com/insights/worlds-top-economies/>',
'<https://www.investopedia.com/university/stocks/>',
'<https://www.investopedia.com/university/options/>',
'<https://www.investopedia.com/university/economics/>',
'<https://www.investopedia.com/markets/>',
'<https://www.investopedia.com/markets/watchlist/>',
'<https://www.investopedia.com/markets/stocks/aapl/>',
'<https://subscribe.technologyreview.com/ecom/mtr/app/live/subcustserv?pagemode=start&org=MTR&publ=TR&php=Y>',
'<https://go.technologyreview.com/newsletters/the-download/>',
'<https://go.technologyreview.com/newsletters/chain-letter/>',
'<https://go.technologyreview.com/newsletters/the-algorithm/>',
'<https://go.technologyreview.com/newsletters/fwd-economy>',
'<https://go.technologyreview.com/newsletters/the-airlock>',
'<https://go.technologyreview.com/newsletters>',
'<https://events.technologyreview.com/emtech/>',
'<https://events.technologyreview.com/emtech/digital/>'

'https://events.technologyreview.com/emtech/next/','
'https://events.technologyreview.com/presents/business-of-blockchain/2019/','
'https://learner.coursera.help/hc/','
'https://itunes.apple.com/app/apple-store/id736535961?pt=2334150&ct=Coursera%20Web%20Promo%20Banner&mt=8','
'http://play.google.com/store/apps/details?id=org.coursera.android','
'https://about.coursera.org/','
'https://about.coursera.org/leadership','
'https://www.coursera.community','
'https://www.coursera.org/about/partners','
'https://building.coursera.org/developer-program/','
'https://www.coursera.community/beta-tester-news-announcements-20/join-the-coursera-beta-tester-community-47','
'https://www.coursera.community/gtc-news-announcements-17/join-the-coursera-global-translator-community-45','
'https://blog.coursera.org','
'https://www.amazon.com/Deep-Learning-Adaptive-Computation-Machine/dp/0262035618/ref=sr_1_1?ie=UTF8&qid=1472485235&sr=8-1&keywords=deep+learning+book','
'https://groups.google.com/forum/#!forum/deeplearningbook','
'https://github.com/goodfeli/dlbook_notation','
'https://docs.google.com/document/d/1ABlp7FluwZ0B82_fjNOFVQ2uOZkfuF8elbofhZmNXag/edit?usp=sharing','
'https://searchenterpriseai.techtarget.com','
'https://users.techtarget.com/registration/searchEnterpriseAI/LoginRegister.page','
'https://users.techtarget.com/registration/searchEnterpriseAI/Register.page','
'https://www.techtarget.com/network','
'https://searchenterpriseai.techtarget.com/info/news','
'https://searchenterpriseai.techtarget.com/features','
'https://searchenterpriseai.techtarget.com/tips','
'https://searchenterpriseai.techtarget.com/answers','
'https://whatis.techtarget.com/resources/Buyers-Guides','
'https://www.sas.com/en_us/home.html','
'https://www.sas.com/profile/user/contact.htm?locale=en_us&returnURL=https://www.sas.com/en_us/insights/analytics/deep-learning.html','
'https://www.sas.com/en_si/home.html','
'https://www.sas.com/es_ar/home.html','
'https://www.sas.com/en_au/home.html','
'https://www.sas.com/de_at/home.html','
'https://www.sas.com/en_be/home.html','
'https://www.sas.com/pt_br/home.html','
'https://www.sas.com/en_ca/home.html','
'https://www.sas.com/es_cl/home.html','
'https://deeplearning.mit.edu',

'https://github.com/lexfridman/mit-deep-learning',
'https://selfdrivingcars.mit.edu/deeptraffic/',
'https://www.youtube.com/watch?v=1L0TKZQcUtA&list=PLrAXtmErZgOeiKm4sgNOknGvNjby9efdf',
'https://twitter.com/lexfridman',
'https://www.youtube.com/channel/UCSHZKyawb77ixDdsGog4iWA',
'https://lexfridman.us2.list-manage.com/subscribe?u=2bd2649a06183adee599340a1&id=0c5b4f2d63',
'https://lexfridman.com',
'https://aws.amazon.com/?nc2=h_lg',
'https://console.aws.amazon.com/support/home?nc2=h_ql_cu',
'https://portal.aws.amazon.com/gp/aws/developer/registration/index.html?nc2=h_ct&src=default',
'https://docs.aws.amazon.com/index.html?nc2=h_ql_doc',
'https://aws.amazon.com/marketplace/?nc2=h_ql_mp',
'https://aws.amazon.com/ar/?nc1=h_ls',
'https://aws.amazon.com/id/?nc1=h_ls',
'https://aws.amazon.com/de/?nc1=h_ls',
'https://aws.amazon.com/?nc1=h_ls',
'https://aws.amazon.com/es/?nc1=h_ls',
'https://aws.amazon.com/fr/?nc1=h_ls',
'https://www.ibm.com/cloud/why-ibm',
'https://www.ibm.com/cloud/products',
'https://www.ibm.com/cloud/solutions',
'https://www.ibm.com/cloud/garage',
'https://www.ibm.com/cloud/pricing',
'https://www.ibm.com/cloud/partners',
'https://www.ibm.com/blogs/bluemix',
'https://cloud.ibm.com/docs',
'https://www.ibm.com/cloud/support',
'https://cloud.ibm.com/registration',
'https://dataplatform.cloud.ibm.com/registration/stepone?apps=data_science_experience&context=wdp',
'https://www.scientificamerican.com/store/subscribe/scientific-american-magazine/',
'https://www.scientificamerican.com/',
'https://www.scientificamerican.com/store/shopping-cart/',
'https://www.scientificamerican.com/my-account/forgot-password/',
'https://www.scientificamerican.com/my-account/register/',
'https://www.scientificamerican.com/page/newsletter-sign-up/',
'https://www.scientificamerican.com/the-sciences',
'https://www.scientificamerican.com/mind',
'https://www.scientificamerican.com/health',
'https://www.scientificamerican.com/tech',
'https://www.datacamp.com/community/podcast',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=1',

'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=2',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=3',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=4',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=5',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=6',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=7',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=8',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=9',
'https://campus.datacamp.com/courses/deep-learning-in-python/basics-of-deep-learning-and-neural-networks?ex=10',
'https://www.deeplearning.ai/deep-learning-specialization/',
'https://www.deeplearning.ai/ai-for-everyone/',
'https://www.deeplearning.ai/tensorflow-from-basics-to-mastery/',
'https://www.deeplearning.ai/bootcamp/',
'https://www.deeplearning.ai/ai-career-program-for-university-graduates/',
'https://www.deeplearning.ai/ai-career-program-for-experienced-engineers/',
'https://www.deeplearning.ai/events/',
'https://www.deeplearning.ai/forums/',
'https://www.deeplearning.ai/blog/',
'https://www.deeplearning.ai/about-us/',
'https://software.intel.com/en-us/user/login?destination=node/740582',
'https://software.intel.com/registration?lang=en-us',
'https://www.intel.ai/investing-in-the-pytorch-developer-community/',
'http://en.paddlepaddle.org/en',
'https://www.intel.ai/ngraph/',
'https://01.org/mkl-dnn/overview',
'https://www.intel.ai/plaidML/',
'https://www.intel.ai/hardware/',
'https://software.seek.intel.com/AIWebinarSeries?registration_source=IDZ',
'https://software.intel.com/en-us/ai/docs?search_api_views_fulltext=¤t_page=0&value=20780',
'https://www.intel.ai/publications/',
'https://www.nature.com/my-account/alerts/subscribe-journal?list-id=16',
'https://mts-nrg.nature.com/cgi-bin/main.plex',
'https://idp.nature.com/authorize/natureuser?client_id=grover&redirect_uri=https%3A%2F%2Fwww.nature.com%2Farticles%2Fs41576-019-0122-6',
'http://orcid.org/0000-0001-9579-2909',

'<http://orcid.org/0000-0002-2419-1943>',
'https://idp.nature.com/login/natureuser?client_id=grover&redirect_uri=http://www.nature.com/articles/s41576-019-0122-6',
'<https://idp-saml-nature-federated-login.live.cf.public.nature.com/saml/login?idp=https://idp.eduserv.org.uk/openathens&targetUrl=http://www.nature.com/articles/s41576-019-0122-6>',
'https://idp.nature.com/login/federated?redirect_uri=http://www.nature.com/articles/s41576-019-0122-6',
'<https://kundajelab.github.io/dragonn/tutorials.html>',
'<https://www.kaggle.com/sudalairajkumar/winning-solutions-of-kaggle-competitions>',
'<https://keras.io/>',
'https://en.wikipedia.org/wiki/Feedforward_neural_network',
'<http://h2o.ai/resources>',
'<https://htmlpreview.github.io/?https://github.com/ledell/sldm4-h2o/blob/master/sldm4-deeplearning-h2o.html>',
'<https://www.h2o.ai/blog/h2o-tensorflow-on-aws-gpu/>',
'<https://github.com/h2oai/sparkling-water/blob/master/py/examples/notebooks/TensorFlowDeepLearning.ipynb>',
'<https://www.youtube.com/watch?v=zGdXaRug7LI>',
'<https://www.youtube.com/watch?v=LM255qs8Zsk>',
'<https://www.youtube.com/watch?v=KWdkVoKJG3U>',
'https://en.wikipedia.org/wiki/Modulo_operation',
'<http://www.idg.pl/mirrors/CRAN/web/packages/gbm/vignettes/gbm.pdf>',
'http://premolab.ru/pub_files/pub88/qhkDNEyp8.pdf',
'<https://github.com/fastai/course-v3>',
'<http://www.fast.ai>',
'<http://docs.fast.ai>',
'<https://github.com/fastai/fastai/blob/master/README.md>',
'<http://course18.fast.ai>',
'<https://forums.fast.ai/t/recommended-python-learning-resources/26888>',
'<http://www.fast.ai/2017/11/16/what-you-need>',
'<https://www.fast.ai/2019/01/24/course-v3>',
'<https://docs.microsoft.com/azure/virtual-machine-scale-sets/virtual-machine-scale-sets-use-low-priority>',
'<https://forums.fast.ai>',
'<https://www.discourse.org/about>',
'<http://ufldl.stanford.edu/tutorial/>',
'<http://ufldl.stanford.edu/tutorial/StarterCode>']

"Big Data Analytics" neighbourhood set:

['<https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>',
'<https://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>',
'https://www.sas.com/en_us/insights/analytics/big-data-analytics.html',

'<https://www.ibm.com/analytics/hadoop/big-data-analytics>',
'<https://www.qubole.com/big-data-analytics/>',
'<https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>',
'<https://www.ngdata.com/what-is-big-data-analytics/>',
'<https://analyticstraining.com/5-skills-need-know-become-big-data-analyst/>',
'https://en.wikipedia.org/wiki/Big_data',
'<https://mixpanel.com/blog/2018/12/11/big-data-analytics/>',
'<https://www.techopedia.com/definition/28659/big-data-analytics>',
'https://www.webopedia.com/TERM/B/big_data_analytics.html',
'<https://www.marutitech.com/big-data-analytics-need-business/>',
'<https://www.sisense.com/product/big-data-analytics/>',
'<https://databricks.com/glossary/big-data-analytics>',
'<https://www.teradata.com/Outcomes/Big-Data>',
'<https://www.edx.org/course/big-data-analytics-adelaide-analyticsx>',
'<https://bigdataanalytics.mit.edu/>',
'<https://www.datamation.com/big-data/big-data-analytics.html>',
'<https://www.tibco.com/products/tibco-spotfire/big-data-analytics>',
'<https://www.informationbuilders.com/big-data-analytics>',
'<https://bdataanalytics.biomedcentral.com/>',
'<https://www.eweek.com/big-data-and-analytics>',
'<https://www.edureka.co/big-data-and-analytics>',
'<https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/>',
'<https://www.alteryx.com/solutions/analytics-need/advanced-analytics/database-analytics>',
'<https://www.extension.harvard.edu/course-catalog/courses/crn/15499>',
'https://www.tutorialspoint.com/big_data_analytics/',
'<https://www.linkedin.com/jobs/big-data-analytics-jobs>',
'<https://www.zdnet.com/topic/big-data/>',
'<https://www.informatica.com/products/big-data.html>',
'<https://www.ibmbigdatahub.com/>',
'<https://www.informationweek.com/big-data-analytics.asp>',
'https://www.splunk.com/en_us/solutions/solution-areas/big-data.html',
'<https://cloud.google.com/products/big-data/>',
'<https://cloud.google.com/solutions/big-data/>',
'<https://www.talend.com/solutions/information-technology/big-data-analytics/>',
'<https://pivot3.com/solution/big-data-analytics/>',
'<https://searchbusinessanalytics.techtarget.com>',
'<https://users.techtarget.com/registration/searchBusinessAnalytics/LoginRegister.page>',
'<https://users.techtarget.com/registration/searchBusinessAnalytics/Register.page>',
'<https://www.techtarget.com/network>',
'<https://searchbusinessanalytics.techtarget.com/info/news>',
'<https://searchbusinessanalytics.techtarget.com/features>',
'<https://searchbusinessanalytics.techtarget.com/tips>',
'<https://searchbusinessanalytics.techtarget.com/answers>',

'<https://whatis.techtarget.com/resources/Buyers-Guides>',
'https://www.sas.com/en_us/home.html',

'https://www.sas.com/profile/user/contact.htm?locale=en_us&returnURL=https://www.sas.com/en_us/insights/analytics/big-data-analytics.html',
'https://www.sas.com/en_si/home.html',
'https://www.sas.com/es_ar/home.html',
'https://www.sas.com/en_au/home.html',
'https://www.sas.com/de_at/home.html',
'https://www.sas.com/en_be/home.html',
'https://www.sas.com/pt_br/home.html',
'https://www.sas.com/en_ca/home.html',
'https://www.sas.com/es_cl/home.html',
'<https://www.ibm.com/us/en>',
'<https://www.ibm.com/sitemap/us/en>',
'<https://www.ibm.com/analytics/why-ibm>',
'<https://www.ibm.com/account/reg/signup?formid=urx-14629>',
'<https://www.ibm.com/marketplace/big-replicate>',
'<https://www.ibm.com/account/reg/signup?formid=urx-32397>',
'<https://developer.ibm.com/hadoop/>',
'https://twitter.com/IBMAalytics?ref_src=twsr%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor',
'<http://www.ibmbigdatahub.com/technology/hadoop-and-spark>',
'<https://www.simplilearn.com/resources>',
'<https://www.simplilearn.com/resources/agile-and-scrum>',
'<https://www.simplilearn.com/resources/big-data-and-analytics>',
'<https://www.simplilearn.com/resources/digital-marketing>',
'<https://www.simplilearn.com/resources/it-security-management>',
'<https://www.simplilearn.com/resources/it-service-management>',
'<https://www.simplilearn.com/resources/project-management>',
'<https://www.simplilearn.com/resources/salesforce-training>',
'<https://www.simplilearn.com/resources/virtualization-and-cloud-computing>',
'<https://www.simplilearn.com/resources/career-fast-track>',
'<https://www.simplilearn.com/resources/enterprise>',
'<https://www.ngdata.com/privacy-policy-terms-of-use/>',
'<https://www.ngdata.com/>',
'<https://www.ngdata.com/schedule-demo/>',
'<https://www.ngdata.com/platform/>',
'<https://www.ngdata.com/ngdata-cdp-4-0/>',
'<https://www.ngdata.com/services/>',
'<https://www.ngdata.com/solutions/>',
'<https://www.ngdata.com/resources/>',
'<https://www.ngdata.com/client-stories/>',
'<https://analyticstraining.com/>',
'<https://analyticstraining.com/category/machine-learning-ai/>',

'<https://analyticstraining.com/category/postgraduate-program-in-data-science-machine-learning/>',
'<https://analyticstraining.com/category/internet-of-things/>',
'<https://analyticstraining.com/category/careers-in-analytics/>',
'<https://analyticstraining.com/category/hr-analytics/>',
'https://analyticstraining.com/category/tools_techniques/',
'<https://analyticstraining.com/category/big-data-analytics/>',
'<https://analyticstraining.com/category/student-speak/>',
'<https://analyticstraining.com/all-time-videos/>',
'<https://analyticstraining.com/category/corporate-training/>',
'<https://en.wiktionary.org/wiki/practitioner>',
'<http://turbulence.pha.jhu.edu>',
'<https://scholar.google.com/scholar?cites=581440451196546619&sciodt=0,21>',
'<http://www.martinhilbert.net/WorldInfoCapacity.html>',
'<https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/>',
'<http://www.economist.com/node/15557443>',
'<http://www.nature.com/nature/journal/v455/n7209/full/455001a.html>',
'<http://adsabs.harvard.edu/abs/2008Natur.455....1.>',
'<http://adsabs.harvard.edu/abs/2011Sci...331..703R>',
'<http://gigaom.com/2008/11/09/mapreduce-leads-the-way-for-parallel-programming/>',
'<https://books.google.com/books?id=zxNglqU1FKgC>',
'<https://www.facebook.com/techopedia>',
'<https://www.linkedin.com/company/techopedia>',
'<https://twitter.com/techopedia>',
'https://www.techopedia.com/reg/managing-multiple-database-roles-how-many-hats-do-you-wear/33762?utm_source=techopedia&utm_medium=hellobar',
'https://www.techopedia.com/reg/how-to-continuously-monitor-and-analyze-mysql-and-mariadb-with-ideras-sql-diagnostic-manager/33743?utm_source=techopedia&utm_medium=hellobar',
'https://www.techopedia.com/resources?utm_source=techopedia&utm_medium=hellobar',
'<https://www.techopedia.com/14/33469/best-ways-to-extend-endpoint-management-and-security-to-mobile-devices>',
'<https://www.webopedia.com>',
'<https://twitter.com/WebopediaTech/>',
'<https://www.facebook.com/webopedia>',
'<https://www.webopedia.com/TERM>',
'https://www.webopedia.com/DidYouKnow/_index.asp',
'https://www.webopedia.com/quick_ref/index.asp',
'https://www.webopedia.com/Top_Category.asp',
'<https://www.webopedia.com/guides/index.asp>',
'<https://www.webopedia.com/Blog>',
'<https://www.webopedia.com/slideshows>',
'<https://www.marutitech.com/business-chatbot>',
'<https://www.marutitech.com/>',
'<https://www.marutitech.com/bot-development/>',

'https://www.marutitech.com/it-strategy-and-technology-consulting/','
'https://www.marutitech.com/web-and-mobile-development/','
'https://www.marutitech.com/big-data-and-analytics/','
'https://www.marutitech.com/salesforce-services/','
'https://www.marutitech.com/blog/','
'https://www.marutitech.com/ebooks/','
'https://www.marutitech.com/casestudy/','
'https://www.marutitech.com/white-paper/','
'https://databricks.com/ ',
'https://databricks.com/product',
'https://databricks.com/product/unified-analytics-platform',
'https://databricks.com/data-science-solutions#interactive-workspace',
'https://databricks.com/product/databricks-delta',
'https://databricks.com/mlflow',
'https://databricks.com/product/genomics',
'https://databricks.com/databricks-enterprise-security',
'https://databricks.com/product/pricing',
'https://databricks.com/product/azure',
'https://databricks.com/aws',
'https://twitter.com/datamation',
'https://www.facebook.com/pages/Datamation/43209128573',
'https://www.linkedin.com/company/datamation-com?trk=biz-companies-cym',
'https://www.datamation.com/rss.xml',
'https://www.datamation.com',
'https://www.datamation.com/big-data',
'https://www.datamation.com/big-data/big-data-companies.html',
'https://www.datamation.com/applications/15-top-data-analytics-tools.html',
'https://www.datamation.com/data-center/50-top-open-source-tools-for-big-data-1.html',
'https://www.datamation.com/big-data/big-data-technologies.html',
'https://www.datamation.com/big-data/big-data-startups.html',
'https://docs.tibco.com',
'https://support.tibco.com',
'https://community.tibco.com',
'https://docs.tibco.com/pub/spc/4.0.0/doc/html/index.htm',
'https://www.tibco.com/sites/tibco/files/resources/wp-big-data-analytics-turning-insight-into-action.pdf',
'https://www.tibco.com/sites/tibco/files/resources/ds-spotfire-terr-hadoop-final.pdf',
'https://www.facebook.com/TIBCO',
'https://twitter.com/tibco',
'http://education.informationbuilders.com/ ',
'http://forums.informationbuilders.com/eve/forums',
'https://techsupport.informationbuilders.com/public/tc-library.html',
'https://techsupport.informationbuilders.com/ ',
'https://www.informationbuilders.com/blog',

'<https://www.facebook.com/informationbuilders>',
'<https://twitter.com/infobldrs>',
'<https://www.linkedin.com/company/information-builders/>',
'<https://www.youtube.com/user/informationbuilders>',
'<https://www.instagram.com/informationbuilders/>',
'<https://www.biomedcentral.com>',
'<https://www.biomedcentral.com/login>',
'<http://www.editorialmanager.com/bdan>',
'<https://bdataanalytics.biomedcentral.com/articles/10.1186/s41044-017-0028-2>',
'<http://blogs.biomedcentral.com/on-biology/2016/07/01/big-data-analytics-editor-qa/>',
'<https://www.biomedcentral.com/add-article-alert?journalId=41044>',
'<https://twitter.com/BioMedCentral>',
'<https://www.biomedcentral.com/p/bmc-impact>',
'<https://twitter.com/eWEEKNews>',
'<http://www.facebook.com/pages/eWEEK/49650732622>',
'https://www.linkedin.com/company/eweek-washington-bureau?trk=company_name',
'<https://www.youtube.com/user/eweeknews>',
'<https://www.eweek.com/big-data-and-analytics/why-telecoms-are-at-leading-edge-of-analytics-innovation>',
'<https://www.eweek.com/Authors/chris-preimesberger>',
'<https://www.eweek.com/storage/toshiba-memory-launches-new-sas-ssds-for-hpe-servers>',
'<https://www.eweek.com/networking/why-cisco-customers-should-embrace-new-smart-licensing>',
'<http://www.edureka.co>',
'<http://www.edureka.co/all-courses>',
'<http://www.edureka.co/big-data-and-analytics>',
'<https://www.edureka.co/blog/dataframes-in-spark/>',
'<https://www.edureka.co/blog/rdd-using-spark/>',
'<https://www.edureka.co/blog/spark-architecture/>',
'<https://www.edureka.co/blog/category/big-data-analytics/>',
'<https://www.edureka.co/community>',
'<https://itunes.apple.com/in/app/edureka/id1033145415?mt=8>',
'<https://play.google.com/store/apps/details?id=co.edureka.app>',
'<https://www.facebook.com/edurekaIN>',
'<https://www.forbes.com/worlds-billionaires/>',
'<https://www.forbes.com/billionaires/>',
'<https://www.forbes.com/forbes-400/>',
'<https://www.forbes.com/self-made-women/>',
'<https://www.forbes.com/china-billionaires/>',
'<https://www.forbes.com/india-billionaires/>',
'<https://www.forbes.com/indonesia-billionaires/>',
'<https://www.forbes.com/korea-billionaires/list/>',
'<https://www.forbes.com/thailand-billionaires/list/>',
'<https://www.forbes.com/japan-billionaires/list/>',
'<https://pages.alteryx.com/Alteryx-Demo-Q4-2016-W.html>',

'<https://pages.alteryx.com/alteryx-trial.html>',
'<https://www.alteryx.com/contact-us>',

'https://www.alteryx.com/resources?field_industry_target_id=All&field_department_target_id=All&field_resource_type_target_id=151&title=',
'<https://community.alteryx.com/t5/Support-Page/bd-p/SupportPage>',
'<http://licenses.alteryx.com/>',
'<https://community.alteryx.com/t5/Alteryx-Academy/ct-p/alteryx-academy>',
'<https://community.alteryx.com/>',
'<https://www.tutorialspoint.com/whiteboard.htm>',
'<https://www.tutorialspoint.com/netmeeting.php>',
'<https://www.facebook.com/tutorialspointindia>',
'<https://plus.google.com/u/0/116678774017490391259/posts>',
'<https://www.twitter.com/tutorialspoint>',
'<https://www.linkedin.com/company/tutorialspoint>',
'<https://www.youtube.com/channel/UCVLbzhxVTiTLiVKeGV7WEBg>',
'<https://store.tutorialspoint.com/>',
'<http://tools.tutorialspoint.com/bda>',
'https://www.linkedin.com/legal/cookie-policy?trk=guest_job_search_cookie_policy_click',
'https://www.linkedin.com/jobs?trk=guest_job_search_logo',
'https://www.linkedin.com/uas/login?emailAddress=&fromSignIn=&trk=guest_job_search_signin',
'https://www.linkedin.com/start/join?source=jobs_registration&trk=guest_job_search_join',
'https://www.linkedin.com/jobs/view/amazon-web-services-big-data-solutions-and-development-engineering-redshift-at-amazon-932739163?position=1&pageNum=0&trk=jobs_jserp_job_listing_text',
'https://www.linkedin.com/jobs/view/senior-big-data-architect-at-perficient-1240206771?position=2&pageNum=0&trk=jobs_jserp_job_listing_text',
'https://www.linkedin.com/jobs/view/lead-big-data-engineer-at-diligente-technologies-1237299606?position=3&pageNum=0&trk=jobs_jserp_job_listing_text',
'https://www.linkedin.com/jobs/view/senior-software-engineer-big-data-at-five9-1231765696?position=4&pageNum=0&trk=jobs_jserp_job_listing_text',
'https://www.linkedin.com/jobs/view/data-analyst-at-techdigital-corporation-1162376133?position=5&pageNum=0&trk=jobs_jserp_job_listing_text',
'https://www.linkedin.com/jobs/view/senior-data-engineer-big-data-at-zillow-group-1087038147?position=6&pageNum=0&trk=jobs_jserp_job_listing_text',
'https://www.linkedin.com/jobs/view/big-data-engineer-at-amazon-web-services-aws-1171379447?position=7&pageNum=0&trk=jobs_jserp_job_listing_text',
'<http://www.zdnet.com.cn/>',
'<http://www.zdnet.fr/>',
'<http://www.zdnet.de/>',
'<http://www.zdnet.co.kr/>',
'<http://japan.zdnet.com/>',
'<https://www.zdnet.com/>',
'<http://www.techproresearch.com/>',

'<http://www.zdnet.com/article/vpn-services-2018-the-ultimate-guide-to-protecting-your-data-on-the-internet/>',
'<https://www.techrepublic.com/resource-library/whitepapers/>',
'<https://www.techrepublic.com/forums/>',
'<https://www.zdnet.com/user/login/?appId=1101>',
'<https://www.informatica.com/legal/cookie-policy.html>',
'<https://www.informatica.com/>',
'<https://www.informatica.com/solutions.html>',
'<https://www.informatica.com/solutions/move-to-the-cloud.html>',
'<https://www.informatica.com/solutions/move-to-the-cloud/accelerate-business-insights.html>',
'<https://www.informatica.com/solutions/move-to-the-cloud/deliver-fast-cloud-innovation.html>',
'<https://www.informatica.com/solutions/move-to-the-cloud/integrate-cloud-applications.html>',
'<https://www.informatica.com/solutions/move-to-the-cloud/manage-hybrid-data-complexity.html>',
'https://www.informatica.com/lp/6-ways-to-deliver-business-values-cdw_3460.html',
'<https://www.informatica.com/solutions/data-governance-and-compliance.html>',
'<http://twitter.com/IBMAalytics>',
'<https://www.linkedin.com/company/ibm-big-data-&-analytics>',
'<https://www.facebook.com/IBManalytics>',
'<https://www.ibmbigdatahub.com/blog/ai-everywhere-modern-predictive-analytics>',
'<https://www.ibmbigdatahub.com/podcast/making-data-simple-spark-hadoop-and-more-jean-georges-perrin>',
'<https://www.ibmbigdatahub.com/blog/tackle-customer-churn-spss-modeler>',
'<https://www.ibmbigdatahub.com/blog/jpmc-teams-ibm-data-science-and-ai-elite>',
'<https://www.splunk.com>',
'http://login.splunk.com/page/sso_redirect?type=portal',
'<https://answers.splunk.com/>',
'<http://docs.splunk.com>',
'<https://www.splunk.com/page/securityportal>',
'<https://answers.splunk.com>',
'<https://www.splunk.com/blog/homepage.html>',
'https://www.splunk.com/en_us/community.html',
'<http://conf.splunk.com/>',
'<http://dev.splunk.com/>',
'<http://docs.splunk.com/Documentation>',
'<https://cloud.google.com/>',
'<https://cloud.google.com/why-google-cloud/>',
'<https://cloud.google.com>',
'<https://cloud.google.com/security/>',
'<https://cloud.google.com/open-cloud/>',
'<https://cloud.google.com/customers/>',
'<https://cloud.google.com/analyst-reports/>',
'<https://talend.channeltivity.com/Preview/BecomeAPartner/>',
'<https://community.talend.com/>',
'<http://help.talend.com/>',

'http://investor.talend.com/','
'https://www.talend.com/','
'https://fr.talend.com/','
'https://de.talend.com/','
'https://it.talend.com/','
'https://jp.talend.com/','
'https://es.talend.com/']