ECE/CS 498 DS HW 1
Spring 2020

Name: Gowtham Kuntumalla

Netid: gowtham4

Registration Status: registered for 4 credits, ECE 498

A. Data Structure to Parse Raw Log File

- Provide a (i) diagram and (ii) brief explanation of the data structure you used to parse the raw log file

```
(('1506816069251', 'firefox', '13179', '0x282235aae', 'R', 'minor', '50'), [['</usr/lib/x86_64-linux-gnu/libcairo.so.
2.11400.10', '16727808', '686943'], ['</lib/x86_64-linux-gnu/libglib-2.0.so.0.5400.1', '16748032', '660132'], ['</li
b/x86_64-linux-gnu/libpthread-2.26.so', '16746496', '483292'], ['</usr/lib/x86_64-linux-gnu/libX11.so.6.3.0', '167488
00', '108902'], ['</lib/x86_64-linux-gnu/libc-2.26.so', '16767232', '7501']])
(('1506816074664', 'firefox', '13179', '0x10fb420b4', 'W', 'minor', '23'), [['</lib/x86_64-linux-gnu/libglib-2.0.so.
0.5400.1', '16657664', '64606'], ['</usr/lib/x86_64-linux-gnu/libxcb.so.1.1.0', '16756992', '607425'], ['</lib/x86_64
-linux-gnu/libpthread-2.26.so', '16759296', '459484'], ['</lib/x86_64-linux-gnu/libc-2.26.so', '16723456', '528024'],
['</usr/lib/x86_64-linux-gnu/libgdk-x11-2.0.so.0.2400.31', '16689664', '40592'], ['</usr/lib/x86_64-linux-gnu/libX11.
so.6.3.0', '16659456', '385253'], ['</usr/lib/x86_64-linux-gnu/libcairo.so.2.11400.10', '16731648', '161826']])
(('1506816094793', 'thunderbird', '5914', '0x3e9e414bb', 'W', 'minor', '11'), [['</usr/lib/x86_64-linux-gnu/libxcb.s
o.1.1.0', '16716800', '619231']])
(('1506816166103', 'watchdog', '7518', '0x2a537754b', 'R', 'minor', '17'), [['</lib/x86_64-linux-gnu/libglib-2.0.so.
0.5400.1', '16755200', '381200'], ['</usr/lib/x86_64-linux-gnu/libX11.so.6.3.0', '16756480', '490547'], ['</lib/x86_6
4-linux-gnu/libc-2.26.so', '16723456', '463886'], ['</usr/lib/x86_64-linux-gnu/libcairo.so.2.11400.10', '16720640',
'79241'], ['</usr/lib/x86_64-linux-gnu/libgdk-x11-2.0.so.0.2400.31', '16706816', '124380'], ['</lib/x86_64-linux-gnu/
libpthread-2.26.so', '16758272', '751319'], ['</usr/lib/x86_64-linux-gnu/libxcb.so.1.1.0', '16721152', '184707']])
(('1506816180376', 'thunderbird', '5914', '0x1bf19b90a', 'W', 'minor', '12'), [['</usr/lib/x86_64-linux-gnu/libgdk-x1
1-2.0.so.0.2400.31', '16675328', '592929']])
```

- Dictionary to hold a list of backtraces for each pagefault
  - key = tuple (pagefault details) -- Tuple
  - value = list (backtraces) -- 2D list

B.a. Time Range Covered By Data

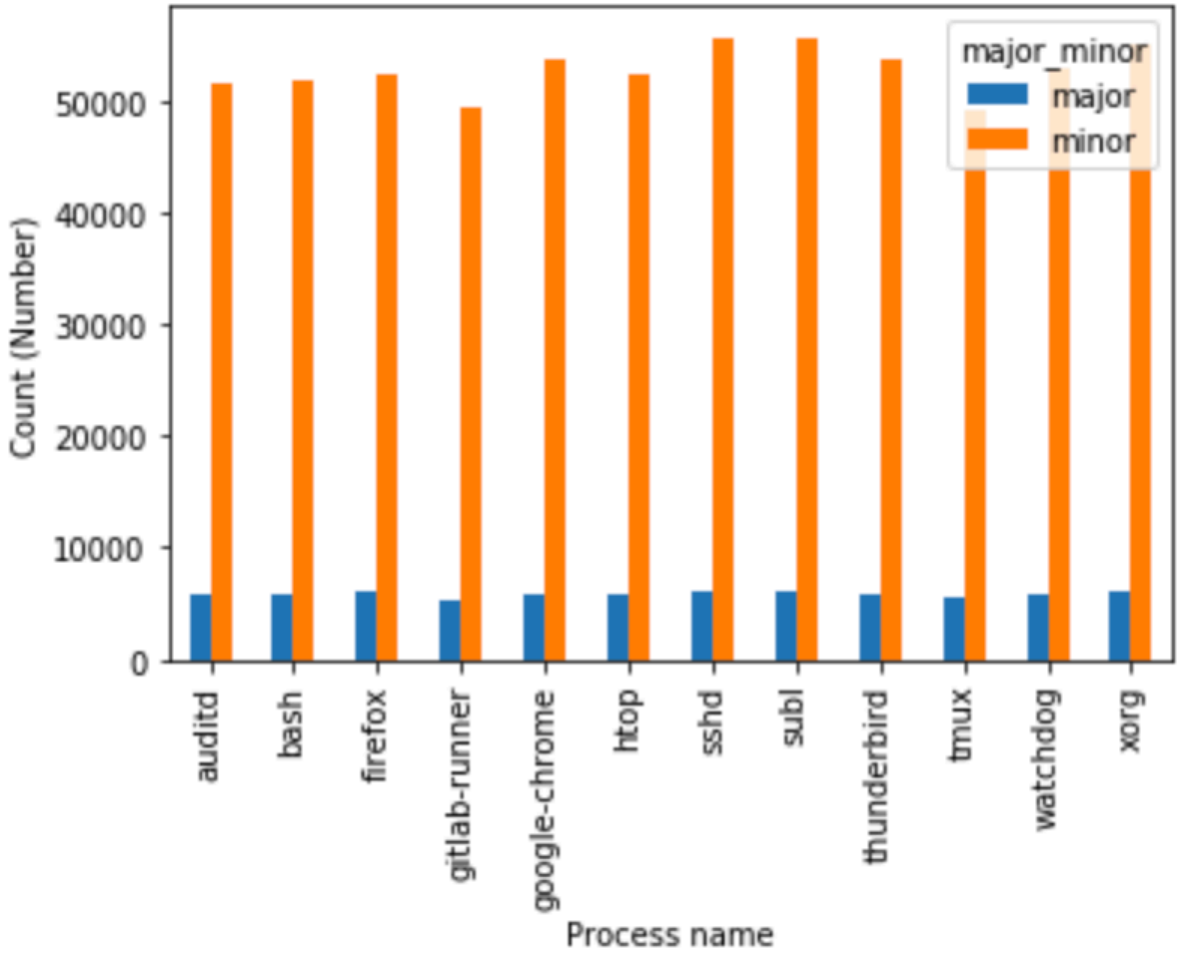
- Start Time: 2017-10-01 00:01:09.251000
- End Time: 2018-01-07 18:59:50.839000
- Total Duration: Timedelta('98 days 18:58:41.588000')

B.b. Unique Processes

- The number of unique processes :12
- The name of each process: ['firefox' 'thunderbird' 'watchdog' 'auditd' 'subl' 'gitlab-runner' 'sshd' 'google-chrome' 'bash' 'tmux' 'xorg' 'htop']
- The number of times each process was executed:

1. auditd 228982
2. bash 229904
3. firefox 233452
4. gitlab-runner 218405
5. google-chrome 238107
6. htop 232215
7. sshd 246903
8. subl 245982
9. thunderbird 237590
10. tmux 219329
11. watchdog 234938
12. xorg 243924

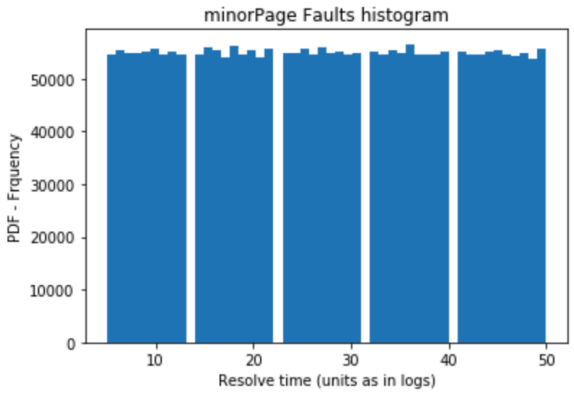
B.c. Major and Minor Page Faults



B.d. Time to Resolve Page Faults

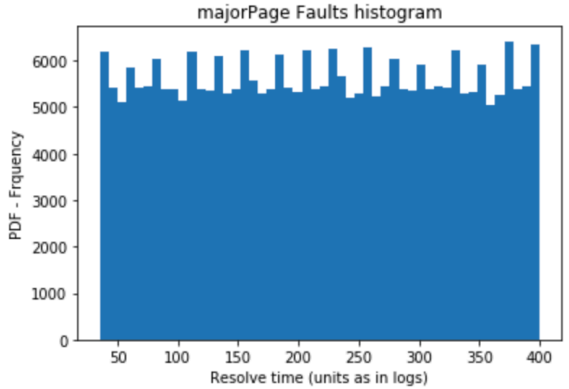
aggregate results of minor page faults are:

proc_name	mean	std
auditd	27.522918	13.291775
bash	27.429991	13.285091
firefox	27.586621	13.292283
gitlab-runner	27.401360	13.241882
google-chrome	27.486030	13.217610
htop	27.359186	13.283209
sshd	27.541563	13.300029
subl	27.426151	13.238984
thunderbird	27.458199	13.254275
tmux	27.421387	13.280019
watchdog	27.615395	13.272543
xorg	27.528463	13.263153



aggregate results of major page faults are:

proc_name	mean	std
auditd	218.191273	105.493006
bash	219.380514	105.209861
firefox	219.960241	104.477581
gitlab-runner	213.306051	105.639431
google-chrome	218.813443	104.972038
htop	219.088798	104.744666
sshd	216.805671	105.636435
subl	215.073090	106.021568
thunderbird	219.667928	106.878117
tmux	219.742614	105.805424
watchdog	215.778328	105.352504
xorg	216.892856	105.984679



C.a. Class Priors

'firefox':	0.082941079376643 48,
'thunderbird':	0.08451199244140381,
'watchdog':	0.08372369005202226,
'auditd':	0.08137016631188317,
'subl':	0.08786014320352431,
'gitlab-runner':	0.07761078921306362,
'sshd':	0.08782456999100709,
'google-chrome':	0.08480084692704361,
'bash':	0.08171451500904983,
'tmux':	0.0777786947761449,
'xorg':	0.08690108939406013,
'htop':	0.08296242

C.b. – C.c. : Predictions

- Given that the page fault was major, which process was it most likely caused by?

- **Process that is likely to occur: subl , 0.0921**
- Given that the page fault was from a read access, which process was it most likely caused by?
- **Process that is likely to occur: subl , 0.0923**

C.d. Appropriate Model

- In 2 sentences or less, explain which model taught in class could be used for classifying the process given information about the fault's (i) severity and (ii) access type.

Ans) We would use Naive Bayes Classifier for classifying. It is a decent first estimate considering the fact: independence of read/write operations and type of fault (major/minor)