

Machine Configuration for the experiments ----

OS – Ubuntu 20.04 64 bit kernel version 5.11.0-25-generic

CPU – AMD Ryzen 5 4600H 3.0 Ghz 6 Cores 12 Threads

RAM – 8 GB

PYTHON

Tool	Accuracy (%)	Average Response Time (in Milli Seconds)	Storage requirement (in MBs)
PyCln	99.29	0.1755	Not Applicable
AutoFlake	99.29	0.2280	Not Applicable

Both PyCln and AutoFlake rely on code analysis rather than matching with a white list, hence Storage requirement is not applicable on them

Executions of both tools are as:

pycln -a -d 'file location without quotes'

autoflake --remove-all-unused-imports --recursive 'file location without quotes'

3 Projects have been selected from GitHub at random consisting of 48 .py files (including init.py files as well)

To calculate metrics, total number of imports are obtained from all .py files and then based on whether an import is actually used and whether the tool correctly removes/flags the unused ones the following are defined:

AU – Import actually used

AUN – Import actually unused

DU – Import detected as used (better say not flagged as unused) by the tool

DUN – Import detected as unused by the tool

In essence a confusion matrix is formed

Total imports = 283 (each individual module imported is consider.

Consider the statement “ import sys,math,time “. This statement is equivalent to 3 separate imports

AU/DU ---> a True Positive (detected and is actually used in the file)

AU/DUN ---> a False Negative (actually used but incorrectly flagged as unused)

AUN/DU ---> a False Positive (actually unused but incorrectly missed by the tool)

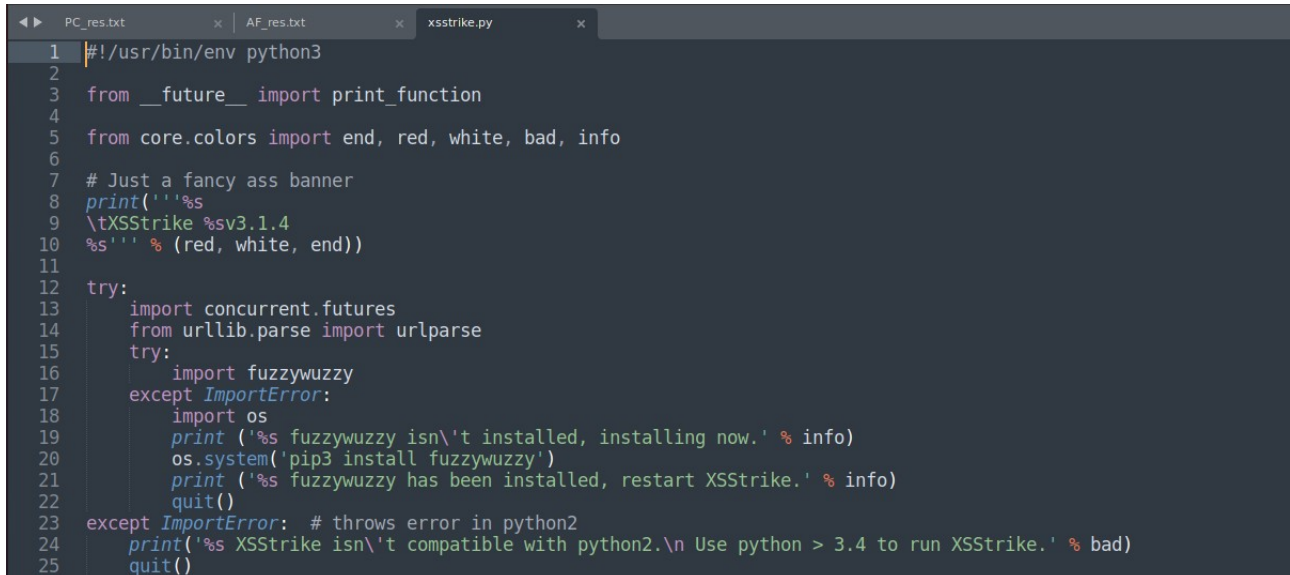
AUN/DUN ---> a True Negative (actually unused and correctly flagged as unused)

Accuracy = $\frac{TP+TN}{\text{Total Obserations}}$

PyCln	DU	DUN
AU	259	0
AUN	2	22

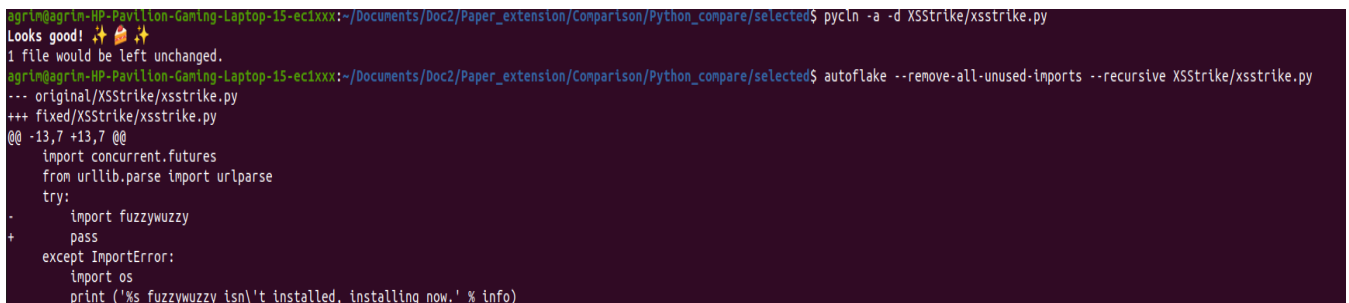
AutoFlake	DU	DUN
AU	258	0
AUN	2	23

**** Both produce near identical results but there are differences. One example is of the case of the file XSSStrike/xsstrrike.py. Particularly, the line 16 of this file is treated differently by AutoFlake and PyCln.



```
1#!/usr/bin/env python3
2
3from __future__ import print_function
4
5from core.colors import end, red, white, bad, info
6
7# Just a fancy ass banner
8print('%s
9\tXSSStrike %sv3.1.4
10%s' % (red, white, end))
11
12try:
13    import concurrent.futures
14    from urllib.parse import urlparse
15    try:
16        import fuzzywuzzy
17    except ImportError:
18        import os
19        print ('%s fuzzywuzzy isn\'t installed, installing now.' % info)
20        os.system('pip3 install fuzzywuzzy')
21        print ('%s fuzzywuzzy has been installed, restart XSSStrike.' % info)
22        quit()
23except ImportError: # throws error in python2
24    print('%s XSSStrike isn\'t compatible with python2.\n Use python > 3.4 to run XSSStrike.' % bad)
25    quit()
```

Image from XSSStrike/xsstrrike.py. Line no. 16 is the focus.



```
agrin@agrin-HP-Pavilion-Gaming-Laptop-15-ec1xxx:~/Documents/Doc2/Paper_extension/Comparison/Python_compare/selected$ pycln -a -d XSSStrike/xsstrrike.py
Looks good! 🎉 🎉 🎉
1 file would be left unchanged.
agrin@agrin-HP-Pavilion-Gaming-Laptop-15-ec1xxx:~/Documents/Doc2/Paper_extension/Comparison/Python_compare/selected$ autoflake --remove-all-unused-imports --recursive XSSStrike/xsstrrike.py
--- original/XSSStrike/xsstrrike.py
+++ fixed/XSSStrike/xsstrrike.py
@@ -13,7 +13,7 @@
     import concurrent.futures
     from urllib.parse import urlparse
     try:
-        import fuzzywuzzy
+        pass
     except ImportError:
         import os
         print ('%s fuzzywuzzy isn\'t installed, installing now.' % info)
```

Output of PyCln and AutoFlake differ in this case, as is evident from above image. PyCln does not flag the import at line 16 as unused while AutoFlake flags it.

**** Both the tools miss the case where when a file A.py imports from B.py as:
from B import *
and not all the modules imported in B are used up by A.

Case in point XSSStrike/core/log.py file's imports. The log.py file imports all modules from colors.py in the same directory (by from.colors using *).Now colors.py itself imports and uses the modules **sys**,**os** and **platform**. However, logs.py uses only the sys module imported from colors.py . Therefore despite 3 imports only 1 is used , 2 are unused, but both the tools don't flag this.

```

log.py
1 import logging
2 from .colors import *

```

Log.py's imports, only sys module is directly utilized (exactly at line 162 and 167 of the file).

```

log.py
1 import sys
2 import os
3 import platform

```

colors.py's imports , all of these are used.

```

agrin@agrin-HP-Pavillon-Gaming-Laptop-15-ec1xxx:~/Documents/Doc2/Paper_extension/Comparison/Python_compare/selected$ pycln -a -d XSStrike/core/log.py
Looks good! 🎉👏🎉
1 file would be left unchanged.
agrin@agrin-HP-Pavillon-Gaming-Laptop-15-ec1xxx:~/Documents/Doc2/Paper_extension/Comparison/Python_compare/selected$ autoflake --remove-all-unused-imports --recursive XSStrike/core/log.py

```

Both PyCln and AutoFlake don't flag the imports in log.py