Better Code, Better Sharing: On the Need of Analyzing Jupyter Notebooks

**Authors**; It was introduced by Fernando Pérez , Jiawei Wang who was from Monash University , Li Li from Monash University and Andreas Zeller from CISPA Helmholtz Center for Information Security.

**Introduction**; Jupyter notebooks is an software which is used for new or basic programmers who cannot code properly have to work on scientific field it makes code easy for the students and the scientist to work on their field and also store them on computer , jupyter is an open source environment where dozens of languages can be implemented it was made from I Python it is also awarded many times this type of notebook helped many students and new programmers to be advance in programming .following people who felt that there is a need for this kind of software and were chosen to introduced jupyter . jupyter can also be described as open source web application allowing users to write documents containing important text for them ,equations and visualizations , and also as well coding which is executable , Jupyter now has became very popular as it provides workbook for sharing information just like GitHub , now a days we can say it is compulsory notebook for scientist for data scientist to analyze although jupyter is also best for poor programming as it aims to get skills in programming

Summary

Jupyter notebook is an open-source web application which points towards the scientist and students to get advance in coding and making the scientific data analysis easier for the customer ,jupyter data is proceed by code which is converted into .py and then output is given by Jupiter although the code will not appear in python scripts , its called the Jupyter processing chain after that you can save the coding or your work in jupyter library as notebook it is referred as “library code” main idea of jupyter is to work easily on “statistics”. if Jupiter selected 1982 contains total of 202,332 lines of python code it explains that the code if distributed gives median and average lines of 62 and 102.5 lines , which contains 10 code cells

Further more jupyter converts python code in pythons scripts it shows that ideally jupyter is made to practice on python as 1982 jupyter corresponds to 202,332 lines of python code if PEP8 is run than it gives 73,371 errors which shows that notebook are not well aligned with the recommended code practices as this table shows,

Table

Python remark Notebook remark

E501 line too long E231 missing whitespace after ,, ;, or: E231 missing whitespace after ,, ;, or : E501 line too long

W291 trailing whitespace W293 blank line contains whitespace W293 blank line contains whitespace W291 trailing whitespace

E111 indentation is not a multiple of four E225 missing whitespace around operator E201 whitespace after ( E251 unexpected spaces around keyword parameter equals

E265 block comment should start with # E703 statement ends with a semicolon

E302 expected 2 blank lines, found 0 E261 at least two spaces before inline comment

E225 missing whitespace around operator E265 block comment should start with # comment

E251 unexpected spaces around keyword / parameter equals E128 continuation line under-indented for visual indent

Methodology

Now comes the part for methodology jupyter illustrates the working process , It builds an abstract syntax tree (ATS) for the involving code

Example import numpy as np

a = np.arange(15)

print(a)

print(a.shape)

this code with each variable is associated with special node indicating it is introduced into context or we can say it is referenced by the context and secondly jupyter performs transversal over the AST which separates all the variables that are associated with the context now jupyter application works as if a variable is stored but not loaded subsequently , it will be consider as unused variable by the flag . if it is recalled then now it will be considered as used . now this experiment shows that 803 notebooks out of 1982 contains code with unused variables

Now the team jupyter notebooks contain code which access the deprecated functions of the library as it if for the educational purposes but the team says that deprecated functions should be avoided .now in order to check jupyter team checks that if the deprecated functions are used or not they make them rely on ground truth of deprecated functions although the ground truth functions are not directly connected to the deprecated functions but the library maintainer is used to describe the changes if the deprecated functions , by simplifying these notes user will easily be able to collect a ground truth of deprecated functions.

Aim and result

Jupiter notebooks as mentioned above help students and Scientist to get advance in coding it is considered as educating application or we can say it is made only for beginners as it will harm the community if not changed but it is also mandatory for poor programming skills and technical debts .

Jupyter is used for the educational purposes as it helps the learners

1. Enforcing good coding styles.  
   jupyter teams made this application to help learners get some better skills in coding and points out mistakes which might help to make better programming skills and styles
2. Improving code quality and reliability  
   automated tools made within the application helps to point out poor coding style which makes the code more reliable and it also recommend to fix the error and also indicates the duplicated codes and inefficient algorithm and much more

Jupyter application team further aims to first investigate the necessities of the application by researching and views by the community .