

APIScanner - Towards Automated Detection of Deprecated APIs in Python Libraries

Aparna Vadlamani, Rishitha Kalicheti, Sridhar Chimalakonda
Research in Intelligent Software & Human Analytics (RISHA) Lab
Dept. of Computer Science & Engineering
Indian Institute of Technology Tirupati
Tirupati, India
{cs17b005, cs17b014, ch}@iittp.ac.in

Abstract—Python libraries are widely used for machine learning and scientific computing tasks today. APIs in Python libraries are deprecated due to feature enhancements and bug fixes in the same way as in other languages. These deprecated APIs are discouraged from being used in further software development. Manually detecting and replacing deprecated APIs is a tedious and time-consuming task due to the large number of API calls used in the projects. Moreover, the lack of proper documentation for these deprecated APIs makes the task challenging. To address this challenge, we propose an algorithm and a tool *APIScanner* that automatically detects deprecated APIs in Python libraries. This algorithm parses the source code of the libraries using abstract syntax tree (ASTs) and identifies the deprecated APIs via *decorator*, *hard-coded warning* or *comments*. *APIScanner* is a Visual Studio Code Extension that highlights and warns the developer on the use of deprecated API elements while writing the source code. The tool can help developers to avoid using deprecated API elements without the execution of code. We tested our algorithm and tool on six popular Python libraries, which detected 838 of 871 deprecated API elements. Demo of *APIScanner*: https://youtu.be/1hy_ugf-iek. Documentation, tool, and source code can be found here: <https://rishitha957.github.io/APIScanner>.

Index Terms—Deprecated APIs, Python Libraries, API Evolution, Visual Studio Code Extension

I. INTRODUCTION

Python is one of the popular dynamic programming language that has gained immense popularity due to its extensive collection of libraries, including popular modules for machine learning and scientific computing¹. Due to reasons such as feature improvements and bug repairs, python libraries are frequently updated. Most API changes include moving *methods* or *fields* around and *renaming* or *changing* method signatures [1]. These changes may induce compatibility issues in client projects [2]. It is recommended to follow the *deprecate-replace-remove* cycle to enable developers to adapt to these changes smoothly[3]. In this process, APIs that are no longer supported are first labeled as deprecated, and then the deprecated APIs are replaced with their substitution messages to help developers transition from deprecated APIs to new ones [4]. The deprecated APIs are gradually removed from the library in future releases. Unfortunately, this process is not always followed, as discovered by several studies [5], [6],

making it difficult for both library maintainers and developers. Ko et al. have analyzed the quality of documentation for resolving deprecated APIs [7]. Researchers have proposed techniques to automatically update deprecated APIs [8], [9]. However, most of them are for static programming languages such as *Java*, *C#* and *Android SDKs*. Python being a typical dynamic programming language, exhibits different API evolution patterns compared to *Java* [2]. Hence it motivates the need for new techniques and tools to detect deprecated APIs.

Deprecated APIs in Python libraries are mainly declared by *decorator*, *hard-coded warning*, and *comments* [10]. Nevertheless, it was discovered that library maintainers use varied and multiple strategies for API deprecation, leading to inconsistency in the implementation of libraries as well as their automated detection [10]. In addition, nearly one-third of the deprecated APIs in Python is not included in the official library documentation, making it hard for developers using libraries to limit the use of deprecated APIs [10].

To avoid the usage of deprecated APIs during new software development, developers should be aware of deprecating APIs in the project, motivating the need for this research. Hence, given the rise in popularity of Python and the number of deprecated APIs used in Python projects, we propose a novel algorithm that uses the source code of the Python libraries to get a list of deprecated APIs. This list is further used to detect deprecated APIs in Python projects. This paper contributes (i) an algorithm for deprecated API detection and (ii) a Visual Studio Code extension, *APIScanner*². We believe that *APIScanner* might assist developers to detect deprecated APIs and help them avoid searching through API documentation or on forums such as Stack Overflow. As a preliminary evaluation, we tested our algorithm and tool on six popular Python libraries [11] that are commonly used in data analytics, machine learning, and scientific computing. The initial results are promising with 90% API deprecation detection, with potential for application beyond these libraries.

II. APPROACH

Wang et al. [10] investigated that inconsistency in the adopted deprecation strategies makes it a harder task to use

¹<https://www.tiobe.com/tiobe-index/>

²<https://marketplace.visualstudio.com/items?itemName=Rishitha.apiscanner>

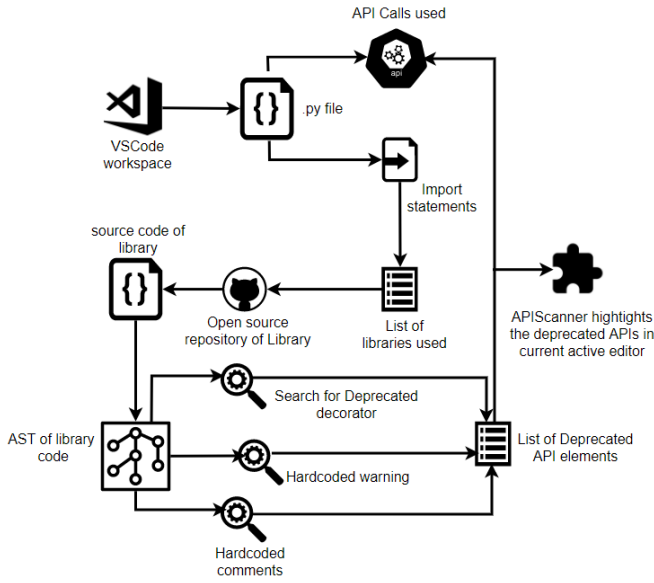


Fig. 1. Approach for Detecting Deprecated API Elements in Python Libraries

automated approaches for managing deprecated APIs and their documentation. In this paper, we propose an approach (as shown in Fig. 1) to automatically detect deprecated APIs in Python libraries and alert developers during API usage in software development. Firstly, we identify the libraries used in the client code from *import* statements. We build an abstract syntax tree (AST) to parse the source code to detect the patterns. The proposed Algorithm 1 is then applied on the ASTs to retrieve a list of deprecated APIs in those libraries. Based on this list, *APIScanner* parses each line of code in the editor, highlights the deprecated elements in the editor. On hovering, the tool also displays a message informing the developer that some element(s) of this API call has been deprecated (as shown in Fig. 2). We developed *APIScanner* as a Visual Studio Code extension as it supports both Python scripts and *jupyter notebooks*³.

```
(a) Using Decorator: in Matplotlib
@api.deprecated("3.3", alternative="Glue('fil')")
class Fil(Glue):
    def __init__(self):
        super().__init__('fil')

(b) Using Comments: in Sklearn
class GradientBoostingClassifier(args):
    """
    .. criterion : {'friedman_mse', 'mse', 'mae'}..
    .. deprecated: 0.24 'criterion='mae' is deprecated and will be removed in
    version 0.26. Use 'criterion='friedman_mse' or 'mse' instead, as trees
    should use a least-square criterion in Gradient Boosting
    """

(c) Using Hardcoded Warnings: in Pandas
class Series(args):
    def __init__(self, args):
        if dtype is None:
            warnings.warn("The default dtype for empty Series will be 'object'
            instead of 'float64' in a future version", DeprecationWarning, stacklevel=2)
```

Listing 1. Examples of methods of deprecation strategies adopted in Python libraries which are deprecated through a) *decorator*, b) *comments* c) *hard-coded warning*

A. Detecting Deprecated API Elements through Source Code

We parse the source code of the library to generate an AST and denote it as P_{AST} . Examples of Python APIs deprecated

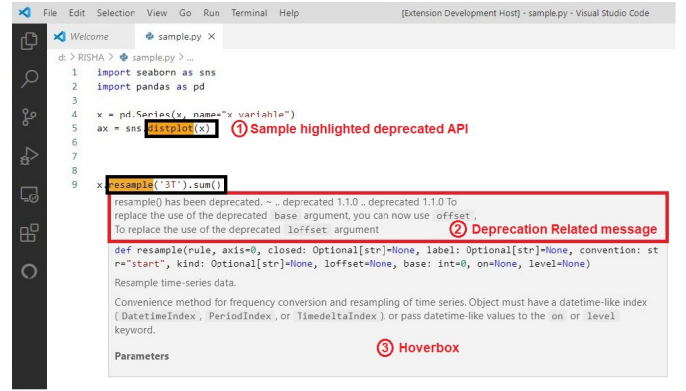


Fig. 2. Snapshot of *APIScanner*. The black boxes indicate deprecated APIs highlighted by *APIScanner*. The red box indicates the message shown by *APIScanner* on hovering over the highlighted deprecated APIs.

by *decorator*, *hard-coded warnings*, and *comments* are shown in listing 1. Structure of AST helps to realize the relationship between class declaration and function definition with *decorator*, *hard-coded warnings*, and *comments*. We traverse through each node N_{AST} in the AST and generate P_{AST} using Depth-First Search (cf. Line-2). Whenever we encounter a class definition node, we extract the *doc-string* of that particular class. If the *doc-string* contains the deprecate keyword (such as (b) in Listing 1), we generate the Fully Qualified API name of the class by appending the class name to the directory path. We also append the deprecation message to L_D (cf. Line-13) along with a list of decorators associated with the class. If there is a deprecated decorator (such as (a) in Listing 1) in the extracted list, we add the fully qualified name of the class and any description provided to list L_D (cf. Line-16). Similarly, when we encounter the function definition node, we extract the list of decorators associated with it. If there is a deprecated decorator in the extracted list, we add a fully qualified name of the function to list L_D (cf. Line-6). For each function call node in N_{AST} (cf. Line-7), we verify if *DeprecationWarning* or *FutureWarning* are passed as arguments (such as (c) in Listing 1) and add its fully qualified name to list L_D , which is the final generated list of deprecated API elements.

III. EVALUATION

A. Libraries Selection

To evaluate our approach, we applied it on six popular third-party Python libraries that were identified by Pimentel et al [11]. However, this approach is not limited to the selected libraries and could be applied to other Python libraries as well.

- *NumPy*: Array programming library [12].
- *Matplotlib*: A 2D graphics environment [13].
- *Pandas*: Data analysis and manipulation tool [14].
- *Scikit-learn*: Machine learning library for Python [15].
- *Scipy*: Library for scientific and technical computing [16].
- *Seaborn*: Data visualization based on *matplotlib* [17].

³<https://jupyter.org/>

Algorithm 1: Detecting Deprecated API Elements in Python Libraries

Input: P , Python Library Code
Output: L_D , List of Deprecated API Elements

```

1 Function Detect_Deprecated_API():
2    $L_D \leftarrow \{\}$ 
3    $P_{AST} \leftarrow \text{parseCode}(P)$ 
4    $\text{/* Traverse each node in } P_{AST} \text{ using BFS */}$ 
5   for  $N_{AST} \in P_{AST}$  do
6     if  $\text{isFunctionDefNode}(N_{AST})$  then
7        $D = N_{AST}.Decorators$ 
8       if  $\text{isDeprecatedDecorator}(D)$  then
9          $L_D.add(\text{getFullyQualifiedName}(N_{AST}.Name))$ 
10         $\text{/* Traverse each Node in } N_{AST} \text{ */}$ 
11        for  $Node \in N_{AST}$  do
12          if  $\text{isFunctionCallNode}(Node)$  and
13             $\text{isDeprecationWarning}(Node)$  then
14               $L_D.add(\text{getFullyQualifiedName}(N_{AST}.Name))$ 
15          else if  $\text{isClassDefNode}(N_{AST})$  then
16             $doc_{str} = N_{AST}.Docstring$ 
17            if  $\text{doc}_{str}.hasDeprecationKeyword()$  then
18               $L_D.add(\text{getFullyQualifiedName}(N_{AST}.Name))$ 
19             $D = N_{AST}.Decorators$ 
20            if  $\text{isDeprecatedDecorator}(D)$  then
21               $L_D.add(\text{getFullyQualifiedName}(N_{AST}.Name))$ 
22  return  $L_D$ 

```

B. Results

Table I summarizes the total number of deprecated API elements detected by the Algorithm 1 and the total number of deprecated API elements found in the source code of the Python libraries. We manually counted the number of deprecated API elements present in the source code of the libraries. From Table I, we can observe that the algorithm has detected more than 90% of the deprecated APIs. In the case of *Matplotlib*, only 65% of the deprecated APIs could be detected since *Matplotlib* deprecates many of its parameters using a custom warning function which does not have any parameters indicating if it is a *DeprecationWarning* or not. In such cases, the proposed algorithm could not detect the deprecated API elements.

In the case of *Scikit-learn*, *Numpy* and *Pandas*, some of the functions that are used to deprecate parameters or parameter values or deprecation warnings induced by other libraries are also captured. Hence, the number of deprecated API elements detected by the algorithm is higher than the actual number of deprecated APIs. Whereas in the case of *Scipy* and *Seaborn*, some of the parameters are deprecated without using any of the three deprecation strategies, which could not be detected by the algorithm. Hence, the number of deprecated API elements detected by the algorithm for *Scipy* and *Seaborn* are lower than the actual number of deprecated APIs.

IV. LIMITATIONS AND THREATS TO VALIDITY

APIScanner detects deprecated APIs through *decorator*, *warning* or *comments*. Any other deprecated APIs that are not implemented through the above three strategies cannot

Library Name	LOC	Total No. of Deprecated API elements identified using Algorithm 1	Total No. of Deprecated API elements in source code
<i>Scikit-learn</i>	388.1k	487	438
<i>Matplotlib</i>	982.5k	169	254
<i>Numpy</i>	145.6k	39	36
<i>Pandas</i>	668.9k	66	59
<i>Scipy</i>	725.62k	46	49
<i>Seaborn</i>	83.7k	31	35

TABLE I
EVALUATION OF RESULTS OBTAINED USING OUR ALGORITHM

be detected by the algorithm. Moreover, the algorithm finds the function or class in which a parameter is deprecated but the exact parameter deprecated may not be mentioned in the deprecation message displayed by the extension due to the inconsistent deprecation strategies adopted by the library maintainers. APIs deprecated without using the *DeprecationWarning* and *FutureWarning* as parameters in the warning function cannot be detected by the algorithm. APIs deprecated using single-line comments and not using the doc-strings also cannot be detected by the algorithm. Further, a major prerequisite for our approach is the availability of source code of libraries. We can mitigate the threat due to inconsistent deprecation strategies if we can ensure that the documentation is structured and well maintained for Python libraries.

Finally, since the results are evaluated manually, there may be human errors. Hence, we have carefully reviewed and validated some of the results using release notes to mitigate this potential threat. We plan to extend the evaluation of the tool using release notes and API documentation.

V. RELATED WORK

In the literature, several studies on deprecated APIs for different environments have been done to analyze and tackle the challenges posed by the deprecation of APIs in libraries.

Robbes et al. [5], [6] studied the reactions of developers to the deprecation and the impact of API deprecation on the *Smalltalk* and *Pharo* ecosystem. Ko et al. [7] examined 260 deprecated APIs from eight Java libraries and their documentation and observed that 61% of deprecated APIs are offered with replacements. Similarly, Brito et al. [18] conducted a large-scale study on 661 real-world Java systems and found that replacements are provided for 64% of the deprecated APIs. In another study [4] conducted on *Java* and *C#* projects, they have observed that an average of 66.7% of APIs in *Java* projects and 77.8% in *C#* projects were deprecated with replacement messages. In 26 open-source Java systems over 690 versions, Zhou et al. [19] analysed the history of deprecated APIs and observed that deprecated API messages are not well managed by library contributors with very few deprecated APIs being listed with replacements. Li et al. [3] characterized the deprecated APIs in Android Apps parsing the code of 10000 Android applications. Zhang et al. [2] have observed a significant difference in evolution patterns of Python and *Java* APIs and also identified 14 patterns in

which Python APIs evolve. Wang et al. [10] observed that library contributors do not properly handle API deprecation in Python libraries. To this end, there is a need for approaches and tools to automatically detect deprecated API elements in Python projects.

Several approaches have been proposed in the literature for other ecosystems to migrate from deprecated APIs [20], [9], [8]. Yaoguo Xi et al. [20] proposed an approach and built a tool *DAAMT* to migrate from deprecated APIs in Java to their replacements if recorded in the documentation. Fazzini et al. [9] developed a technique *AppEvolve* to update API changes in Android Apps by automatically learning from examples before and after-updates. Haryono et al. [8] proposed an approach named *CocciEvolve* that updates using only a single after-update example. However, tools that handle deprecated APIs in Python projects have not been developed, which motivated us towards the development of *APIScanner*.

VI. CONCLUSION AND FUTURE WORK

Considering the extensive use of deprecated APIs during software development and lack of proper documentation for deprecated APIs, we proposed an approach to automatically detect deprecated APIs in Python libraries during the development phase of the project. In this paper, we presented a novel algorithm and a tool called *APIScanner* that detects deprecated APIs. The algorithm identifies the APIs deprecated via *decorator*, *hard-coded warning* or *comments* by parsing the source code of the libraries and generated a list of deprecated APIs. *APIScanner* used this list and searched for the use of deprecated APIs in the current active editor. The tool highlights deprecated APIs in the source code along with further deprecation details. *APIScanner* thus aims to help developers detect deprecated APIs during the development stage and avoid searching through API documentation or forums such as Stack Overflow. Highlighting the use of deprecated APIs in the editor might help developers to address and replace them. The proposed algorithm identified 838 out of 871 API elements across six different Python libraries.

As future work, our goal is to strengthen the tool with release-specific information and develop a better user interface (such as different colors) to indicate the severity of the deprecation. We also plan to improve the documentation of deprecated APIs through the information obtained from the algorithm. We plan to extend the tool to provide a feature to migrate from the deprecated API to its replacement. We aim to improve the tool's accuracy by extracting APIs that are deprecated using the custom deprecation strategies. Finally, we plan to conduct extensive developer studies on the usage of the approach and the tool with more libraries.

REFERENCES

- [1] D. Dig and R. Johnson, "The role of refactorings in api evolution," in *21st IEEE International Conference on Software Maintenance (ICSM'05)*, 2005, pp. 389–398.
- [2] Z. Zhang, H. Zhu, M. Wen, Y. Tao, Y. Liu, and Y. Xiong, "How do python framework apis evolve? an exploratory study," in *2020 IEEE 27th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 2020, pp. 81–92.
- [3] L. Li, J. Gao, T. F. Bissyandé, L. Ma, X. Xia, and J. Klein, "Characterising deprecated android apis," in *Proceedings of the 15th International Conference on Mining Software Repositories*, ser. MSR '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 254–264.
- [4] G. Brito, A. Hora, M. T. Valente, and R. Robbes, "On the use of replacement messages in api deprecation: An empirical study," *Journal of Systems and Software*, vol. 137, pp. 306 – 321, 2018.
- [5] R. Robbes, M. Lungu, and D. Röthlisberger, "How do developers react to api deprecation? the case of a smalltalk ecosystem," in *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering*, ser. FSE '12. New York, NY, USA: Association for Computing Machinery, 2012.
- [6] A. Hora, R. Robbes, N. Anquetil, A. Etien, S. Ducasse, and M. T. Valente, "How do developers react to api evolution? the pharo ecosystem case," in *Proceedings of the 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, ser. ICSME '15. USA: IEEE Computer Society, 2015, p. 251–260.
- [7] D. Ko, K. Ma, S. Park, S. Kim, D. Kim, and Y. L. Traon, "Api document quality for resolving deprecated apis," in *Proceedings of the 2014 21st Asia-Pacific Software Engineering Conference - Volume 02*, ser. APSEC '14. USA: IEEE Computer Society, 2014, p. 27–30.
- [8] S. A. Haryono, F. Thung, H. J. Kang, L. Serrano, G. Muller, J. Lawall, D. Lo, and L. Jiang, "Automatic android deprecated-api usage update by learning from single updated example," 2020.
- [9] M. Fazzini, Q. Xin, and A. Orso, "Automated api-usage update for android apps," in *Proceedings of the 28th ACM SIGSOFT International Symposium on Software Testing and Analysis*, ser. ISSTA 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 204–215.
- [10] J. Wang, L. Li, K. Liu, and H. Cai, "Exploring how deprecated python library apis are (not) handled," in *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2020, pp. 233–244.
- [11] J. a. F. Pimentel, L. Murta, V. Braganholo, and J. Freire, "A large-scale study about quality and reproducibility of jupyter notebooks," in *Proceedings of the 16th International Conference on Mining Software Repositories*, ser. MSR '19. IEEE Press, 2019, p. 507–517.
- [12] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del R'io, M. Wiebe, P. Peterson, P. G'erard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020.
- [13] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
- [14] W. McKinney, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, S. van der Walt and J. Millman, Eds., 2010, pp. 51 – 56.
- [15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [16] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright et al., "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.
- [17] M. Waskom and the seaborn development team, "mwaskom/seaborn," Sep. 2020.
- [18] G. Brito, A. Hora, M. T. Valente, and R. Robbes, "Do developers deprecate apis with replacement messages? a large-scale analysis on java systems," in *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, vol. 1, 2016, pp. 360–369.
- [19] J. Zhou and R. J. Walker, "Api deprecation: A retrospective analysis and detection method for code examples on the web," in *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, ser. FSE 2016. New York, NY, USA: Association for Computing Machinery, 2016, p. 266–277.
- [20] Y. Xi, L. Shen, Y. Gui, and W. Zhao, "Migrating deprecated api to documented replacement: Patterns and tool," in *Proceedings of the 11th Asia-Pacific Symposium on Internetwork*, ser. Internetwork '19. New York, NY, USA: Association for Computing Machinery, 2019.