

الماسح الآمن لرمز الاستجابة السريعة

Secure scanner for QR code

Participant

Suliman Aldubaikhi

Supervisor

Professor Dr. Suliman Alsuhibany

مركز الابتكار والملكية الفكرية

مشروع حاضنة ابتكار23

1. Introduction

QR code (Quick Response) is a two-dimensional version of the barcode, typically made up of black and white pixel patterns, used for storing URLs or other information for reading by the camera. QR codes consist of black squares and dots arranged in a square grid with a white background, and every QR Code is unique and stores data in numeric, alphanumeric, byte/binary, and kanji formats. QR codes are designed for various needs, such as quickly directing users to a website, downloading apps, adding geolocation, ordering from restaurants, shops, and coffee, acting as an access control system, sharing social media profiles, and payments. However, the use of the QR code has become widespread in our daily life, it has a wide range of applications.

1. Threats and vulnerabilities.

Attackers take advantage of the increased use of QR codes to launch different attacks. This section lists some of these attacks.

* 1. Phishing

Is an activity that tricks people to discover their sensitive information by masquerading as a trustworthy entity [1-2]. A QR code can redirect users to a fake bank that looks exactly like a real bank. A normal user cannot see the differences and type in his or her information and hand it to the attackers [3-4].

* 1. Malware

A malicious QR code can be used to redirect users to a URL containing malware[5]. An early example of malware attacks through QR code was that people were fooled into scanning a QR code and downloading a malicious application, which sent off multiple text messages to a number that charged users $5 per SMS message [6].

1. Proposed solution

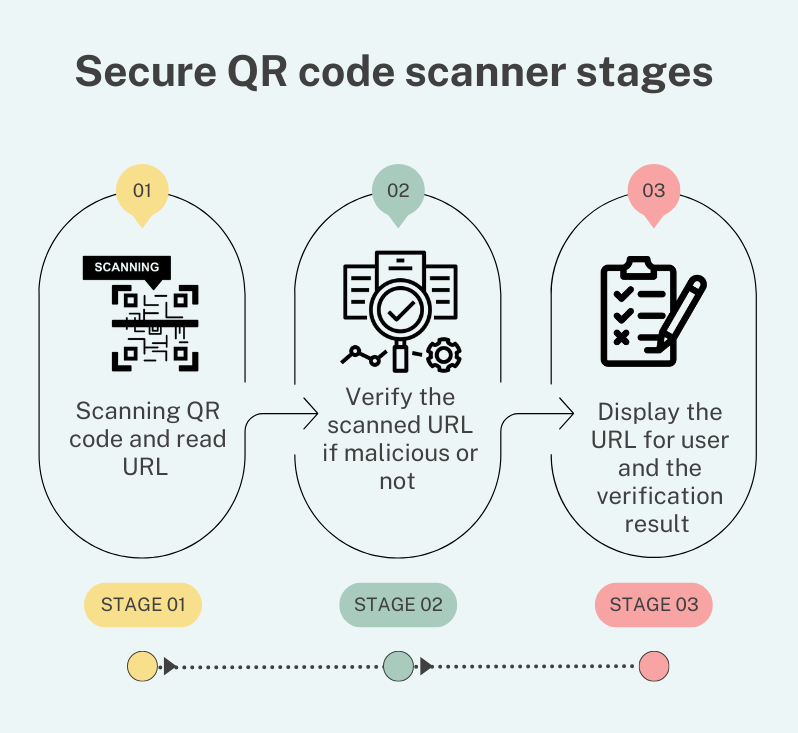


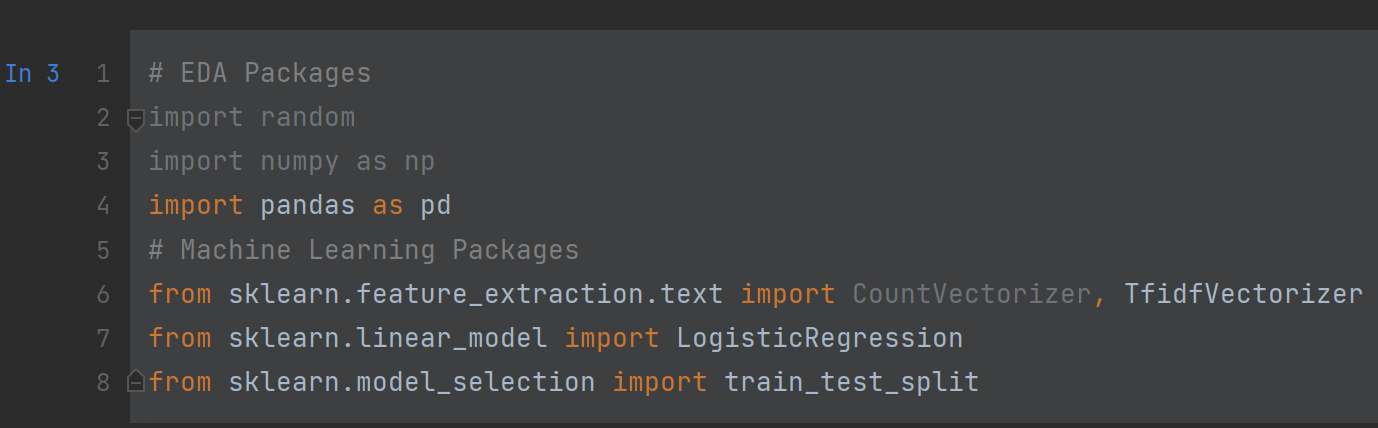
Figure 1 Stages of the QR code scanner

The goal of this project is to address the QR reader threat by adding a URL verification stage to the scanner. Machine Learning will be used for the verification stage. Application stages shown in figure 1.

1. Application
   1. Machine learning

A subfield of artificial intelligence (AI) and computer science called machine learning focuses on using data and algorithms to simulate how humans learn, increasing the accuracy of the system. The expanding area of data science includes machine learning as a key component. In data mining projects, classifications or predictions are made using algorithms that have been trained using statistical techniques [7]. This project uses a logistic regression machine learning model with the URLData dataset to verify whether URLs are malicious or not [8].

The first step, importing machine learning libraries in python language, the code in Figure 2.



*Figure 2. Importing machine learning libraries.*

* 1. URL Dataset

The URL dataset consists of 420,465 rows and 2 columns: URL and label (good or bad). There are 75,643 malicious URLs and 344,821 benign URLs [9]. The data load code in Figure 3.

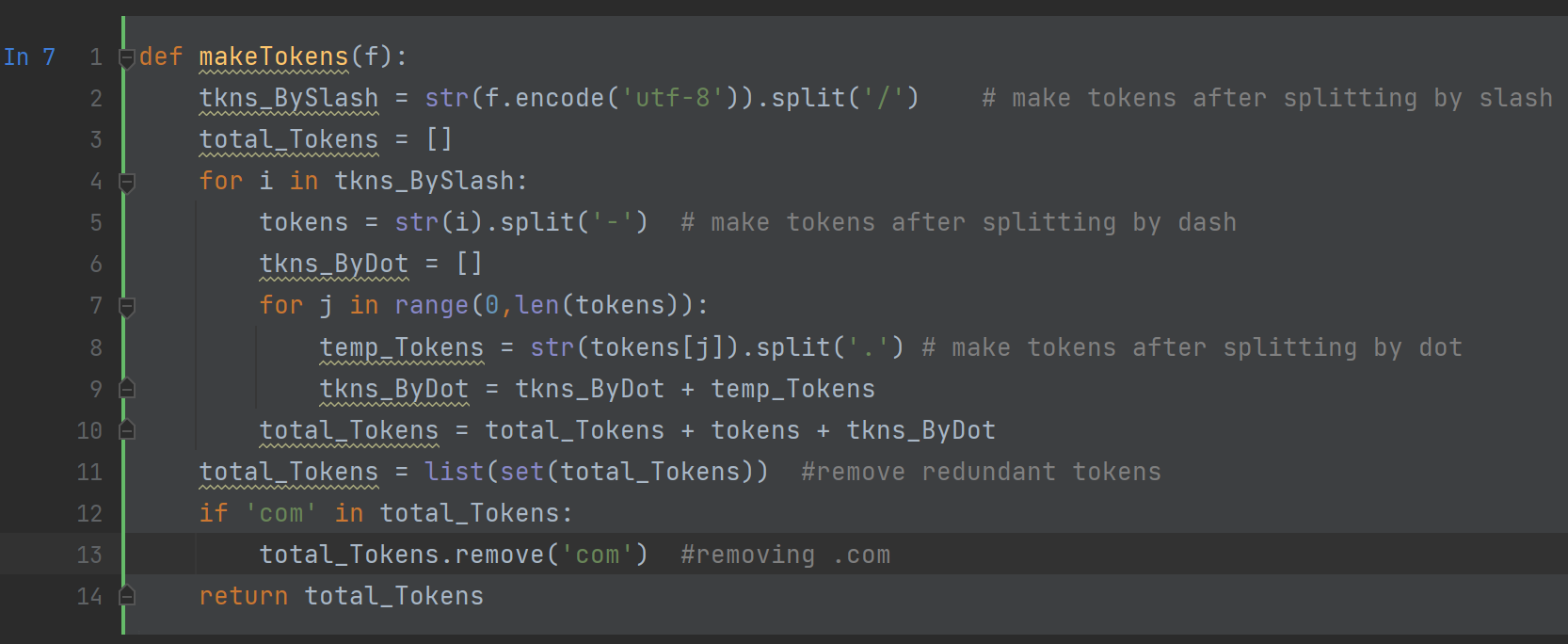


*Figure 3. Loading the dataset.*

### 

### Data Vectorization

A tokenizer creation ,Split ,Remove Repetitions and “Com” by using TfidfVectorizer machine learning library. Code in figure 4.

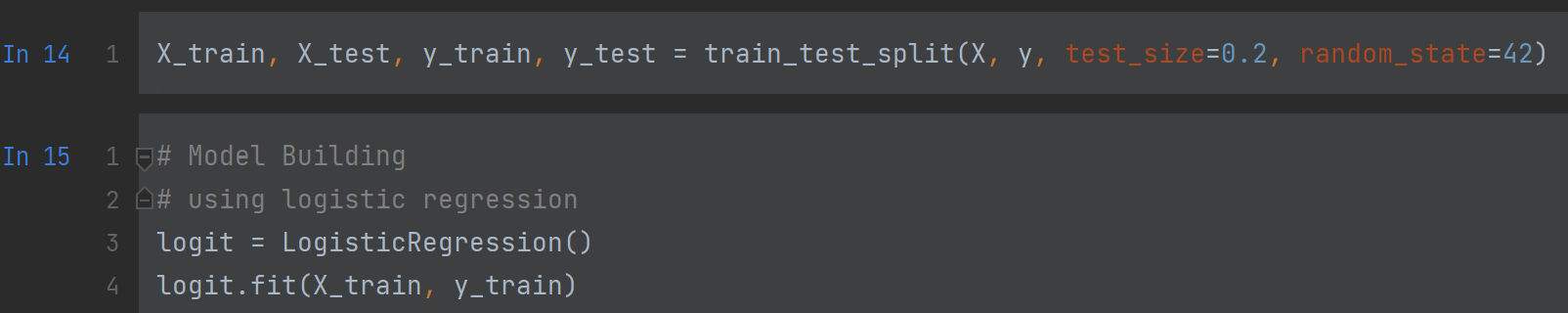


*Figure 4. Create a tokenizer.*

* 1. Logistic Regression Model

It is a supervised machine learning, predictive analytics and classification regularly makes use of this type of statistical model, also referred to as a logit model. Based on a given dataset of independent variables, logistic regression predicts the probability that an event will occur, such as fake or not fake. Given that the result is a probability, the dependent variable's range is 0 to 1 [10].

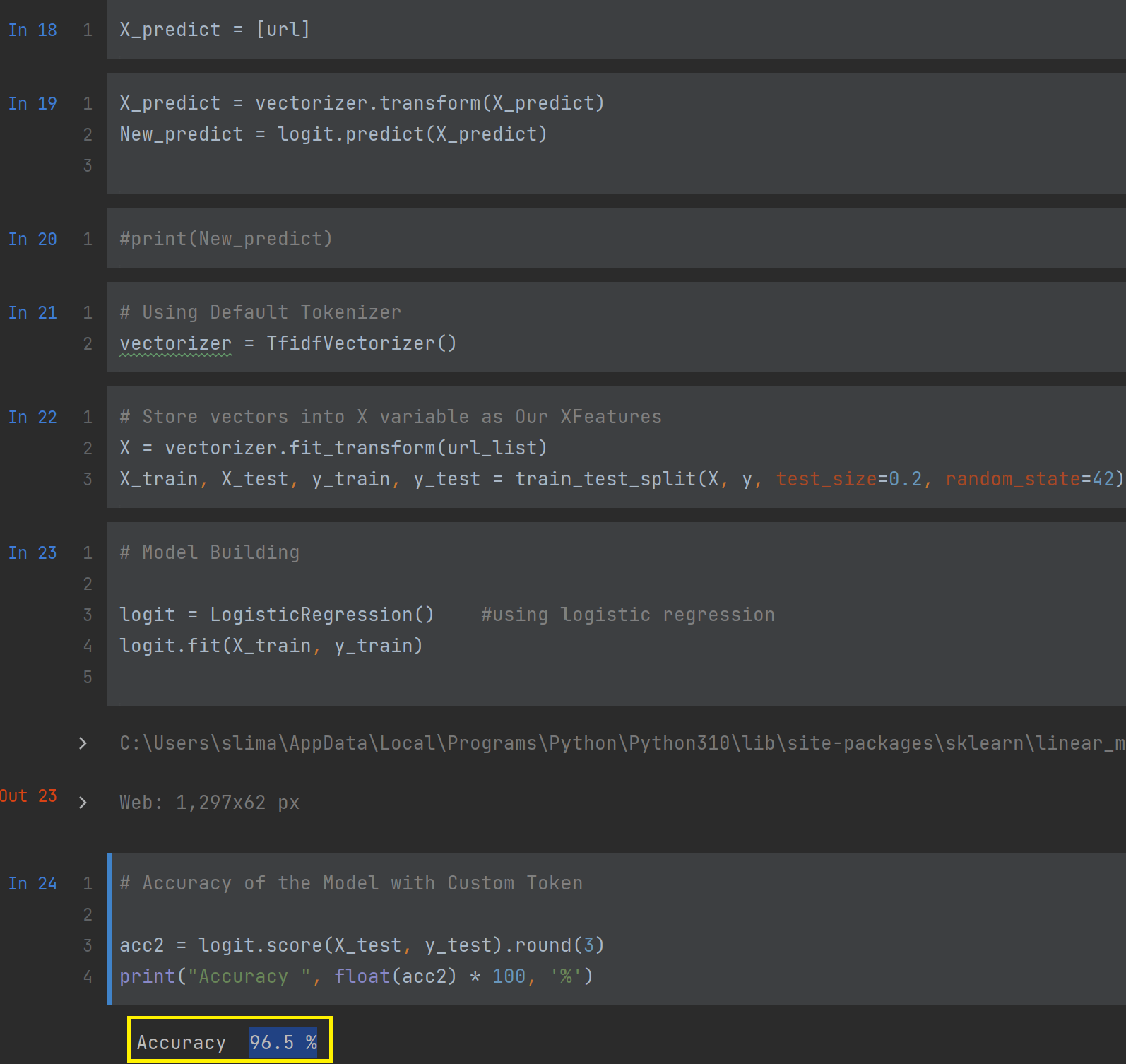
#### Split into training and testing dataset and building the model using logistic regression, code shown in figure 5.



*Figure 5. Training and testing the model.*

* 1. Predicting with the model

Accuracy of logistic regression is 96.5%. In Figure 6, predicting with logistic regression code.



*Figure 6. Building the model.*

* 1. QR code scanner

QR Scanner code in Figure 7 [11]

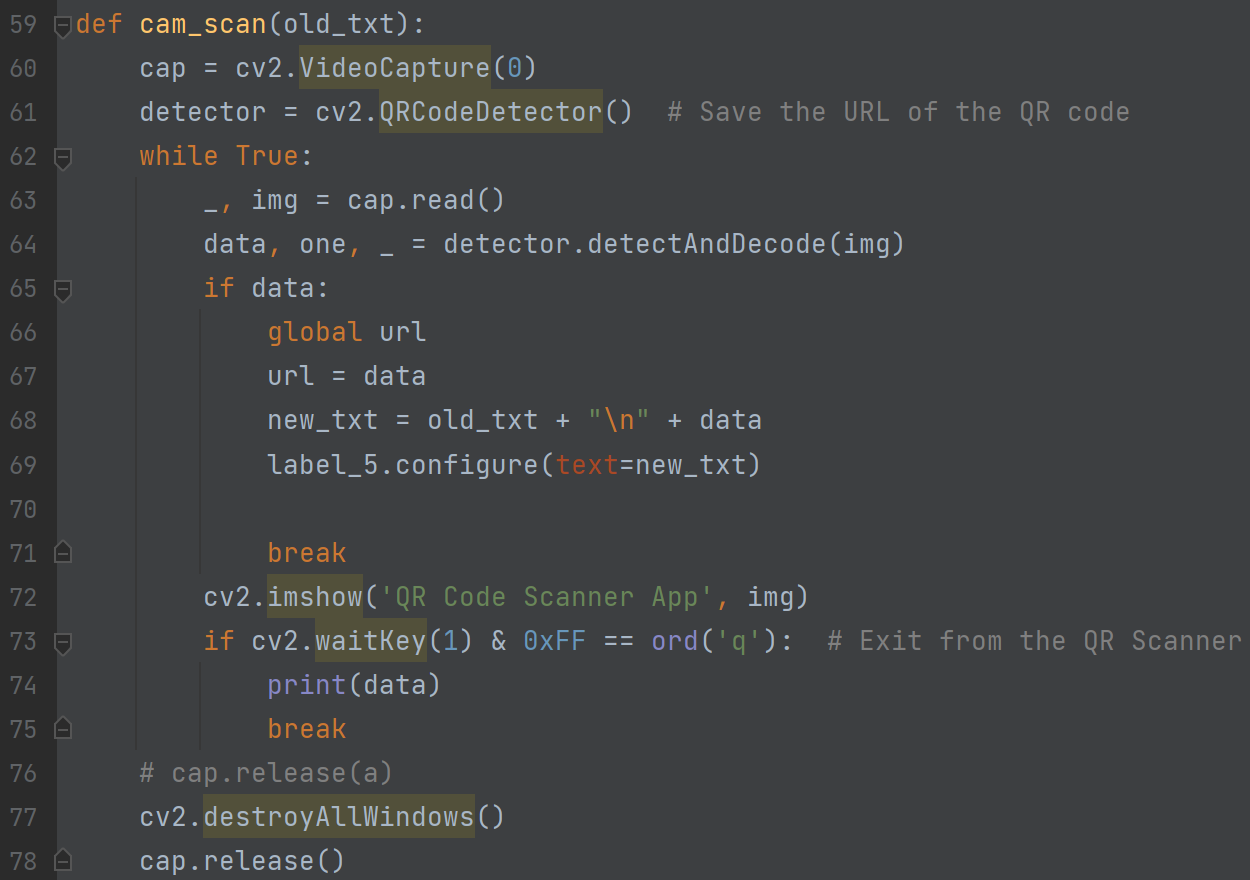
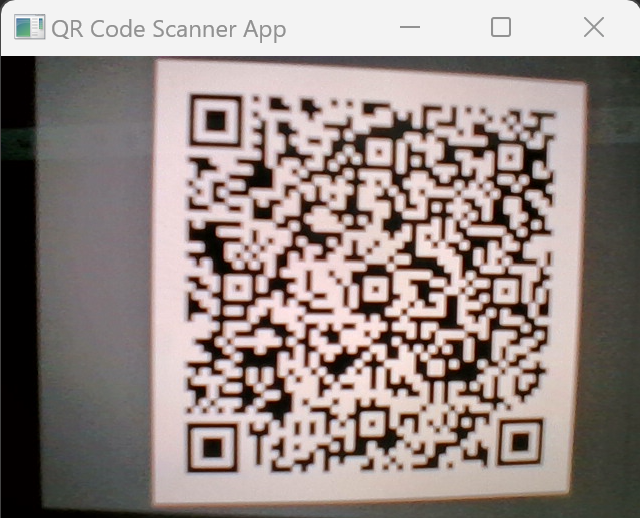
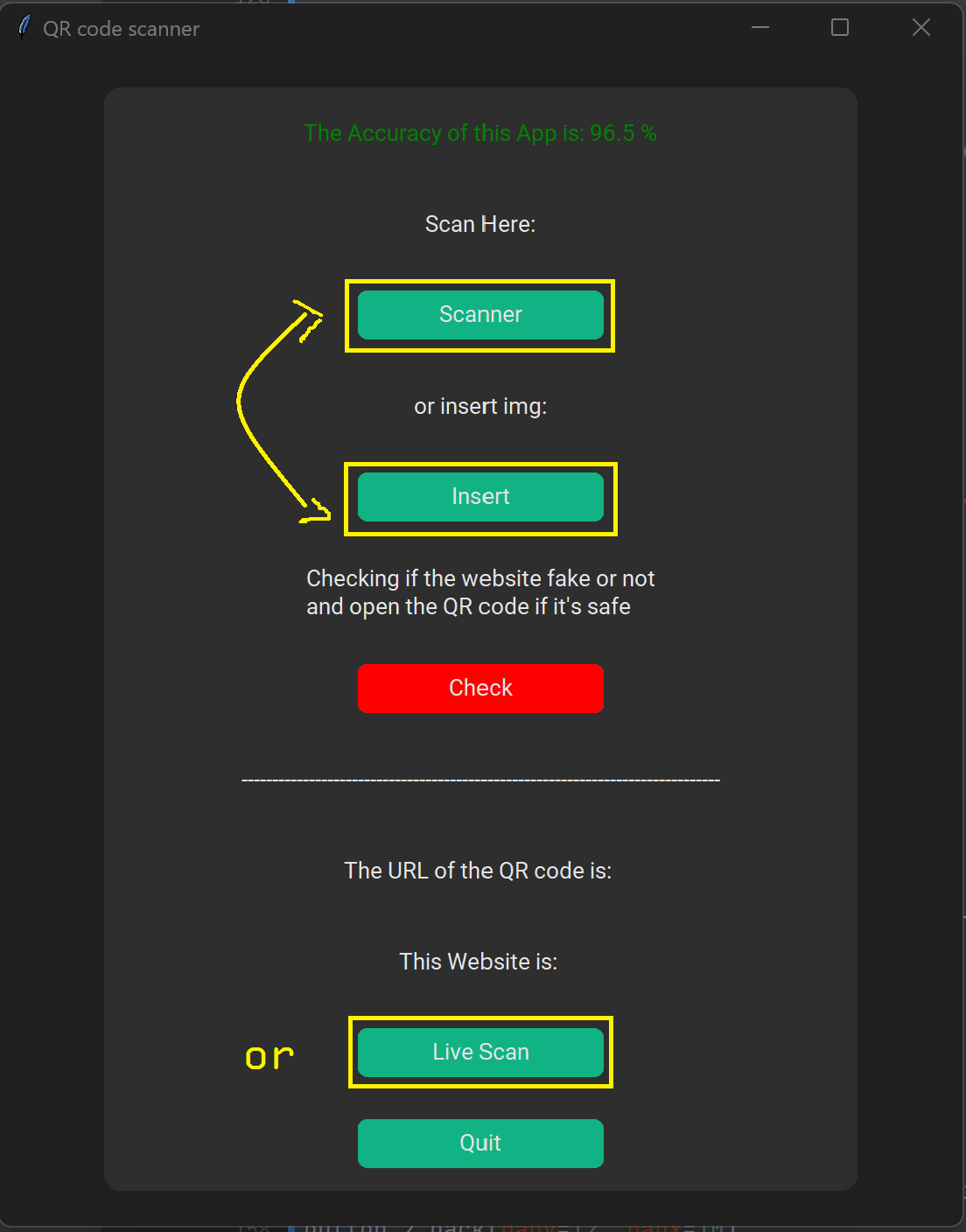


Figure 7

* 1. GUI

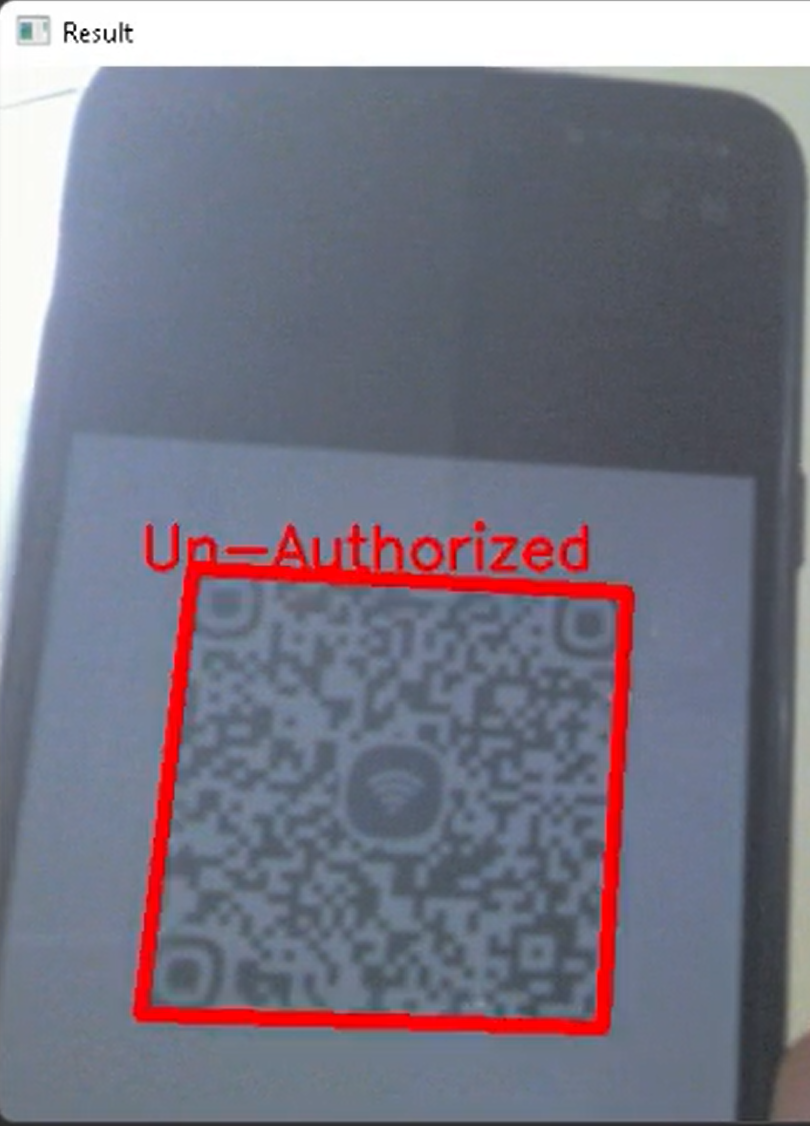
This section shows the application interface [12]. On the main page, the user can select one of two ways to scan and verify the URL as in Figure 8. First way, as shown in Figure 9 , verification will be done during the scanning. Second way, as shown on Figure 10 scanning the URL then clicking the "Check" button for verification.



1- Scan, Insert pic or Live Scan

2- Scanning.

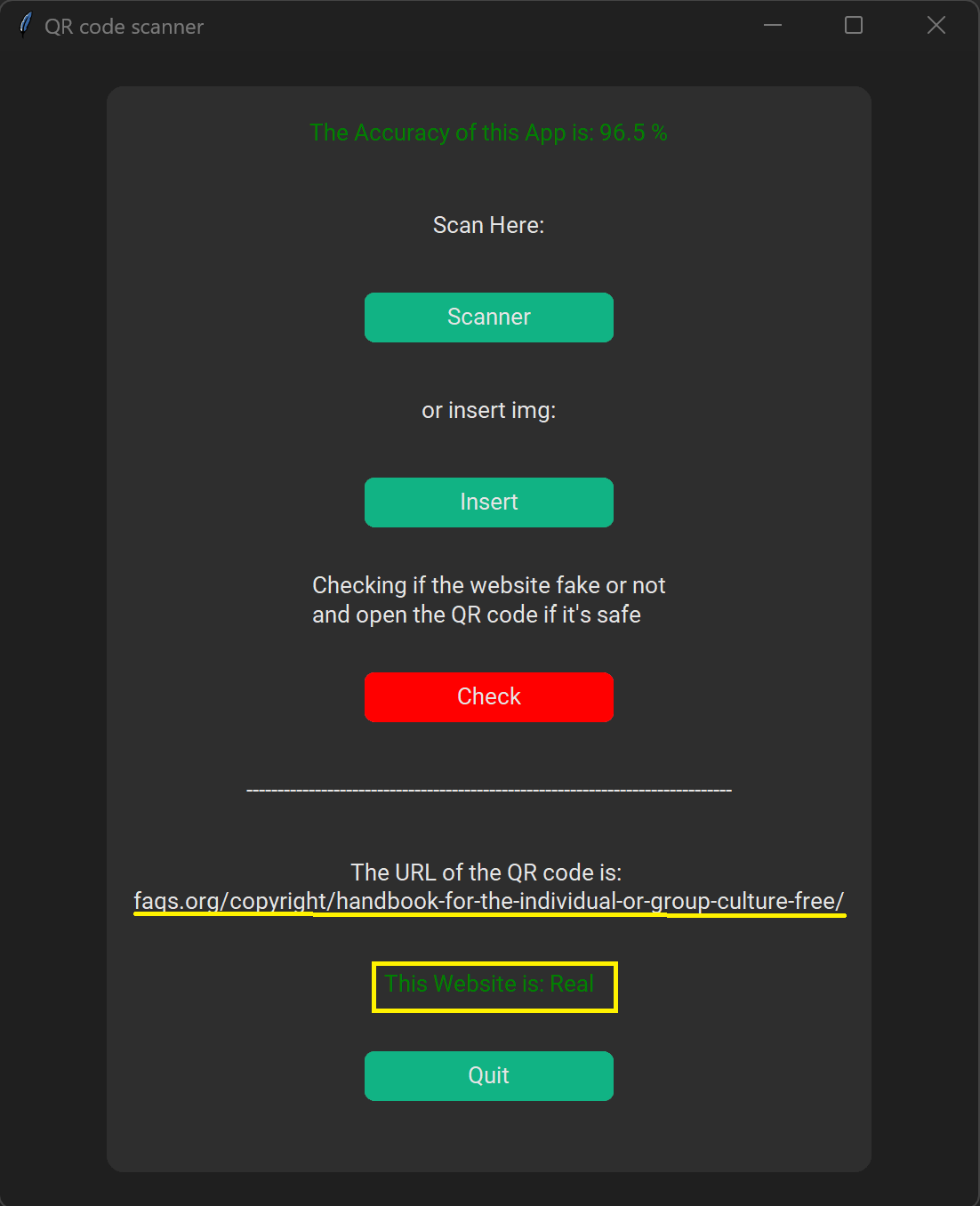
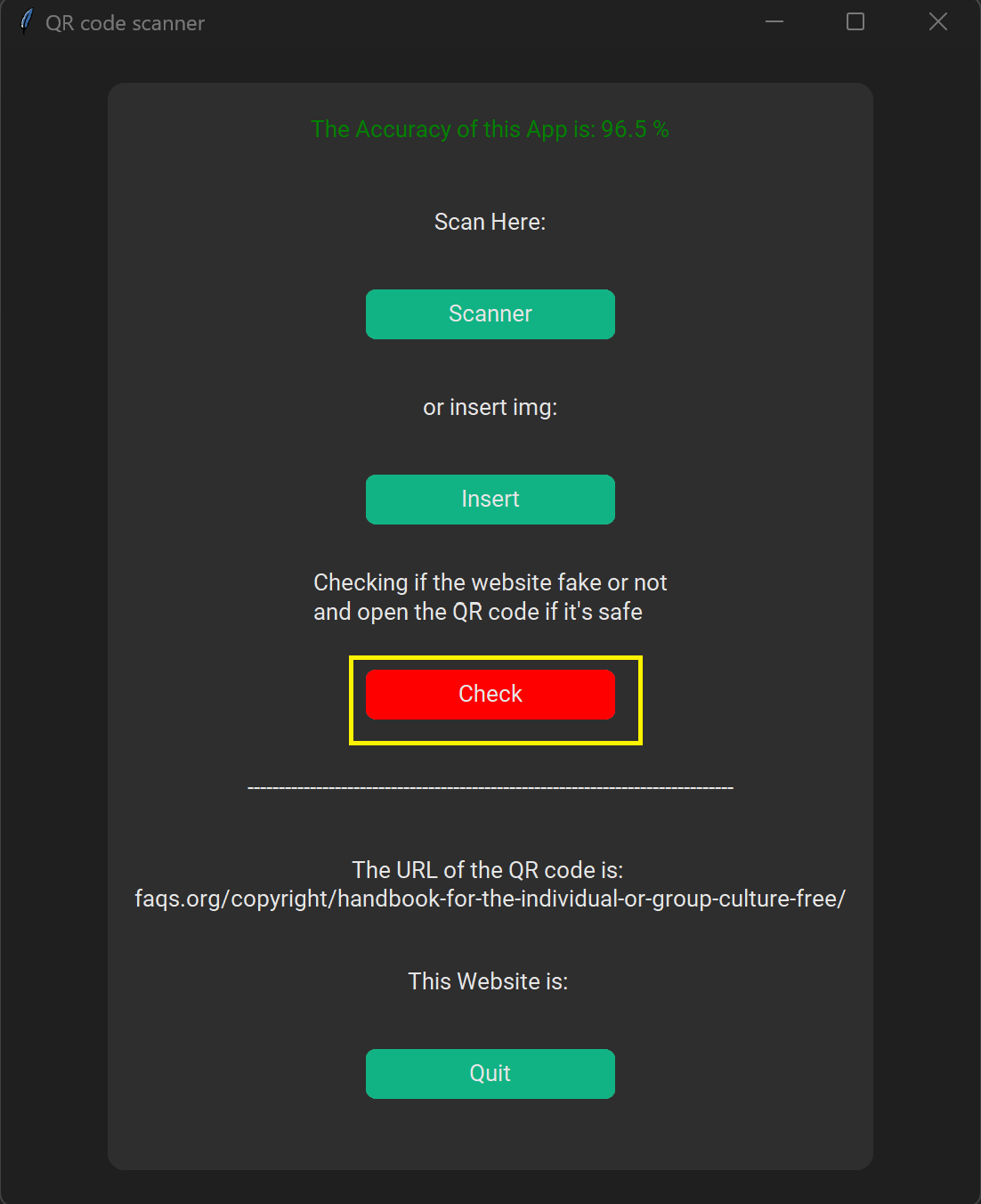
Figure 8



3- Unauthorized QR code

4- Authorized QR code

Figure 9: live scanner, to scan the code without inserting or scanning the code in the app.



5- Checking if the website real or fake

6- The URL of the QR code and the result

Figure 10

1. Conclusion

Significant security issues come with QR codes. Attackers can encrypt harmful links that take users to phishing websites, for example. These malicious QR codes can be printed on tiny stickers and used in place of legitimate ones on billboard advertisements. This project proposed a solution to protect QR code users by adding a URL verification stage to the scanner. URL verification by machine learning Logistic Regression model, with an accuracy of 96.5%.

References

[1] Dhamija, R., Tygar, J. D., & Hearst, M. (2006, April). Why phishing works. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 581-590).

[2] Soni, P., Firake, S., & Meshram, B. B. (2011, February). A phishing analysis of web based systems. In Proceedings of the 2011 International Conference on Communication, Computing & Security (pp. 527-530).

[3] Kieseberg, P., Leithner, M., Mulazzani, M., Munroe, L., Schrittwieser, S., Sinha, M., & Weippl, E. (2010, November). QR code security. In Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia (pp. 430-435).

[4] Felt, A. P., & Wagner, D. (2011). Phishing on mobile devices.

[5]Zhou, Y., & Jiang, X. (2012, May). Dissecting android malware: Characterization and evolution. In 2012 IEEE symposium on security and privacy (pp. 95-109). IEEE.

[6] Monthly Malware Statistics: September 2011. Available: www.securelist.com/en/analysis/204792195/Monthly\_Mal ware\_Statistics\_September\_2011

[7] What is Machine Learning?. (2020). Retrieved 12 November 2022, from <https://www.ibm.com/sa-en/cloud/learn/machine-learning>

[8] Jcharis (2017). *Detecting Malicious Url With Machine Learning In Python.ipynb*. github. <https://github.com/Jcharis/Machine-Learning-In-Julia-JCharisTech.git>

[9] Jesse, J. (2017). *urldata.csv Cleaned urls list*. Github. <https://github.com/Jcharis/Machine-Learning-In-Julia-JCharisTech/blob/master/urldata.csv>

[10] What is Logistic regression? | IBM. (2022). Retrieved 12 November 2022, from <https://www.ibm.com/topics/logistic-regression>

[11] https://t.me/s/akpython

[12] Schimansky, T. (2022). *customtkinter*. github. <https://github.com/TomSchimansky/CustomTkinter.git>