**User Manual for Running Malimg and Malives Genetic Algorithm Notebooks**

**Introduction**

This manual provides instructions for running the Malimg and Malives Genetic Algorithm notebooks. The notebooks are used for malware classification using XGBoost and genetic algorithms. The Malimg notebook is used for classifying the Malimg dataset, while the Malives notebook is used for classifying the Malives dataset.

**Dataset Requirements:**

1. Open Google Drive and create a folder named "malimg\_dataset" for the Malimg dataset and "malware\_dataset" for the Malives dataset.
2. Download the Malives dataset from <https://web.cs.hacettepe.edu.tr/~selman/malevis/>

into the "malware\_dataset" folder.

1. Download the Malmig dataset from [malimg\_dataset9010 | Kaggle](https://www.kaggle.com/datasets/keerthicheepurupalli/malimg-dataset9010/code) into the “malmig\_dataset” folder.

**Minimum System Requirements**

To run these notebooks, you will need a computer with the following minimum system requirements:

Operating System: Windows, macOS or Linux

Processor: Intel Core i3 or higher

RAM: 16GB or higher

Hard Disk Space: 5GB or higher

GPU: NVIDIA GeForce GTX 1050 or higher (optional)

**To run the two notebooks for Malimg and Malware classification using Genetic Algorithm on Google Colab, follow the steps below:**

1. Ensure you have a Google account and are signed in to Google Colab (https://colab.research.google.com/).
2. Open a new notebook by clicking on "File" in the menu bar and selecting "New notebook".
3. In the new notebook, click on "Runtime" in the menu bar and select "Change runtime type". Choose "Python 3" as the runtime type and "GPU" as the hardware accelerator.
4. Copy and paste the code from the notebook into the cells of the new notebook.
5. Run the first cell to mount Google Drive to access the dataset. This requires authorization by following the prompted instructions.
6. The second cell is used to install and import the necessary libraries. Run the cell to install the required packages.
7. Run the cells containing the Genetic Algorithm for Malimg/Malware classification. These cells will take some time to run, depending on the number of generations and other parameters set.
8. Once the Genetic Algorithm has finished running, the resulting best model will be displayed.
9. Use these parameters modify the XGBoost algorithm inthe next cell. On successfully running the cell it displays the accuracy.
10. Run the next cells to print the confusion matrix and then to predict an example using the same model.
11. Remember to change the dataset path and any other necessary parameters for the second notebook.
12. Finally, run the cells containing the Genetic Algorithm for Malware classification using the same steps as the first notebook.

**Minimum requirements:**

* A Google account
* A computer or device with a web browser
* Stable internet connection
* Basic knowledge of Python and Google Colab interface.