Challenge 2

In this challenge, you are provided with a dataset of malicious and benign URLs, and you're asked to use the k-NN algorithm (k = 3) to fit a model on the training data. Then, you're asked to test the fitted model on a test dataset and evaluate the performance of your model. Thus, your task is to prepare a report and submit a PDF file by Tuesday (**10/06/2020 before 6 pm**) considering the below details:

1. Read the CSV file into a data structure that you think is the best. This CSV file contains 1781 malicious and benign URLs collected from different places. The format of this CSV file is as follows:
    a. For each row (i.e., URL), you have 11 different features and a label. If the URL is malicious, the label is 1; otherwise, it is 0.
    b. All features have integer values. However, their scales might be different.
    c. Some features might have missing values or "NA" values. So, you may need to replace these with either 0, or consider removing this feature from the dataset.
2. Once you read this CSV file, you should have a matrix with size: 1781 * 12. Now, import and use a scikit-learn k-NN classifier and fit it on 80% of URLs.
3. In the final step, apply the fitted model on the remaining 20% of URLs and report the precision and accuracy of your model using the confusion matrix.