# Created on Wed May 27 22:13:54 2020

@author: Kofi Agyeman

# UCR CS235 - Data Mining Techniques: Spring 2020

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Phishing detection algorithm: Support Vector Machine implementation

# Data set:

# https://www.kaggle.com/akashkr/phishing-website-dataset

# Sources:

# Parts of implementation inspired by SVM tutorials from scikit learn and

# Towards data science svn from scratch online

@ https://scikit-learn.org/stable/auto\_examples/svm/plot\_rbf\_parameters.html#sphx-glr-auto-examples-svm-plot-rbf-parameters-py

@ https://towardsdatascience.com/svm-implementation-from-scratch-python-2db2fc52e5c2

# CODE FILES:

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# ka\_phishing\_svm\_v1.py :Only this file is required for implementation of the algorithm

# DATA FILES:

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# dataset.csv Properties of the data set can be accessed by running cell 2: Load, Partition and visualize in the code file

# Location: /Data/dataset.csv

# Dataset Source:

# https://www.kaggle.com/akashkr/phishing-website-dataset

# IMPLEMENTATION

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Once the first two code cell blocks are initialized all other cells can be run independently

# Note:The feature selection cell has to be initialized to test subsequent feature selection analysis

# for linear, polynomial and gaussian svm kernels

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

END