

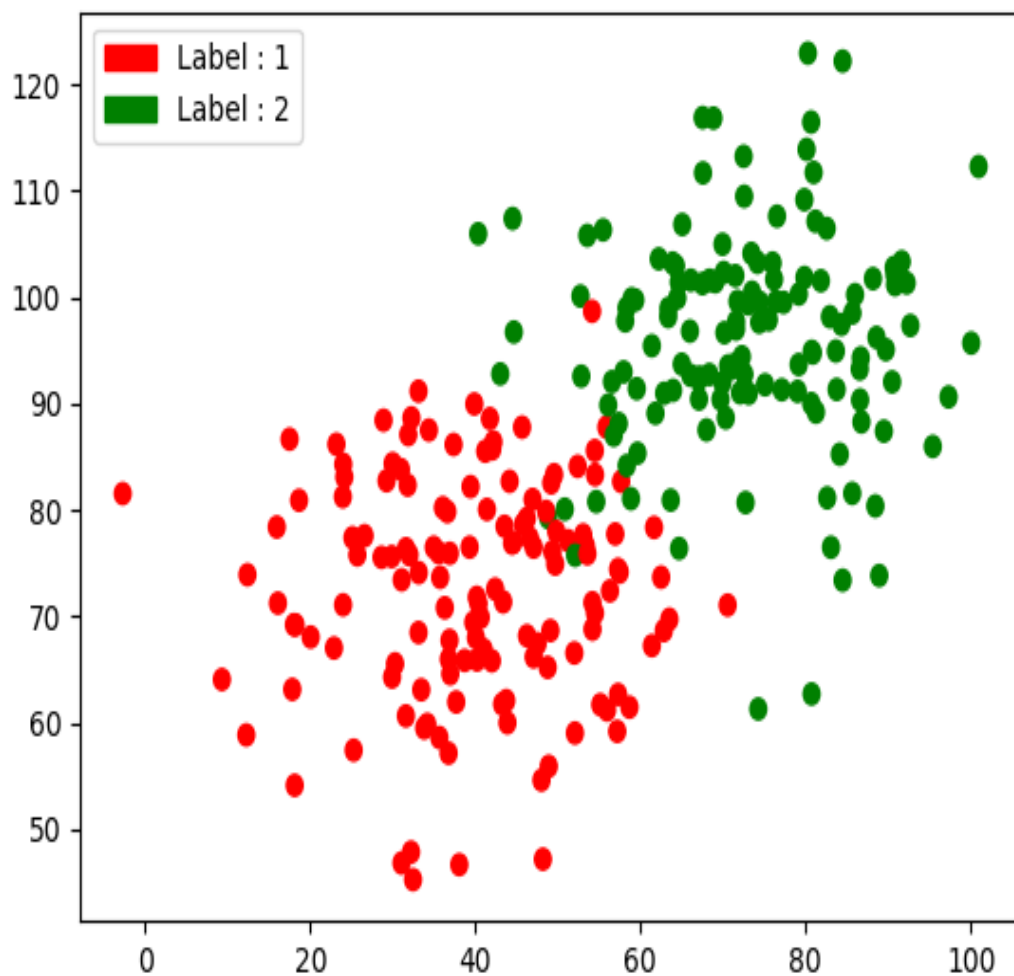
## CSC – 591 Internet of Things Analytics (Project 5: SVMs)

Name :- Harsh Jatinbhai Patel

Student Id :- 200258486

Unity Id :- hpatel8

- ❖ Colored scatter plot to get a general idea of the cluster of points. Label with value 1 is colored as red and label with value 2 is colored as green.
- ❖ Here we can see that the range of values on X-axis is [0-105] and values on Y-axis is [0-125]. Below is the scatter plot with points plotted with their appropriate color.



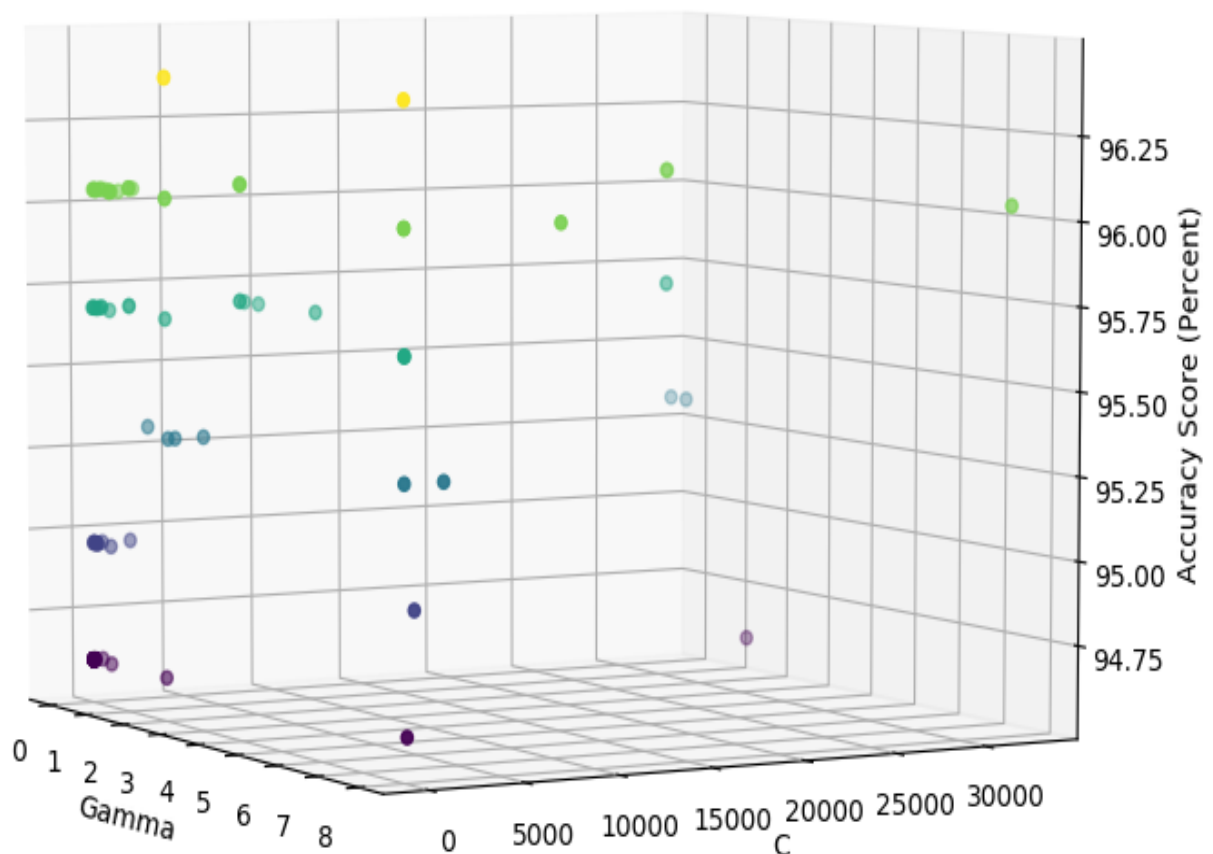
❖ Performing Grid Search on our SVM method and plotting a 3D Plot with Gamma ( $\gamma$ ) on X-axis, C on Y-axis and the Accuracy score in percentage on the Z-Axis.

➤ Here the range of values considered for implementing the Grid Search on the SVM Method are as follows:

- Gamma =  $[2^{-15}, 2^{-14}, \dots, 2^3]$
- C =  $[2^{-5}, 2^{-14}, \dots, 2^3]$
- Kernel Type : RBF

❖ Here I have used the stratified fold method to perform cross validation on folds of our data and have generated 5 folds of the dataset.

❖ Here is the 3D Plot of our Model



❖ Here the accuracy score is in percentage which is on Z-axis, Gamma is on X-axis and C on Y-Axis respectively.

❖ Here we can visually see that the highest accuracy score in our plot is **96.40%**.

❖ We visually check the values and see that for **C = 2.0** and **Gamma = 2.0** we get this score.

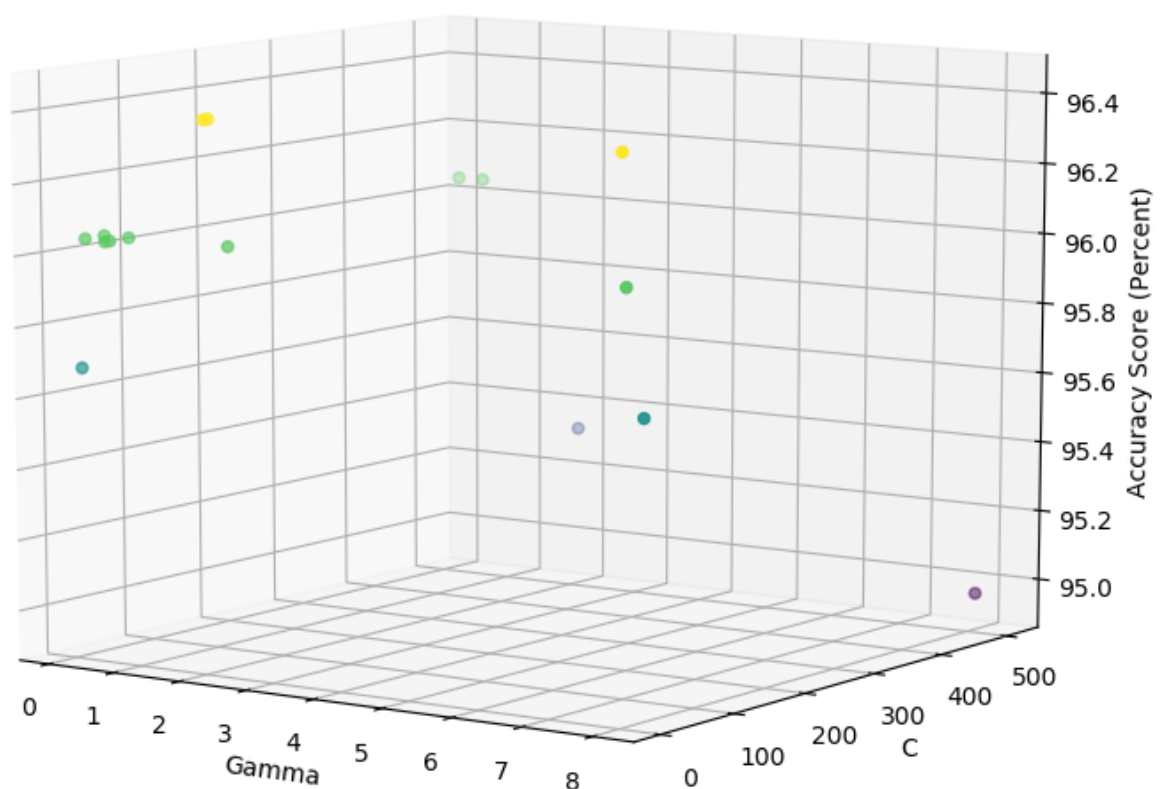
- Now we shall perform a refined search in our grid which means we shall limit our gamma and C values and use those values that perform the best i.e. which gives us a good accuracy and perform the grid search process again.

❖ **Implementing Grid search in a smaller range of values with the following parameters.**

- $C = [2^1, 2^3, 2^5, 2^7]$
- $\text{Gamma} = [2^{-3}, 2^{-1}, 2^1, 2^3]$
- Type of Kernel : RBF

- ❖ Here I selected the best performing C and Gamma values from the above 3D Plot and perform the Grid Search again to get the Best C and Gamma values.

- ❖ 3D Plot of the second iteration of the Grid Search with limited range of values.



- ❖ Here from the above plot we can see that the maximum accuracy score is 96.40% and the C and Gamma values are still C = 2.0 and Gamma = 2.0.

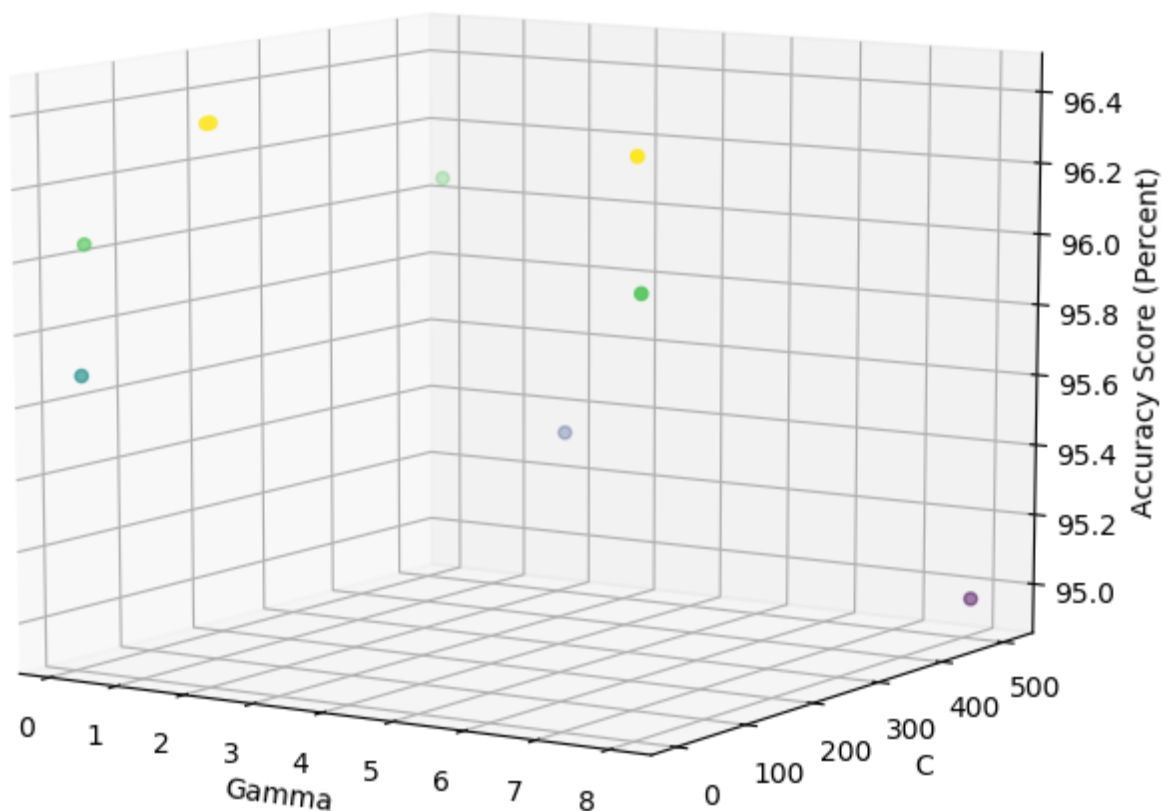
- ❖ Hence here we can see that the values of C and Gamma remained the same even when we limited our range of search and performed the grid search on the limited set of values.

❖ This means that  $C = 2.0$  and  $\text{Gamma} = 2.0$  are the best possible values and the accuracy score is 96.40%.

❖ Now, just to make sure that our values are correct we shall narrow down  $C$  and  $\text{Gamma}$  to a much smaller range of values which perform the best to get the best accuracy score and best values of  $C$  and  $\text{Gamma}$ .

❖ **Range of values used for the third iteration of the Grid Search.**

- $C = [2^1, 2^3, 2^7]$
- $\text{Gamma} = [2^{-3}, 2^1, 2^3]$
- Type of Kernel : RBF



- ❖ Hence by performing the grid search on a much smaller range of values and from the above 3D Plot we can confirm that following are the best results of our Grid Search and SVM implementation.
  - **Best value of C : 2**
  - **Best value of Gamma : 2**
  - **Accuracy Score for C = 2 and Gamma = 2 is 96.40%**
- ❖ So, we can see that to get the best values of C and Gamma we need to perform a grid search till we get the same values of C and Gamma for each iteration of our Grid Search.
- ❖ In this SVM model C is the cost of misclassification on the data, here I observed that on increasing the value of C the accuracy dropped and it also gave me a higher bias and a lower variance.
- ❖ Also on changing the parameters of Gamma I found that small values of Gamma gave a low bias and high variance where as large values of Gamma gave high bias and low variance.
- ❖ Therefore the values of C and Gamma heavily affect our classification in the SVM model. So Cross-validation, Resampling, Grid Search need to be implemented to get the best values of C and Gamma.