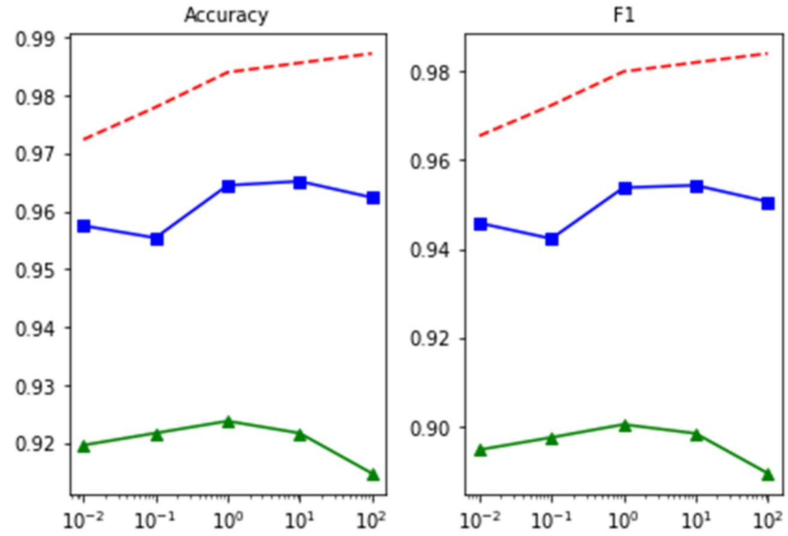
5-fold Cross validation:

- The arrayPrimeNumber and arrayNotPrimeNumber was split into 5 parts – 1 for testing, 3 for training and 1 for validation set.
- Similarly, the arrayLabelPrimeNumbers (consisting of labels for class of prime numbers) and arrayLabelNotPrimeNumbers (consisting of labels for class of not prime numbers) were also split into 5 parts.
- Then, for each of the 5-fold cross validation sets, the values of one part of the primeNumber array and the not prime number arrays were concatenated. Similar operation was performed for the labels array parts corresponding to the labels. Then, to have a permuted combination of data, shuffle function was used on each of the parts. This was done to avoid data skewness.
- Then, C parameters – 0.01,0.1,1,10,100 were used as C param values for the SVM Linear Kernel model. It was initially trained on the training set – combination of 3 split array parts, and then predicted on the training set, validation set and the testing set. The corresponding Accuracy, F1 score for the same were tabulated corresponding to each of the C parameters and the best one was chosen among them.
- The confusion matrix for the best C parameter was obtained for the test set.
- The accuracy and F1 score was plotted for the training set, validation set and the test set data.

```
Feature Engineering
Total length of array of prime numbers and labels of prime numbers
721 721
Length of 5 split arrays of prime numbers and corresponding labels array
145 145
144 144
144 144
144 144
144 144
Total length of array of not prime numbers and labels of not prime numbers
1076 1076
Length of 5 split arrays of not prime numbers and corresponding labels array
216 216
215 215
215 215
215 215
215 215
Length of combined arrays of prime and not prime numbers ; and their labels
361 361
359 359
359 359
359 359
359 359
Length of shuffled combined arrays of prime and not prime numbers ; and their label
s
361 361
359 359
359 359
359 359
359 359
1077
1077
1077
1077
                       Train     Valid      Test
parameter metric
0.01      Accuracy  0.972377  0.957521  0.919668
          F1        0.965481  0.945835  0.894731
0.10      Accuracy  0.977948  0.955432  0.921745
          F1        0.972364  0.942312  0.897506
```

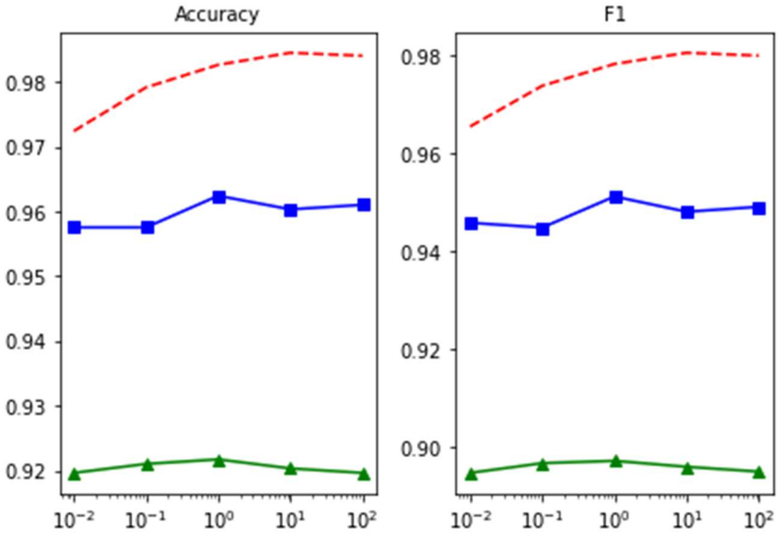
```
1.00      Accuracy  0.983983  0.964485  0.923823
          F1        0.979978  0.953813  0.900437
10.00     Accuracy  0.985608  0.965181  0.921745
          F1        0.982029  0.954333  0.898412
100.00    Accuracy  0.987233  0.962396   0.91482
          F1        0.984053  0.950652  0.889433
```



Without Feature Engineering

```
Total length of array of prime numbers and labels of prime numbers
721 721
Length of 5 split arrays of prime numbers and corresponding labels array
145 145
144 144
144 144
144 144
144 144
Total length of array of not prime numbers and labels of not prime numbers
1076 1076
Length of 5 split arrays of not prime numbers and corresponding labels array
216 216
215 215
215 215
215 215
215 215
Length of combined arrays of prime and not prime numbers ; and their labels
361 361
359 359
359 359
359 359
359 359
Length of shuffled combined arrays of prime and not prime numbers ; and their label
s
361 361
359 359
359 359
359 359
359 359
1077
1077
1077
1077
                        Train     Valid      Test
parameter metric
```

```
0.01      Accuracy  0.972377  0.957521  0.919668
          F1        0.965499  0.945835  0.894731
0.10      Accuracy  0.979109  0.957521  0.921053
          F1        0.973822   0.94485  0.896716
1.00      Accuracy  0.982591  0.962396  0.921745
          F1        0.978281  0.951205  0.897194
10.00     Accuracy  0.984448  0.960306   0.92036
          F1        0.980572  0.948092  0.895975
100.00    Accuracy  0.983983  0.961003  0.919668
          F1         0.97999  0.949089  0.894994
```



Confusion matrix for test label set:

```
Confusion Matrix for C = 1.00 and Testing set 1
[[211   5]
 [ 22 123]]
Confusion Matrix for C = 1.00 and Testing set 2
[[210   6]
 [ 20 125]]
Confusion Matrix for C = 1.00 and Testing set 3
[[208   8]
 [ 17 128]]
Confusion Matrix for C = 1.00 and Testing set 4
[[208   8]
 [ 27 118]]
```