

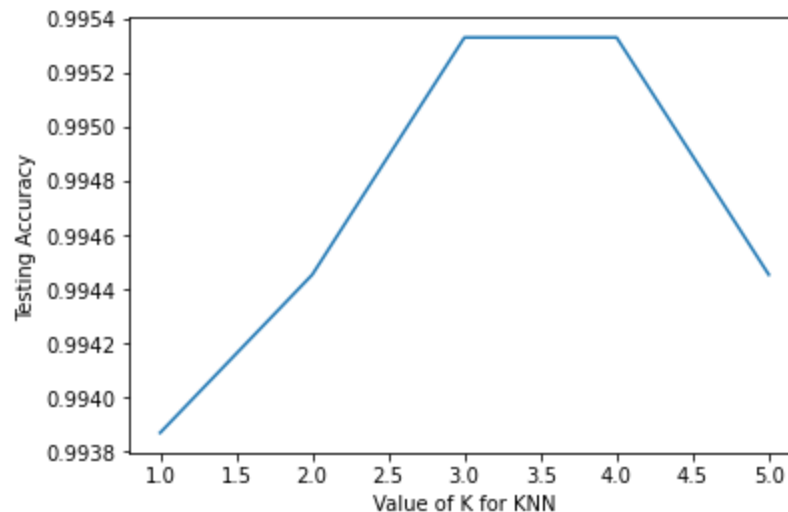
# Assignment Report

**Name:Fazal Ahmad**

**College: IIT Bombay**

## Steps followed during the Assignment:

- 1) The Dataset contains some missing data points so first I remove the rows whose data points are missing. I think removing the rows having missing data points is good because most of the missing data points are for 'forest' labels and already in the data set there are a lot of sample points for 'forest' so I think removing some forest data points will not affect the performance.
- 2) Then I have to train my model by using the cleaned data set. I used K-nearest nearest neighbour so in this model I have to decide the no. of neighbour(n) to get the n for maximum accuracy I plot the graph for different n values and for n=4 the accuracy is maximum.



- 3) Now I have to predict the output by the given 10 input images. To do this first I have to prepare the input data set for prediction so to do this I stack the pixel wise data points in a csv for 10 time intervals. So I used Numpy to convert a grayscale image to 1D array and stack all those 10 1D arrays into a CSV file. Now I am ready to predict the output so I gave this dataframe to my model input and got my output pixel wise labelling.

- 4) Next I have to convert the 1D pixel wise labeling to an image with different color for each label. For this I also used numpy to create a RGB Image.

### **Some Challenges Faced:**

1. Since the TIF Images for input data is large in size so it was not visible using the Imshow command in python so to get the image pixel wise data I have to use Matlab using MAP to see the Input Images.
2. Also in the input images there are NAN data values so I ignored the values but then I don't have the count of NAN rows so I have to keep two pointers to put the right label to right pixel