

# KNN Classifier

**Objective:** Design a KNN Classifier for the given dataset of Dogs, Cats and Pandas.

**Dataset:**

Training Dataset	70%	700 Dogs, 700 Cats, 700 Pandas
Validation Dataset	20%	200 Dogs, 200 Cats, 200 Pandas
Test Dataset	10%	100 Dogs, 100 Cats, 100 Pandas

**Algorithm:**

1. Resize all 3000 Images into 32\*32\*3 images.
  - a) `cv2.resize(image, (32,32))` is used to resize the images.
2. Convert resized images into row vectors.
  - a) 3-dimensional images( dim: 32 X 32 X 3) are converted into row vectors (dim: 3072 X 1). All values in 3-d matrix are sequentially appended to rowVector.
  - b) This is done to simplify labelling and other math calculations.
3. Label all 3000 row vectors.
  - a) One more element is appended to each row vector to store label of images.

Label	Animal inside Image
'd'	Dog
'c'	Cat
'p'	Panda

- b) Each row vector now has 3072 raw pixel values and label of image.
4. Divide the dataset into training(70%), validation(20%) and test datasets(10%).
    - a) Walk through all the files using “os.walk()” function.
  5. Calculate L1 Distance of each image in Validation dataset from all images in Training dataset.
  6. Pick the best K images and vote the labels. Label voted here is a predicted label.
    - a) Pick a value of K. (eg: 40)
    - b) Images with least K distances are picked and respective labels are voted.
    - c) Predicted label = Voted Label
  7. Repeat step 6 for different values of K and build confusion matrix with predicted labels and actual labels for all values of K

8. Calculate Accuracy for all K. Pick the one with highest accuracy.

<u>Value of K</u>	<u>Accuracy(%)</u>
40	69.33
100	70.00
120	70.44
130	69.99
140	70.66
160	71.11
200	71.11
320	71.33
<b><u>350</u></b>	<b><u>72.22</u></b>
400	71.55
450	71.77

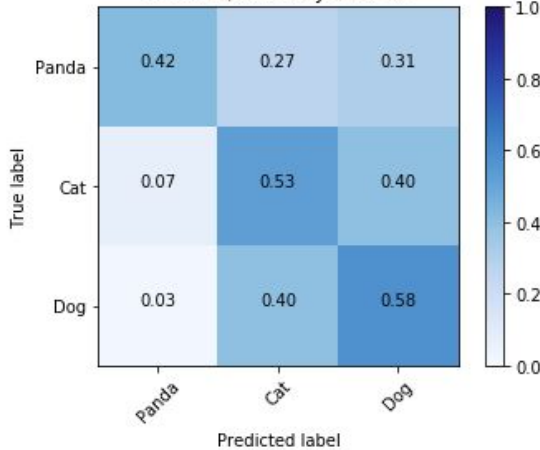
**Note**: There is no need of calculating Precision or recall as we have balanced dataset( Number of Dog Images = Number of Cat Images = Number of Panda Images)

9. Once K is finalized, run code on test dataset with finalized value of K(Here, K = 350).  
10. Calculate Accuracy, Precision, Recall, F-measure for test dataset.

**Results:**

Best value of K = 350.(Accuracy on Validation dataset is 72.22%)

Our KNN Classifier with K = 350 gave following results on test dataset.

Accuracy(%)	67.33																
Precision	0.57																
Recall	0.51																
F-Measure	0.51																
Confusion Matrix	<div><p>K = 350/Accuracy=67.3%</p><table><tr><th></th><th>Panda</th><th>Cat</th><th>Dog</th></tr><tr><th>Panda</th><td>0.42</td><td>0.27</td><td>0.31</td></tr><tr><th>Cat</th><td>0.07</td><td>0.53</td><td>0.40</td></tr><tr><th>Dog</th><td>0.03</td><td>0.40</td><td>0.58</td></tr></table></div>		Panda	Cat	Dog	Panda	0.42	0.27	0.31	Cat	0.07	0.53	0.40	Dog	0.03	0.40	0.58
	Panda	Cat	Dog														
Panda	0.42	0.27	0.31														
Cat	0.07	0.53	0.40														
Dog	0.03	0.40	0.58														