I have applied the K-Nearest Neighbours algorithm (K-NN). K-NN is a supervised multi-class classifier model. While building the K-NN model we have to be careful in data cleaning, outlier treatment and data should be normalized.

Objective:

Suppose new animals are to be introduced into the zoo, but we want to know to which ***Class*** they belong. So, for this identification, we use the K-NN algorithm, as it works on similarity measures. The K-NN model will find the similar features of the new datasets to the animals which are already present in the zoo and based on the features it will classify the animals into the suitable class.

Outline of the dataset.

- The dataset consists of 101 animals from the zoo.
- There are various characteristics to describe the animal.
  All the characteristics, except legs are classified as binary values (1-Present/0-Absent). Except for legs which it is given as a count of legs for a particular animal.
- There are 7 class types.

In order to test the model accuracy.

- I have applied GridSearchCV for hyperparameter tuning and selecting the best estimator with an optimal score. The metric is Minkowski's inequality with p=2 which is equivalent to Euclidean distance.
- I have plotted the plots for accuracy results and error rates.
- We will perform cross-validation, split the data into train and test datasets. Build a model on training dataset and check the training accuracy.

Observations:

- The model accuracy is very good for the lower values of K and as the value increases the accuracy goes on decreasing. Also, the error rate increases with higher values of K.
  - After hyperparameter tuning, the optimal value of K is 3.
- While evaluating the model performance the accuracy score using train_test_split is found to be 94%.
- From the classification report, we observe precision and recall for each ***Class*** and the confusion matrix gives a clear idea of why the accuracy is not 100% as some of the animals belonging to the actual class are marked into a different class.