**Task-3:**

The task requires using the dataset provided in this link - <https://www.kaggle.com/rajanand/rajyasabha>. This dataset contains information regarding the questions asked in the Rajya Sabha. The task is to **predict the ministry**, provided with all the other details. You are expected to provide different approaches that would be feasible for this problem. We are expecting you to justify your choice for the model, hyper-parameters, etc. It is strongly encouraged to support your arguments through exploratory analysis of the data.

A few points to note:

1. The ministry in question appears in the initial part of the answer for every question. But, this part of the answer will be stripped off, during testing of the idea.
2. Please do not suggest finding the ministry using the name of the Minister/Ministers in charge.

<< Add your solutions here >>

As this is a multiclass classification problem(predicting the ministry), a problem of this type can be solved in various methods like neural networks(single or multi-layer perceptron), Decision trees,k-nearest neighbors, SVM etc. But I will take the neural network approach here as the dataset is large (and neural networks perform well in large datasets). The main part of this process will be pre-processing the data than building the neural network.First we have to know number of inputs tokens into the neural network. So we combine all questions and answers into a single text file, remove stop words , stem the words , remove all special characters using the nltk library and common string.