Week 9:

1. Started building the dataset. The dataset is created using tweets from different sources. We have fetched data separately for different scenarios. We have 2 columns, one for the text and the other is for the label (0, 1, 2, 3, 4).
   1. Pakistan, Nepal, Chile, California Earthquakes for disaster tag (Label 1)
   2. Ebola virus outbreak, H1N1 2009 outbreak and influenza outbreaks (Label 2)
   3. Pulwama attacks, Peshawar attacks, uri attacks and mumbai attacks (Label 3)
   4. Nepal floods, India floods, Hurricane Mexico, Philippines Typhoon and Cyclone Pam (Label 4)
   5. Normal everyday tweets about randomly chosen topics to train our model for tweets not belonging to any category (Label 0)
2. Started coding preprocessor function.