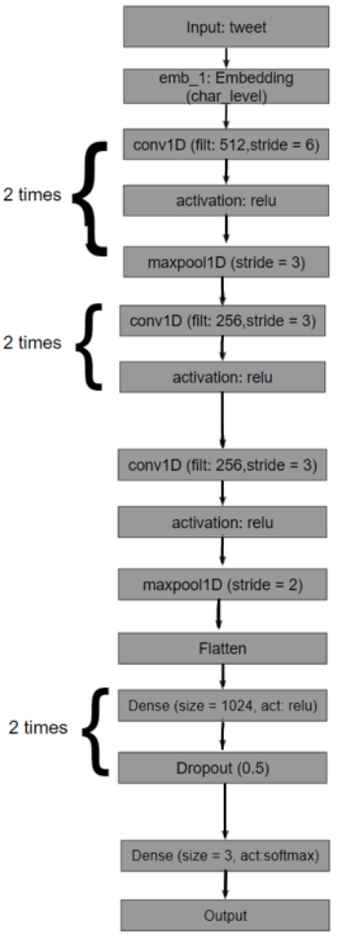
NLP Assignment 3: Sentiment Analysis

Harshil Jain (Roll Number – 17110060)

Link to Github repository: https://github.com/jain-harshil/NLP-Assignment-1/tree/master/NLP_Assignment_3

Data Cleaning: Created a csv of the train and test dataset and created lists of tweets and corresponding labels (negative, neutral and positive)

Description of Architecture: Since the tweets dataset provided contains both Hindi ad English mixed words, it was not possible to create embeddings at word level. So, the embeddings at character level were created. The total unique tokens were 900 after tokenization was done. Since the maximum number of characters of a tweet are 140, all tweets were post padded to make the maximum size of a tweet to length 150. The model has the following architecture which consists of Convolution1D, Maxpooling1D and Dense layers:



Model Architecture

The model was trained for 25 epochs with a batch size of 128 and Adam optimizer, categorical cross entropy loss and learning rate of 0.01 gave an accuracy of 98.70% on the train dataset and an accuracy of 52.17% on the test dataset. The precision, recall and F1 scores are as follows: (Here 0 – negative, 1 – neutral and 2 – positive)

	precision	recall	f1-score
0	0.56	0.48	0.52
1	0.48	0.61	0.54
2	0.58	0.44	0.50