

Attendance: 10%, Continuous evaluation: 70%, Viva-20%**Assignment No. 7**

- i. Download and preprocess the sentiment analysis dataset from <https://www.kaggle.com/snap/amazon-fine-food-reviews>. Download the Glove word vectors from <http://nlp.stanford.edu/data/glove.6B.zip> and extract the 100-dimensional file (glove.6B.100d.txt) from the zipped folder.
- ii. Preprocess the review dataset by considering the column “review score” >3 as positive reviews and others as negative reviews. For training on local machine considers 5000 positive and negative reviews each for the training dataset.
Consider 2000 reviews for the test dataset and validation dataset each. Strip the length of each review sentence (number of words) according to your computation availability.
- iii. Train RNN model with the FC layer applied in the final hidden layer output using the following parameter:

Sr. No:	RNN	RNN Layer	LSTM size	Activation	FC layer	Embedding Layer
1.	LSTM	1	64	Relu	1	Glove
2.	GRU	1	64	Relu	1	Glove

- iv. For the best model above vary the size of RNN : [32, 128]
- v. For the best model above vary the number of stack layers of RNN : [2, 3, 4]. One is done previously.
- vi. For the best model above run a bidirectional RNN model: One is done previously.
- vii. For the best model above try Dropout: 0.1, Recurrent Dropout:0.2, and both together. Explore any other regularization parameter.
- viii. For the best model above consider training a self trainable embedding layer, and one hot encoding layer. Discuss the major differences in performance.
- ix. For the best model above consider training a self trainable embedding layer, and one hot encoding layer. Discuss the major differences in performance.
- x. Compare the number of parameters, training and inference computation time, Training Loss graph (preferably in a single graph), accuracy.

- xi. Write a review of your own and test your model. Save the model checkpoint for later use. [Note: To verify the best model is saved, re-run the notebook and only perform testing]
- xii. For the best model try the Hindi movie review dataset <https://www.kaggle.com/disisbig/hindi-movie-reviews-dataset> (use self trainable embedding layer or any other Hindi Word2Vec representation).
- xiii. Discuss the time required and other practical challenges in training with the whole Amazon review dataset.

Submit a report with results.