**Natural Language Processing (NLP)**

Link-https://www.kaggle.com/nitishputrevu/nlp-snli-dataset

Task1-To determine the inference relationship namely,*entailment, contradiction, or neutral*(SNLI), between the sentence pairs

Steps-

1)Loaded the json files in three variables-train test and dev

2)Dropped sentence pairs from train test and dev which has no tags determining the inference relationship (gold labels).

3)Introduced dictionaries-training\_X, testing\_X and dev\_X which contained both the sentence pairs

4)Introduced lists called training\_Y, testing\_Y and dev\_Y which contained manually encoded representations of the inference relationship between the sentence pairs

*(Entailment-[100000000000000], Contradiction-[010000000000000],  Neutral-[001000000000000]*

*5)Tokenised and padded (post truncation with maxlength 50) each sentence present in the sentence pairs and then represented one sentence pair as the difference between the 2 pairs*

*6)Introduced the model I have used-BIDICTIONAL LSTM*

model = tf.keras.Sequential([

tf.keras.layers.Embedding(1000, 16, input\_length=50),

tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(32,input\_shape=(549367,50),recurrent\_dropout=0.2,return\_sequences=True)),

tf.keras.layers.Flatten(),

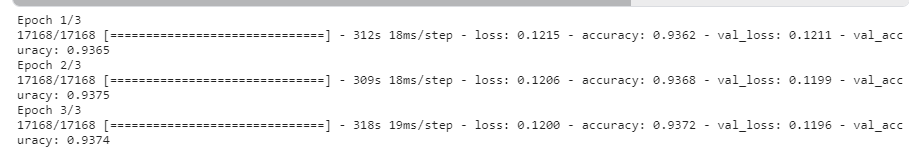
tf.keras.layers.Dense(24, activation='relu'),

tf.keras.layers.Dense(4, activation='softmax')

])

model.compile(loss='sparse\_categorical\_crossentropy',optimizer='adam',metrics=['accuracy'])

7)After fitting,using validation data (dev\_X and dev\_Y),measured accuracy using test data and hence predicted the inference relationship between the sentence pairs

**

Results-

Task2-

1. Introduced a variable-tagged\_sentence which contains sentences with tags from the nltk.corpus.treebank.tagged\_sents(tagset='universal')
2. Separated the sentences and tags in two separate lists N1\_X and N1\_Y respectively
3. Introduced list called num\_N1\_Y which contained manually encoded representations of the inference relationship between the sentence pairs
4. *(*1-. 2-ADJ', 3-'ADP', 4-'ADV', 5-'CONJ', 6-'DET', 7-'NOUN', 8-'NUM', 9-'PRON', 10-'PRT', 11-'VERB', 12-'X'*)*
5. *Used train test split and obtained data for variables train\_X, validation\_X(N1\_X), train\_Y and validation\_Y(Num\_N1\_Y)*
6. *Fitted the data using the same model and obtained predictions of sentences*

*Observations-Acuracy of 75,55% was obtained showing that the model does predict inputs to a certain extent based on parts of speech.*