NAME: Renu Tamsekar

ROLL NO: 3253

TITLE : sentiment analysis using RNN/ LSTM

LAB NO:

PROBLEM STATEMENT :  develop an algorithm and write a program for sentiment analysis of IMDB movie reviews using RNN/LSTM

METHODOLOGY :

* Pre-process the test data by performing tokenization, lower casing, stop word removal, stemming or lemmatization and other NLP techniques.
* Encode the pre-processing data into numerical vectors using one hot encoding, word embedding.
* Split the encoded data into training and validation and testing sets.
* Build the LSTM model using sequential architecture in keras.
* Compile the model using a suitable loss function, optimizer and evaluation metrics.
* Train the LSTM model on training data using BPTT. Use the validation set to monitor the performance.
* Evaluate performance of model on test data.

APPLICATIONS:

* Prediction problems.
* Machine Translation.
* Speech Recognition.
* Language Modelling and Generating Text.
* Video Tagging.
* Generating Image Descriptions.
* Text Summarization.
* Call Center Analysis.

RESULTS :

* Accuracy of 83.06% is achieved.
* Reviews have been classified as positive or negative.

OBSERVATION :

* LSTM can work better at classifying the model compared to RNN since accuracy of RNN is less than LSTM.

CONCLUSION :

* LSTM is a part of RM architecture that is designed to solve the vanishing gradient problem.
* The LSTM has 3 main components: input gate, forget gate, output gate
* The input gate contains how much new information is allocated into the cell. It is a sigmoidal layer and outputs a value between 0 to 1. If its closer to 0, then no new information and If 1 then new information is added.
* The forget gate controls how much information is allowed to be retained in the cell. It is a sigmoidal layer and outputs a value between 0 to 1. 0 means old information can be forgotten and 1 means old information is retained.
* The output gate controls how much output information is allowed from the cell. It’s a sigmoidal layer that outputs a value between 0 to 1. It also has a tanh layer that provides a vector of new candidate values. The output gate combines these 2 components to provide the output.

PRINT OF CODE AND OUTPUT