For Movie Reviews, A Sentiment Analysis using Long Short Term Memory Networks

R.Pushpakumar[a], Sakunthala prabha.K.S[b] and P N Karthikayan[c]

[a] Information Technology, VelTech Rangarajan Dr.Sagunthala R&D Institute of
Science and Technology, Chennai, India
[b]Information Technology, VelTech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India
[c]Information Technology, VelTech Rangarajan Dr.Sagunthala R&D Institute of
Science and Technology, Chennai, India

_____

Abstract: The sentimental or opinion details in web pages rises in parallel with the development of web 2.0. Capturing such sentimental information from the web 2.0, is a challenging task. The problem is addressed in this paper by introducing a new paradigm for implementing Long Short-Term Memory Networks. Word2Vec Model is used to transform sentimental contents into input embedding vectors. These input embedding vectors are then fed into the LSTM, which predicts the content's sentiment details. The suggested approach is put to the test on two well-known datasets: IMDB and Amazon Analysis. In the IMDB dataset, the suggested approach achieves an accuracy of 0.87, and in the Amazon analysis dataset, it achieves an accuracy of 0.89.

1. Introduction

Opinion mining and inclination analysis are two terms that can be used to describe sentiment analysis. It involves subjectively evaluating, processing, triggering, and inferring text with emotion. It has a wide variety of applications in the industry, ranging from demand forecasting based on blog and message expressions to determining consumer or public satisfaction or disappointment from feedback. Word vectorization is the process of converting words into vectors. A Word vector algorithm uses a large corpus to create a vector mapping of similar values and simply the word appearing in the same sense. It feeds these vectors into other machine learning languages including neural networks, SVM, and so on, with the intention of allowing machines to comprehend human languages. The skip-gram models aid learning more quickly than the continuous bag-of-words model. The input layer, secret layer, and output layer are the key components of a recurrent neural network. There will be a relationship between the front and the back for certain text data, as well as a temporal relationship between the data. RNNs are useful for collecting sequential information because their design includes loops that enable them to transfer information collected from previous outputs.

The list of applications contained on the web decreased in the initial stages because there were only two choices available, such as NLP libraries for a python and R. Such programs are Seniti Power, Rapid Miner, Lingpipe and others with certain usage restrictions on large datasets, trail times, free version time constraints, all those kinds of restrictions would not be scalable for large datasets to be used. The advantage of using sentiment analysis is that we can use it for both academic and commercial purposes as an open source tool, and the program is free and has a lot of developers and a wide community of users. The significant function of two tools for sentiment analysis: NLTK for the python library and R sentiment for R. The parameter includes performance, computation time, and number of classes in output as some classification. Building and training the NLP model, which is an advanced issue, requires the ability to classify a broader spectrum of emotions.

The English sentence will evaluate and allocate the score to that sentence, and the sentence can be classified by positive, bad, very positive, very negative, and neutral types of feelings. It counts the number of phrases in each sentiment group, and it takes into account degrees of adjectives by measuring the ranking, negation and different. There are only three roles in the library that provide detailed information about a piece of text that measures ratings, which gives an emotion a numerical parameter. Then, measure the feeling where the text was found that returns the emotion. Finally, it returned a matrix of the respective counted feelings in sections of the full text to measure the overall presence of emotion. It takes the tests from a web source for text testing instruments. It has a wide range of applications in the industry from forecasting markets based on the expressions on the blogs and the messages to identify the customer or public satisfaction or dissatisfaction from the reviews. It helps to understand the people's emotions for the businesses where a customer can express their thoughts and feelings more openly than ever before by using automatic analyzing customer feedback from the survey responses on social media about the brands and their services. So, it is the customers able to listen to the products to meet their needs. It's calculable that eightieth of the world's information is unstructured; in alternative words, it's unorganized.

They are a standard-based system that employs a variety of human-made guidelines to assist in the evaluation of emotion, extremity, or partner degree. These criteria could include a variety of etymological

procedures such as stemming, auto ionisation, grammatical type marking, and parsing. Vocabularies The paradigm restores a positive slant if the number of positive word appearances is greater than the number of negative word appearances, and vice versa. If the numbers are equal, the system will come to an impartial conclusion. Rule-based systems are extremely simple because they don't consider the terms that are joined during a grouping. Obviously, cutting-edge measurement techniques are used, and modern guidelines provide support for the new articulations and jargon. Adding new principles, however, can have an effect on previous results, causing the whole system to become muddled. Since rule-based systems often require adjusting and maintenance, they should be examined on a regular basis. Abstract investigation, evaluation survey, and decision extraction are only a few of the many names for estimation test. The conclusion examination's primary goal is to separate details from knowledge in a content framework.

A individual, the general public, or a customer summarises the sentiment analysis. It briefly describes and contrasts a number of traditional approaches. The aim of text classification is to minimise the intrusion caused by the goal text in the study, and it also serves to classify subjective text by dividing emotion into multiple categories based on people's emotional expressions for the purpose of analysing specific circumstances. The analysed text has been divided into three levels: chapter-level, paragraph-level, and sentence-level. Text processing methods can vary depending on the length of the text.

The assignment at this level is to decide the general assessment of the report. Slant examination at the report level accepts that each archive communicates assessments on a solitary substance. It will give a general rating to a subject. In numerous applications, the client has to know the parts of substances that enjoyed and loathed. Looking at elements isn't effectively material at the report level like conversations, sites, and news stories. It is critical to move at the sentence level; for example, each sentence expresses an evaluation. The record level and the sentence level have no distinction. The assumption investigation calculation comprehends language word by word from the unique situation and word request. Be that as it may, our dialects are inconspicuous, nuanced, endlessly mind boggling, and trapped with the feeling. Unsophisticated assumption examination strategies ascertain feeling, extremity by coordinating words back to a word reference of words hailed as "good," "unbiased," or "negative." This methodology is excessively reductive. The data and adulterates our grammatically intricate, lexically rich language. We tune in to a whole sentence and drive, implying that is the gestalt, or more noteworthy than the amount of the individual words. Additionally, we parse approaching words through the intricate latticework of deep-rooted social learning.

In this paper, we focused to classify the movie review by using the RNN (Recurrent Neural Network). Our fundamental problem in sentiment analysis is classifying the sentiment repulsion that when a piece of text given, then the problem is to classify the written text into a specific sentiment polarization as positive, negative, or neutral. There is a vast quantity of data accessible online that might assist people and organizations in deciding processes however at a similar time gift several challenges as organizations and people decide to analyze and comprehend the collective opinion of others. It is impractical to manually realize opinion sources on-line, extract sentiments from them, then precise them during a normal format. In recent years social networking sites, blogs, and review sites as give a great deal of data. Immeasurable individuals express unsuppressed opinions concerning varied product options and their nuances. This forms vigorous feedback that is of importance not solely to the businesses developing the product, however additionally to their rival's alternative and several other potential customers.

2. Literature Review

Mass et.al [1] They used sentiment analysis in their expected methodology, as well as their current sentiment analysis methods. Machine learning approach, lexicon-based mostly approach, and hybrid approach square test the three major categories. Words, bigram, tri-gram, a part of expression, and polarity are some of the choices used in the machine learning method. There are numerous supervised-based techniques available. We used Pak and Paroubek's SVM and NB classifier, a replacement sub-graph-based example derived from grammar dependency trees, created by Pak and Paroubek. They put the model to the test with a series of video game reviews from France. As a result, the aim is to show that SVM (Support Vector Machine) classifier victimisation options designed from sub-graphs derived from dependency trees provide better results than older unigram-based systems. As a tool for document classification, many machine learning methods are used. NB, most entropy, KNN, and SVM are all excellent methods. They must catch the attention of researchers in deep learning methods as a result of significantly outperforming traditional methods. [10].

Socher et al [2]. They did an experiment exploitation sentiment analysis on Twitter. With the increase within social networking, there has been a surge of user-generated content. Microblogging sites have countless individuals sharing their thoughts daily. They planned a way of sentiment from a preferred real-time microblogging service, Twitter, wherever users post period of time reactions to and opinions concerning everything. Their aim is to work out the linguistics orientation of the opinion words in tweets. Microblogging platforms area unit employed by completely different individuals to precise their opinion concerning different quite topics, so the precious sources of individuals opinions. Agarwal et al. approached a task of mining sentiment from Twitter, as a three-method task of classifying sentiment into positive, negative, and neutral categories.

They experimented with a feature that supported a tree kernel for this method they designed a replacement tree illustration for tweets. The feature-based model use a hundred options and also the unigram model uses over ten thousand options. The model is been given a median performance. Firstly, they pre-process the tweets so it's been extracted by opinion intensifiers and opinion words then it'll be calculated by the evaluation module and also the score of the adjective cluster and verb cluster. By exploiting the equation, they calculate the sentiment strength of the tweet, that the tweet has been classified as positive, negative, and neutral. They finally all over that the microblogging sites like Twitter offers Associate in Nursing unprecedented opportunity to make and use theories and technologies that search and mine for sentiments and this novel represent the approach of sentiment analysis on Twitter knowledge to uncover the information they extracted the opinion words within the tweets primarily based on corpus-based methodology was wont to notice the linguistics orientation of adjectives at the side of the dictionary-based methodology to seek out the verbs and also the adverbs of linguistics orientation. However, this has been thought-about an image and it has been evaluated as a preliminary image and it's been showing some improper accuracy as result and there try to develop this project with another option.

Wawre et al [3]. They did a piece on investigation of the Brazilian Portuguese LIWC wordbook for notion examination. This has been assessed by 2 styles of estimation assets for the Portuguese language as assessment dictionary and sentilex. In this work, they assessed the use of Brazilian Portuguese LIWC wordbook opinion grouping in Brazilian Portuguese writings. They looked at the asset against 2 elective feeling assets available for Portuguese: the assessment dictionary and in this manner the sentilex. The 2 examinations are presented as characteristic and extraneous assessment. Their point is to create new experiences with respect to anyway these vocabularies would be useful in slant examination and their fundamental qualities. The algorithmic program was used in this undertaking is tantamount to the SO-CAL portray which figures the individual extremity for each word inside the vocabulary so summarizes these polarities structure the content extremity. In the event that the add is zero, the content is evaluated as impartial, in the event that the add is bigger than nothing, at that point the content has been delegated positive, in any case, negative. The LIWC keeps on claiming intense to mark negative (28.95%), notwithstanding, it has a high score for the positive classification (74.48%). These outcomes are could accept that the LIWC wordbook performs higher demonstrating inspiration and pessimism. At long last, they presumed that this examination chips away at the reason of dictionary principally based assumption partner lysis by leading the 2 investigation: an inborn assessment, by measure the arrangement contrasted and 2 unique vocabularies; related an external examination, by measure the vocabulary sway in a surpassing supposition order task.

Pennington et.al [4] recommended new product options mining that made use of grammar dependency information in a novel way by discriminating between nominal and non-nominal words. Non-nominal terms would be treated as linguistics neighbours of connected nominal terms, and nominal linguistics structures would be parsed using the concept of dependency trees as part of the expected model. Linguistic structure parsing can produce a narrow-minded pair stream with few nominal words, similar to their linguistics neighbours. On the basis of fine-grained product attributes, co-clustering methodology by resolving is also not heritable. The planned structure was found to have advantages in terms of average cluster entropy, mental uncertainty, and manual analysis.

A tale generative theme model, the Joint feature/Sentiment (JAS) model, for separating together angles moreover as viewpoint subordinate supposition dictionaries from on-line client audits was projected at [4]. Partner angle subordinate assumption vocabulary noticed viewpoint explicit assessment words with their perspective mindful supposition polarities with pertinence to a chose aspect. Analysis unconcealed the JAS model's viability in learning aspect subordinate supposition vocabularies and separated dictionaries reasonable qualities once utilized in reasonable errands. Li established fine-grained product options extraction using the simplest custom-built language models as labelling For opinion options mining, a threshold-normalized sentence-level word model is typically recommended. Matrix resolution was used to resolve the extraction of opinion choices. Machine learning was used to extract appropriate, fine-grained opinion choices. Extraction of product options is treated as a sequence labelling task using Conditional Random Fields, a type of discriminative learning model. (CRFs) for attempt it was prompt by Huang et al., (2015)[11]. It integrated options for sections of speech and sentence form into a CRF learning protocol. It implemented spatial structure context-based similitude metrics for calculating similitude between product attribute expressions for product feature categorization, as well as linguistics knowledge-based similitude metrics. It is planned to develop a cost-effective graph pruning-based categorization protocol for classifying the feature expressions collection into various linguistics groups. Experiments showed that the strategy was superior to other approaches in terms of efficacy. [6].

Maximum Entropy could be a likelihood distribution estimation technique that's extensively used for Natural Language process tasks. The instinct that spurs most Entropy characterization is that we should consistently assemble a model that catches the frequencies of individual joint-highlights, without making any unjustifiable suppositions. The essential principle of most Entropy is that the probability distribution ought to be uniform once there's no pre-knowledge. During this paper Maximum Entropy classification is employed to estimate the polarity of Chinese reviews. There's no totally different from different learning technique, the outputs of machine learning technique are relied on the given coaching Data-set of input. Once victimization most Entropy classification, the first step is to induce the constraints that characterize the class-specific expectations for the

distribution from labeled coaching data-set for the model distribution. Figure one is that the most Entropy model for Chinese reviews classification that consists of a coaching method and a testing method. Within the coaching method, a tagged coaching data-set is employed to derive a collection of constrains for building the model that characterizes the class-specific expectations for the distribution. Finally, we tend to use the final reiterative Scaling algorithmic rule to seek out the utmost Entropy distribution that's in line with the given constraints. Within the method of classification, the testing data-set is denoted in options, and so the review classification is obtained by victimization the classifier [7].

The input to recursive neural tensor networks may be any length of expression. They use word vectors and a parse tree to define a concept, then use the same tensor-based composition function to compute vectors for the tree's higher nodes. [8].

3. Proposed Method

In this paper, we proposed sentiment analysis to predicting the data from a dataset is IMDB database for the review from the people or the information from the viewers. Then, we are pre-processing the data by tokenizing and normalizing the data into vectors and removing the casing characters, negation handling, and removing it. Then, we will do feature extraction by using the skip-gram which is used to extract the words into index mapping to make the word which will make a vector that is equal to the total number of the words present in the data and it will assign a value for the index of the particular word. We use LSTM for our proposed technique which is to recollect each snippet of data through time and the valuable in time arrangement expectation simply because of the component to recall past contribution too and this additionally called Long Short-Term Memory. To provide the accurate value for providing the positive reviews about the movie or a product that will be provided to the respective renders.
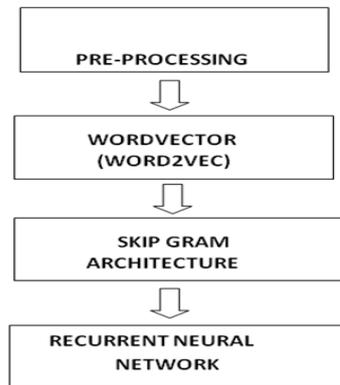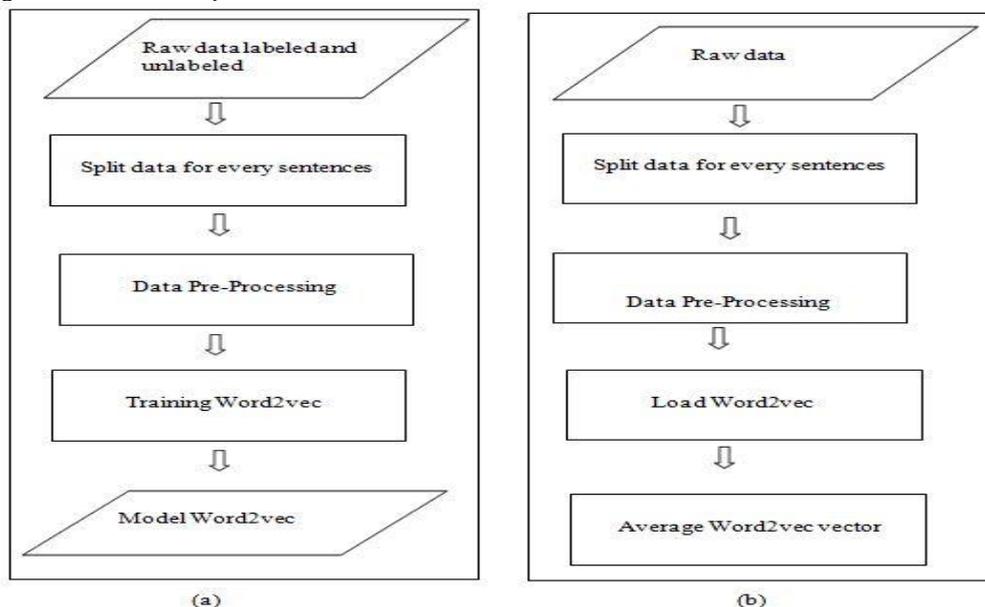

Fig 1 function module

3.1 Pre-Processing

Tokenizing and Normalization complete the pre-processing. Tokenizing is the process of translating text into tokens before converting it to vectors. It completes normalisation by casing the characters, treating negations, and deleting them. Stop words are the most popular words that aren't applicable to the data and don't add any additional significance to the expression.

Flowchart of word2vec (a) is used to the training process and (b) is used to make sample vectors.

Fig 2. Pre-Processing

Word2vec is a computational tool for learning a single word from a text corpus rapidly and efficiently. As a result of the need to make neural-network based embedding training more effective, it has since become the defector standard for creating pre-trained word embeddings.

The study entailed analysing the learned vectors and experimenting with vector math on word representations. For example, removing the "man-ness" from "king" and replacing it with "women-ness" yields the word "queen," which captures the analogy "the king is to queen as man is to woman." Vector-oriented reasoning based on offsets between words is possible because the representations are amazingly good at capturing syntactic and semantic regularities in language, and each relationship is represented by a relation-specific vector offset. [5].
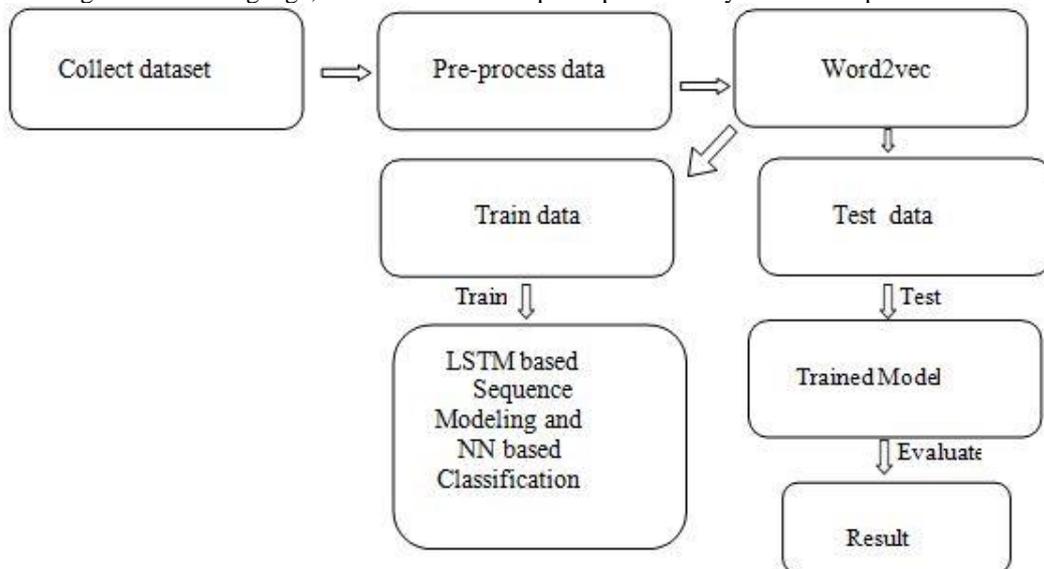

Fig 3 Word2Vec

Word2vec is a set of similar models for creating word embeddings. These models are two-layered shallow neural networks that have been trained to recreate the meaning of words or data.

These models are extremely successful at understanding the meaning and relationships between words. In the vector space, related terms are grouped together, while dissimilar words are spread out.

Skip-Gram

The neural networks start with a single word and then try to predict the words around it. To train the data and construct the vectors, the neural network has one input layer, one hidden layer, and one output layer. We'll use a single hidden layer to train a neural network to perform a task, but we won't use the neural network to perform the task we trained it on.

In NLP, The first step is to assign unique indices to all words by using a single hot encoding for a word from our dictionary. This can be achieved by grouping the words alphabetically ascending or descending, or by arranging them in some random order, with the intention of mapping all of the words to a single index.
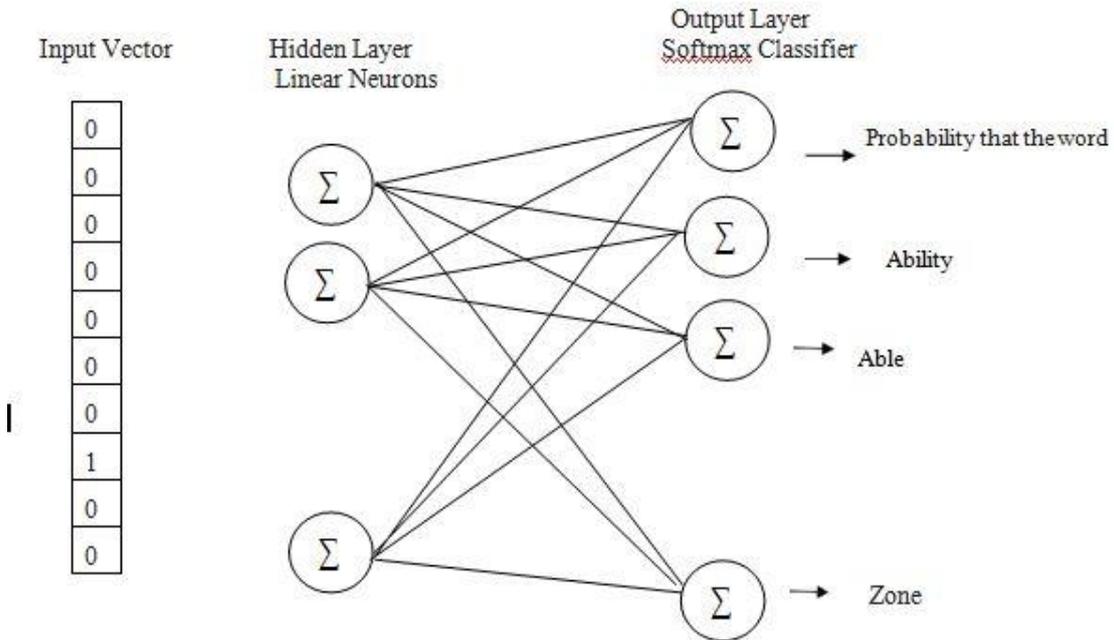
Fig 4 Working model of skip gram

Working process of Skip gram:

The words are recovering into a vector exploitation one hot coding. The component of those vectors is [1, V]. The word w (t) is passed to the stowed away layer from |v| neurons.

1. The shrouded layer plays out the internal item between weight vector W [v, N] and furthermore the information vector w(t). In this, we will reason that the (t) toss of W [v, N] are the yield (H [1, N]).

2. To remember there's no initiation perform utilized at the concealed layer thusly the H [1, k] will be given to the yield layer.

3. The yield layer can apply an internal item between H [1, N] and W [N, v] and can give the North American country the vector U.

4. Now, to look out the probability of each vector, we'll utilize the softmax perform. As each cycle gives yield vector U that is of 1 hot coding kind.

5. The word with the best probability is that the outcome and on the off chance that the foreseen word for a given setting position isn't right, we'll use back engendering to change our weight vectors W and W'.

This progression is dead for each word w (t) blessing in jargon. What's more, every word w(t) is spent k occasions. In this way, we will see that forward engendering is prepared |v|*k times in each age.

All the more officially, given an arrangement of preparing words w1, w2, w3, . . . , wT , the goal of the Skip-gram model is to boost the normal log probability)training time.

$$\frac{1}{T} \sum_{t=1-c \leq j \leq c, j \neq 0}^{T} \sum \log p(w_{t+j} | w_t)$$

The basic Skip-gram formulation defines $p(w_{t+j})$ using the softmax function:

$$p(wO|wI) = \frac{exp(v'_{wo}{}^T v_{w_t})}{\sum_{w=1}^{W} exp(v'_w{}^T v_{wI})}$$

where v_w and v'_w are the "input" and "output" vector representations of w, and W is the number of words in the vocabulary. To reduce the time complexity from above equation we apply hierarchical softmax function which is formulated as follows.

$$p(w|wI) = \prod_{j=1}^{L(w)-1} \sigma([n(w, j+1) = ch(n(w,j))].v'_{n(w,j)}{}^T v_{wI})$$

In an LSTM neural network, the output from the previous step is fed as input to the current step. It is a neural network that is specialized for processing a sequence of data with the time step-index(t). It's because they have a memory that holds knowledge about previous calculations. It was built with the intention of resolving this problem using a hidden layer. The important feature of LSTM is that it has a secret state that remembers some sequence information.
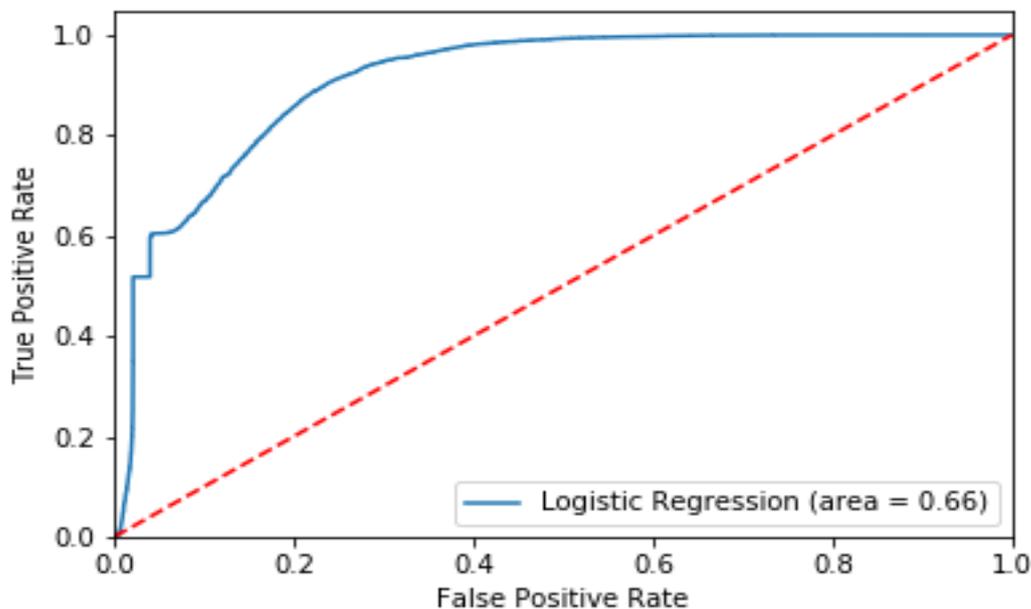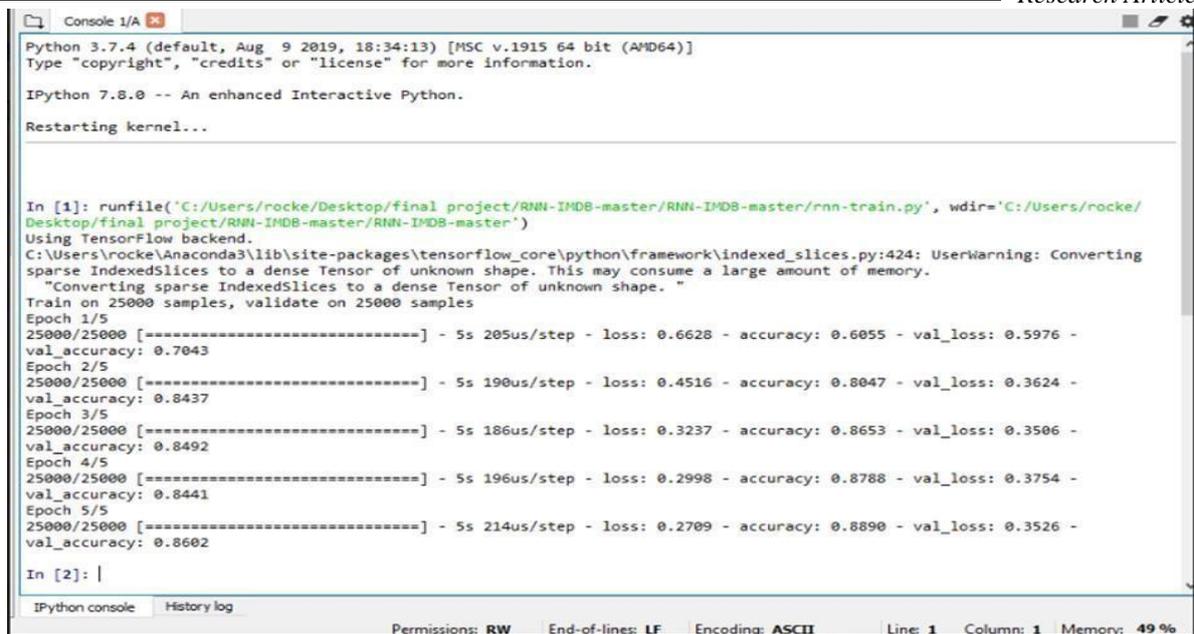
4. Dataset

The IMDB dataset is a web-based database with information about movies, TV shows, and streaming content, including cast, crew, and personal biographies. All web clients can access IMDB's film and skill pages, but to add data to the website, an enlistment interaction is needed. Volunteer patrons have the bulk of the knowledge base. Clients may use the web to apply new material and make changes to existing passages.

The dataset is the huge film survey dataset regularly alluded to as the IMDB dataset. The enormous film survey dataset contains 25,000 profoundly polar moving audits (positive or negative) for preparing and a similar sum again for testing. The issue is to decide if a given moving audit has appositive or negative estimation. The preparation set is equivalent to 25,000 marked audits used to prompt word vectors with our model. At that point it assesses the classifier execution after cross-approving classifier boundaries on the preparation set, again utilizing a straight SVM taking all things together cases. It shows the grouping execution on our subset of IMDB surveys. Our model showed better execution than different methodologies and performed best when connected with a pack of words portrayal. Once more, the variation of our model which used extra unlabeled information during preparing performed best. Contrasts in precision are little, but since our test set contains 25,000 models, the fluctuation of the exhibition gauge is very low. For e.g., a precision increment of 0.1% compares to effectively arranging an extra 25 surveys.

5. Result Analysis

We prepared our model on IMDB informational collection, which it is contains around 50000 film audits, among this 25000 positive conclusions and 25000 negative feelings present. We tried our model around 2000 film audits, among this 1000 positive notions and 1000 negative suppositions, tried effectively. The amazon reviews audit dataset contains item surveys and metadata from Amazon, including 142.8 million surveys. This dataset incorporates audits, item metadata and connections. The testing precision accomplished by the proposed model is 0.86. The figure 1 shows the precision of the proposed model after 5 ages.

```
Console 1/A

Python 3.7.4 (default, Aug  9 2019, 18:34:13) [MSC v.1915 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 7.8.0 -- An enhanced Interactive Python.

Restarting kernel...


In [1]: runfile('C:/Users/rocke/Desktop/final project/RNN-IMDB-master/RNN-IMDB-master/rnn-train.py', wdir='C:/Users/rocke/
Desktop/final project/RNN-IMDB-master/RNN-IMDB-master')
Using TensorFlow backend.
C:\Users\rocke\Anaconda3\lib\site-packages\tensorflow_core\python\framework\indexed_slices.py:424: UserWarning: Converting
sparse IndexedSlices to a dense Tensor of unknown shape. This may consume a large amount of memory.
  "Converting sparse IndexedSlices to a dense Tensor of unknown shape. "
Train on 25000 samples, validate on 25000 samples
Epoch 1/5
25000/25000 [==============================] - 5s 205us/step - loss: 0.6628 - accuracy: 0.6055 - val_loss: 0.5976 -
val_accuracy: 0.7043
Epoch 2/5
25000/25000 [==============================] - 5s 190us/step - loss: 0.4516 - accuracy: 0.8047 - val_loss: 0.3624 -
val_accuracy: 0.8437
Epoch 3/5
25000/25000 [==============================] - 5s 186us/step - loss: 0.3237 - accuracy: 0.8653 - val_loss: 0.3506 -
val_accuracy: 0.8492
Epoch 4/5
25000/25000 [==============================] - 5s 196us/step - loss: 0.2998 - accuracy: 0.8788 - val_loss: 0.3754 -
val_accuracy: 0.8441
Epoch 5/5
25000/25000 [==============================] - 5s 214us/step - loss: 0.2709 - accuracy: 0.8890 - val_loss: 0.3526 -
val_accuracy: 0.8602

In [2]:

IPython console    History log

Permissions: RW    End-of-lines: LF    Encoding: ASCII    Line: 1    Column: 1    Memory: 49 %
```

Fig 5 Precision and Recall graph

Fig 6 Result after Prediction

6. Conclusion

It essentially presents diverse profound learning methodologies and text information for different classes and furthermore sums up and investigates their particular gifts and ability. This technique saves many refined strategies for complex element extraction contrasted and other strategy; the neural organizations take a word thus attempt to foresee the surrounding words. The neural association has one data layer, one covered layer, one yield layer to set up the information and create the vectors. A LSTM reviews every information through time. It's valuable in estimating conjecture solely considering the component to recall past data sources in like way. It consistently insinuates this to a Long memory Irregular Neural Network zone, until it uses with convolution layers to assemble the convincing picture part region. Tests are performed on 25 thousand IMDB picture audits and amazon reviews audits for instructing and subsequently similar numbers for testing. The intricacy of the model has misrepresented, and in this way the progressions inside the outcome are noted. The precision of the proposed strategy is observed above 86% and hence it works better compared to different techniques or models.

References:

1. Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011, June). Learning word vectors for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies (pp. 142-150) 2011.

2. Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013, October). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing (pp. 1631-1642).

3. Wawre, S. V., & Deshmukh, S. N. (2016). Sentiment classification using machine learning techniques. International Journal of Science and Research (IJSR), 5(4), 819-821,2016.

4. Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).

5. Shi, T., & Liu, Z. (2014). Linking GloVe with word2vec. arXiv preprint arXiv:1411.5595.

6. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. arXiv preprint arXiv:1310.4546..

7. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of machine learning research, 12(ARTICLE), 2493-2537.

8. Socher, R., Perelygin, A., & Jean, Y. (2013). WU; JASON CHUANG; CHRISTOPHER D. In MANNING, Andrew Y. NG et Christopher POTTS."Recursive deep models for semantic compositionality over a sentiment treebank". In: Proceedings of the conference on empirical methods in natural language processing (EMNLP). T (Vol. 1631, p. 1642).

9. Muhammad, P. F., Kusumaningrum, R., & Wibowo, A. (2021). Sentiment Analysis Using Word2vec And Long Short-Term Memory (LSTM) For Indonesian Hotel Reviews. Procedia Computer Science, 179, 728-735, 2021.

10. Sadr, H., Pedram, M. M., & Teshnehlab, M. (2021). Convolutional Neural Network Equipped with   Attention Mechanism and Transfer Learning for Enhancing Performance of Sentiment Analysis. Journal of AI and Data Mining, 2021.

11. Zheng, Shuai, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr. "Conditional random fields as recurrent neural networks." In Proceedings of the IEEE international conference on computer vision, pp. 1529-1537. 2015.