International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

Classification of Sentiments on online products using Deep Learning Model – RNN

Lakshmidevi N, M. Vamsikrishna, S. S. Nayak

Abstract: Due to advancement of technology there is a large usage of social media which leads to demand for data in the web. This data is very helpful to categorize the opinions into different sentiments and general evaluating the mood of public. The current research contributions are towards to detect the complete separation of sentence regardless of their aspects. The computational observation of sentiments and opinions stated by people in written language. Examination of defies presented by informal and crisp micro blogging created the origins. The proposed work targets building up a model for conclusion characterization that investigates the product features. It is also addresses domain explicit vocabularies to offer an domain arranged methodology and subsequently dissect and extricate the purchaser opinion towards well known advanced cell marks in the course of recent years. This model describes the use of deep learning model such as recurrent neural network to get better accuracy over traditional machine learning methods such as Random forest, Naïve bayes. The RNN model got training accuracy 97.6% and testing accuracy 95.6% which are much better compared to traditional machine learning models.

Keywords: Sentiment Analysis, SVM, Opinion mining, Recurrent Neural Network, Natural Language Processing, Naive Bayes, Random Forest, Product Reviews.

I. INTRODUCTION

According to the internet statistics 2019[4], there are 4.1 billion internet users in the world. The online medium has been changed into a noteworthy way for persons to express their views and with web-based social networking and there is a prosperity of feeling data accessible. opinion mining is the computational investigation of individuals' feelings, slants, evaluations, and characteristics towards substances, for example, items, administrations, associations, individual's, issues, occasions, subjects, and their traits. It intends to decide the mentality of a speaker as for some subject or the general relevant extremity of the document. It is the process of determining the polarity i.e. Neutral, Positive, Negative feeling toward a subject. People can detect the assumption effectively; be that as it may, it is tedious, conflicting, and expensive in a business setting. Humans have limited capacity to work. Hence, it's just not realistic to have people individually read millions of user review and scores them for sentiment.

Researchers have mainly studied sentiment analysis at three levels of granularity [1]:

1. Document level
2. Sentence level
3. Aspect level

Document level notion classification characterizes a stubborn document as communicating a complete positive, negative or unbiased conclusion. It thinks about the entire archive as the fundamental data unit and accept that the document is known to be obstinate and contain conclusions about a solitary element [2]. Sentence level classification arrangement groups individual sentences in a document. Be that as it may, each sentence can't be thought to be obstinate. For the most part, one must group the sentence as obstinate or not stubborn, which is called Subjectivity order. At that point the obstinate sentences are communicated as either positive or negative sentiments. In this we have three-arrangement issue, that is, to order a sentence as positive, negative or impartial [3].

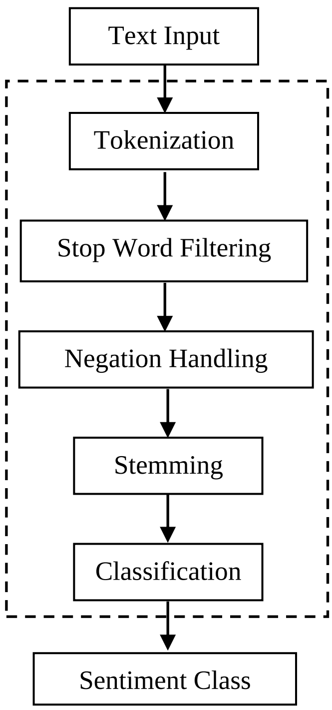

Fig. 1.Tasks in Sentiment Analysis

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Lakshmidevi N*, Research Scholar[1] and Assistant Professor[*1], Centurion University of Technology and Management[1], Paralakhemundi, India and GMR Institute of Technology[*1], Rajam, India. Email: lakshmidevi.n@gmrit.edu.in
Dr.M.Vamsikrishna, Professor, CSE, Chaitanya Institute of Science and Technology, Kakinada, India,Email: vkmangalampalli@gmail.com
Dr.S.S.Nayak, Professor, Centurion University of Technology and Management, Paralakhemundi, India. Email: ssnayak@cutm.ac.in

Over the two, the report and the sentence level conclusion examination this viewpoint level estimation investigation is finely grained. Its center assignment is to concentrate and condense individuals' sentiments on substances and perspectives/highlights of elements called targets. The angle level examination includes few stages which are perspective extraction, element extraction, and viewpoint assessment characterization [5].

At present, opinion investigation dependent on factual machine learning techniques has yielded great outcomes in different applications. In any case, the strategies are basic, which lead to their speculation capacity to deal with the complex classification and the capacity to express the unpredictable capacity to be confined to a limited degree under the state of constrained examples and computational units. The headway of internet technology and machine learning in data recovery make Sentiment Analysis ends up prevalent among specialists. Furthermore, the rising of long range informal communication and sites as a correspondence medium likewise adds to the advancement of research around there. Sentiment mining states to the application of Natural Language Processing, Computational Linguistics, and Text mining to identify and extract particular information in source materials. Machine Learning is commonly used to classify sentiment from text. With the resent advancement in deep learning models for sentiment analysis, the accuracy can be improved further more than traditional machine learning algorithms with recurrent neural networks (RNNs).

A. RNN Architecture

The basic architecture of the recurrent neural network has three layers. The three layers are Embedding, Dense, and LSTM layer.

Recurrent neural networks (RNN) are a kind of system which structures memory through recurrent connections. In feed forward systems, inputs are free of one another. Be that as it may, in RNN, all sources of info are associated with one another. This gives the system to display incredible transient conduct for a period a chance to arrangement which makes it positive for consecutive classification like opinion investigation. As it can be seen in the figure 2, at first,

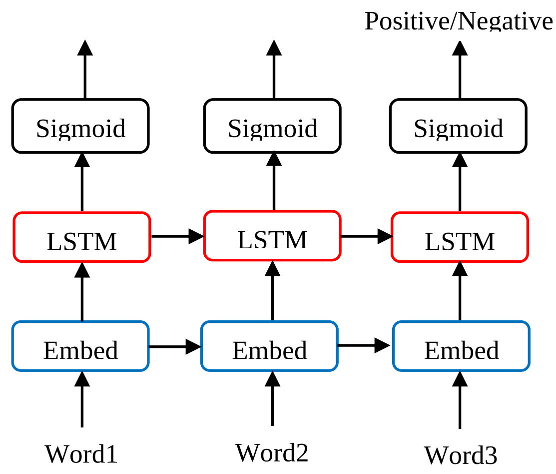

Fig. 2. Architecture of Recurrent Neural Network

It takes the x0 from the arrangement of info and after that it yields h0 which together with x1 is the contribution for the subsequent stage. Along these lines, the h0 and x1 is the contribution for the subsequent stage. Correspondingly, h1 from the following is the contribution with x2 for the subsequent stage, etc. Along these lines, it continues recollecting the unique situation while preparing.

$$h_t = f(W_{hx}X_t + W_{hh}h_t - 1 + b) \quad (1)$$

$$y_t = g(W_{yh}h_t + c) \quad (2)$$

Long Short-Term Memory (LSTM):

An adjustment of RNN with LSTM units, was proposed by the Sepp Hochreiter and Juergen Schmidhuber [ref]. A few errors back engender through normal RNN. These LSTM units are used to help in sidestep these errors. While keeping an increasingly steady errors, they let RNNs continue learning more than a few time steps. LSTMs comprise of data outside the essential flow of the RNN in a valve square. Neural system's nodes get activated by the weights they get. In like manner, LSTM's key pass on or obstruct the information dependent on its weight. From that point forward, these sign are ground with their own arrangements of loads. Thusly, RNN's learning procedure alter these loads that control concealed states and information. Eventually, these cells realize when to give data to get in, a chance to get out or be erased through the subsequent strides of speculating surmises, back proliferating blunder, and adjusting loads by means of inclination drop.

$$f_t = \sigma_g(W_f X_t + U_i h_t - 1 + b_f) \quad (3)$$

$$f_t = \sigma_g(W_f X_t + U_i h_t - 1 + b_i) \quad (4)$$

$$o_t = \sigma_g(W_o X_t + U_o h_t - 1 + b_o) \quad (5)$$

$$c_t = f_t o c_{t-1} + i_t o \sigma_o(W_o X_t + U_o h_t - 1 + b_o) \quad (6)$$

$$h_t = o_g o \sigma_h(c_t) \quad (7)$$

Dense Layer:

A dense network is only a normal layer of neurons in a neural network. Every neuron gets contribution from every one of the others neurons in the past layer as weights, in this way densely connected. The layer has a weight lattice W, a bias vector b, and the activations from the previous layers.

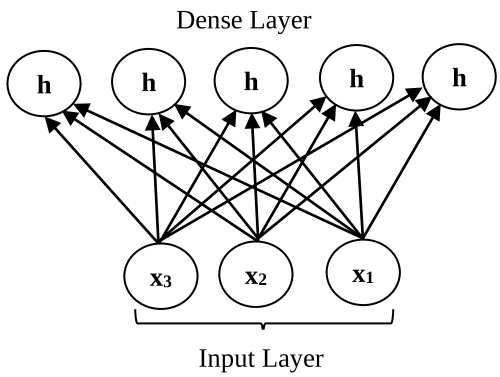

Fig. 3.Dense Layer

Embedding Layer:
The Embedding layer has weights that are found out. In the event that you spare your model to file, this will incorporate loads for the Embedding layer. The yield of the Embedding layer is a 2D vector with one installing for each word in the info succession of words. An implanting is a generally low-dimensional space into which you can decipher high-dimensional vectors. Inserting's make it simpler to do AI on huge data sources like inadequate vectors speaking to words.An embedding can be learned and reused across models.

In this work, we make use of the strong suit of deep learning model in aspect based sentiment analysis on product reviews. In Section 2, we have discussed the related work for sentiment analysis on onle product reviews. Section 3 dedicated to The Proposed model with pre processing, design and training the RNN model. Section 4 is for the results obtained and farther discussions followed by Conclusion.

II. RELATED WORK

The traditional write or online text sentiment analysis is based on sentiment dictionary. This paper combines the advantages of the SVM and CNN together and makes a text sentiment analysis where CNN is utilized for programmed feature and SVM is the text classifier. This model have effectively improved performance of the text analysis[7],[17]. A brief review on sentiment analysis. In this they have explained with different types of the sentiment analysis which are Document-level, Sentence –level, and Word-level sentiment analysis. They initially started discussing about the importance and what is sentiment analysis. They also have proposed different methods that are proposed by the researchers. Different methods are used in different researches process. Diverse natures of sentiment analysis model depends on word, sentence, feature and document is also discussed.This paper proposes the naive bayes method to sentiment analyse the text[8]. The traditional methods such as the SVM, random forest and other machine learning methods are applied to sentiment analysis which have poor classification capability in terms of compound classification problem. As RNN and CNN are two important models in the document and sentence level sentiment analysis. The pre trained model wordvect consists CNN to gain important local features of the text, then features are forwarded to two-layer LSTMs to extract context-dependent features. [9],[18]. The analysis such as the comparison of accuracy of the traditional naïve bayes model with the LSTM method. Initially this discus about the sentiment analysis and then dives into the discussing about the sentiment analysis with LSTM. They have mentioned step by step way to build a proper model using LSTM. They provided the source of the data, discussed about pre-processing the data, and about the modules used such as Tensor flow. Finally they have discussed about the analysis of LSTM that is the problem of long term dependencies, usage of LSTM in RNN, and the metrics to calculate the accuracy. Besides all these this paper also compares the LSTM model to the Naïve Bayes model and prove that the accuracy, F-Score, precision and the recall of the LSTM is greater than the Naïve Bayes[10].

The sentiment analysis, Then they have little bit introduced about RNN and LSTM. In this paper they have discussed about the Bidirectional LSTM. Backward to forward information flows is there in one directional LSTM. On the contrary in Bi-directional LSTM information not only flows backward to forward but also forward to backward using two hidden states. Hence Bi-LSTMs understand the context better. Finally they have compared the accuracy of the Bi-LSTM to the accuracies of the Decision Tree classifier and the SVM[11]. Isidoros Perikos et.al presented a method for aspect based sentiment analysis which depends on classifier gatherings. Idle Dirichlet Allocation is utilized to show subject and regular language preparing procedures are utilized to stipulate conditions on sentence level and oversee relationship among words and points [12]. Profound learning has accomplished commendable achievement in the field of ASC, which will bolster various application spaces [13].

Zulva Fachrina et.al. proposed a method which characterize examples and principles for each angle, and after that attempt a few calculations and highlights for order. For greatest viewpoints, order utilizing SVM with rule-based as one of the highlights gives the best outcome [14], [15].

III. PROPOSED METHOD

A. Data set:

It is possible to write a program to create by design a corpus of product reviews based up on two categories, “positive”, and “negative”. We took Amazon Reviews: Unlocked Mobile Phones dataset of 4,32,120 product reviews in English [6]. The sample raw data is shown in figure 4, It is composed of seven columns that are Name of the Product, Brad Name, cost, Rating, and sentiment. For this work, product review and the product name fields are considered. The value of the sentiment column is 0 if the product review is negative and 1 if the review is positive. The product name gives the information about the company of the product. The product is the review we are most concerned about as this is what we use to analyse the sentiment of the review.

B. Pre-processing

The corpus of reviews and all the resources from dataset to be pre-processed. It is a very important to do modification steps on raw data during pre-processing and this will show direct impact on classifier’s performance. The most important pre-processing is the removal of Unicode, cleaning, stemming, contraction removal, removal of special characters as they had no value that is they are neutral, removal of emotions.

- The typical text pre-processing steps are
- Remove all special characters
- Make all words lowercase
- Remove punctuation
- Tokenize: divide string into a list of substrings.
- Remove words not containing letters
- Remove words containing numbers
- Remove stop words: stop words are a list of high frequency words like, the, to, and also.

- Lemmatize: reduces the dimension of the data by aggregating words that either are the same root or have the same meaning.

The pre-processed data is in the form of text reviews, these to be converted into vectors by allocation of each review with a vector. To do this operation first, the text should be tokenized by fitting Tokenizer class on the data set. As shown in figure 5, the text is converted into lower case and should map these lists of words (tokens) to a list of unique integers for each unique word.

Padding is done on the list of integers to ensure that all reviews have the same length, which is a very important step for preparing data for RNN model. Applying this step would either shorten the review to required integers, or pad them with 0's in case they are shorter. Figure 6 shows reviews after padding.

	A	B	C	D	E	F	G	H
1	Product Name	Brand Name	Price	Rating	Reviews	Review Vote:	sentiment	
2	"CLEAR CLEAN ESN" Spr	Samsung	199.99	5	I feel so LUCKY to have found this used (	1	1	
3	"CLEAR CLEAN ESN" Spr	Samsung	199.99	4	nice phone, nice up grade from my pant	0	0	
4	"CLEAR CLEAN ESN" Spr	Samsung	199.99	5	Very pleased	0	0	
5	"CLEAR CLEAN ESN" Spr	Samsung	199.99	4	It works good but it goes slow sometime	0	1	
6	"CLEAR CLEAN ESN" Spr	Samsung	199.99	4	Great phone to replace my lost phone. T	0	0	
7	"CLEAR CLEAN ESN" Spr	Samsung	199.99	1	I already had a phone with problems... I	1	1	
8	"CLEAR CLEAN ESN" Spr	Samsung	199.99	2	The charging port was loose. I got that sc	0	0	
9	"CLEAR CLEAN ESN" Spr	Samsung	199.99	2	Phone looks good but wouldn't stay chai	0	1	
10	"CLEAR CLEAN ESN" Spr	Samsung	199.99	5	I originally was using the Samsung S2 Ga	0	0	
11	"CLEAR CLEAN ESN" Spr	Samsung	199.99	3	It's battery life is great. It's very respons	0	1	
12	"CLEAR CLEAN ESN" Spr	Samsung	199.99	3	My fiance had this phone previously, bu	0	0	
13	"CLEAR CLEAN ESN" Spr	Samsung	199.99	5	This is a great product it came after two	0	1	
14	"CLEAR CLEAN ESN" Spr	Samsung	199.99	5	These guys are the best! I had a little siti	2	1	
15	"CLEAR CLEAN ESN" Spr	Samsung	199.99	1	I'm really disappointed about my phone	1	1	
16	"CLEAR CLEAN ESN" Spr	Samsung	199.99	5	Ordered this phone as a replacement fo	1	0	
17	"CLEAR CLEAN ESN" Spr	Samsung	199.99	2	Had this phone before and loved it but v	0	0	
18	"CLEAR CLEAN ESN" Spr	Samsung	199.99	5	I was able to get the phone I previously	6	1	
19	"CLEAR CLEAN ESN" Spr	Samsung	199.99	5	I brought this phone as a replacement fc	0	0	
20	"CLEAR CLEAN ESN" Spr	Samsung	199.99	4	I love the phone. It does everything I ne	1	1	
21	"CLEAR CLEAN ESN" Spr	Samsung	199.99	3	unfortunately Sprint could not activate t	0	0	
22	"CLEAR CLEAN ESN" Spr	Samsung	199.99	4	The battery was old & had been over usi	0	1	
23	"CLEAR CLEAN ESN" Spr	Samsung	199.99	4	pros-beautiful screen,capable of runnin	0	0	

Amazon_Unlocked_Mobile

Fig. 4.Raw Data

```
Original revies:

['           excellent service, great price, and amazing friendly staff both in person and through the customer service hotlin
e (in prague), even though i couldn't speak czech!        '
 '           Comfortable seating with adequate space. Clean and bright.Punctual. Organized storage.I like the electrical and U
SB sockets.There was no wifi however; Maybe there were but I did not know. Better have signs indicating the wifi availability.I
will definitely look for this company in my next travels.        ']

--------------------

Reviews after tokenization and sequencing:

[[197, 30, 103, 188, 3, 671, 172, 146, 314, 8, 323, 3, 178, 1, 57, 30, 837, 8, 326, 62, 261, 4, 954, 222, 1274], [85, 1158, 16,
1275, 518, 105, 3, 2097, 195, 1774, 1159, 4, 108, 1, 2098, 3, 1070, 788, 39, 6, 26, 130, 271, 519, 39, 33, 28, 4, 90, 21, 127,
276, 25, 1392, 1775, 1, 130, 2099, 4, 49, 295, 431, 10, 22, 42, 8, 14, 137, 1276]]
```

Fig. 5.Original reviews vs after tokenization and sequencing

All the reviews in the dataset are similarly converted into vectors to find the polarity. As the model cannot understand text directly, converted the text reviews into vectors for feeding into the RNN Network. The first layer in the model is the Embedding Layer followed by LSTM and output layer. Mapping of each review word to the embedding vector space and it must fit for all the reviews.

```
Reviews after tokenization and sequencing:

[[197, 30, 103, 188, 3, 671, 172, 146, 314, 8, 323, 3, 178, 1, 57, 30, 837, 8, 326, 62, 261, 4, 954, 222, 1274], [85, 1158, 16,
1275, 518, 105, 3, 2097, 195, 1774, 1159, 4, 108, 1, 2098, 3, 1070, 788, 39, 6, 26, 130, 271, 519, 39, 33, 28, 4, 90, 21, 127,
276, 25, 1392, 1775, 1, 130, 2099, 4, 49, 295, 431, 10, 22, 42, 8, 14, 137, 1276]]
--------------------

Reviews after padding:

[[ 197   30  103  188    3  671  172  146  314    8  323    3  178    1
    57   30  837    8  326   62  261    4  954  222 1274    0    0    0
     0    0    0    0    0    0    0    0    0    0    0    0    0    0
     0    0    0    0    0    0    0    0    0    0    0    0    0    0
     0    0    0    0    0    0    0    0    0    0    0    0    0    0
     0    0    0    0    0    0    0    0    0    0    0    0    0    0
     0    0    0    0    0    0    0    0    0    0    0    0    0    0
     0    0]
 [  85 1158   16 1275  518  105    3 2097  195 1774 1159    4  108    1
  2098    3 1070  788   39    6   26  130  271  519   39   33   28    4
    90   21  127  276   25 1392 1775    1  130 2099    4   49  295  431
    10   22   42    8   14  137 1276    0    0    0    0    0    0    0
     0    0    0    0    0    0    0    0    0    0    0    0    0    0
     0    0    0    0    0    0    0    0    0    0    0    0    0    0
     0    0    0    0    0    0    0    0    0    0    0    0    0    0
     0    0]]
```

Fig. 6.Reviews after padding

Retrieval Number: A1910109119/2019©BEIESP
DOI: 10.35940/ijeat.A1910.109119
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1, October, 2019

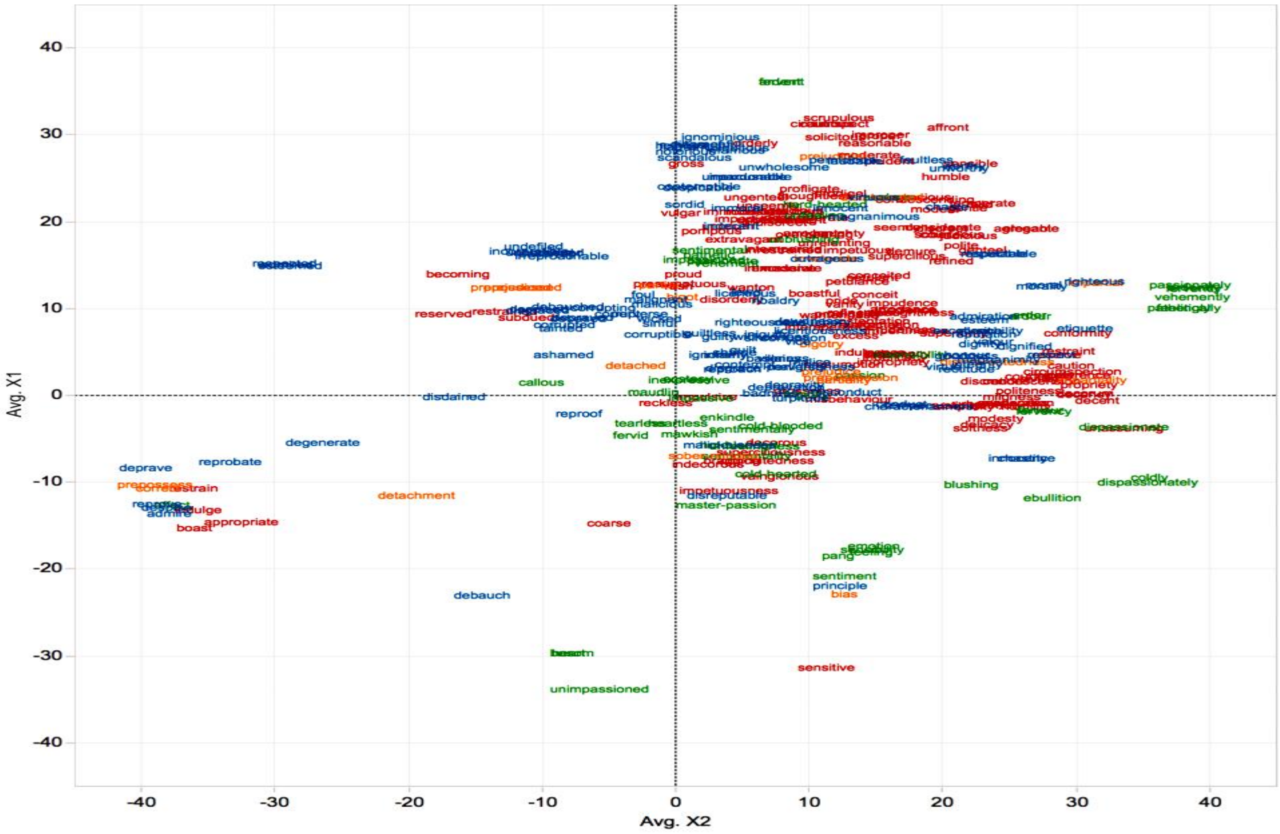

Fig. 7.WordVect mapping of words

So, a word inserting is a class of methodologies for speaking to words and records utilizing a thick vector portrayal. It is an improvement over more the conventional pack of-word model encoding plans where huge inadequate vectors were utilized to speak to each word or to score each word inside a vector to speak to a whole jargon.

The dataset is spitted into training, validation and testing datasets. Initially the dataset is spitted into training and testing sets, furthermore part of the training set is spitted as validation set. Keras framework is used to define and train the model.

Layer (type)	Output Shape	Parameters
embedding_1 (Embedding)	(None, 1504, 128)	256000
spatial_dropout1d_1	(Spatial (None, 1504, 128)	0
lstm_1 (LSTM)	(None, 196)	254800
dense_1 (Dense)	(None, 2)	394
Total parameters: 511,194, Trainable parameters: 511,194, Non-trainable params: 0		

Fig. 8.Layers of RNN

As the embedding layer requires the size of the vocabulary and the length of the sequences, we set the size of the vocabulary as the number of words in tokenizer dictionary + 1 and the input length to 100, where the value of the padding must be the same. Embedding size parameters tells us the dimensions that should be used to represent each word.

```
Train on 894 samples, validate on 64 samples
Epoch 1/10
894/894 [==============================] - 3s 4ms/step - loss: 0.6309 - acc: 0.6264 - val_loss: 0.5542 - val_acc: 0.7812
Epoch 2/10
894/894 [==============================] - 2s 3ms/step - loss: 0.4602 - acc: 0.8177 - val_loss: 0.4140 - val_acc: 0.8438
Epoch 3/10
894/894 [==============================] - 2s 3ms/step - loss: 0.3744 - acc: 0.8289 - val_loss: 0.3000 - val_acc: 0.9062
Epoch 4/10
894/894 [==============================] - 2s 3ms/step - loss: 0.3332 - acc: 0.8456 - val_loss: 0.2963 - val_acc: 0.8750
Epoch 5/10
894/894 [==============================] - 2s 3ms/step - loss: 0.3030 - acc: 0.8714 - val_loss: 0.3110 - val_acc: 0.9062
Epoch 6/10
894/894 [==============================] - 2s 3ms/step - loss: 0.2405 - acc: 0.8926 - val_loss: 0.2965 - val_acc: 0.9375
Epoch 7/10
894/894 [==============================] - 2s 3ms/step - loss: 0.1871 - acc: 0.9620 - val_loss: 0.1190 - val_acc: 0.9688
Epoch 8/10
894/894 [==============================] - 2s 3ms/step - loss: 0.0925 - acc: 0.9810 - val_loss: 0.0862 - val_acc: 0.9844
Epoch 9/10
894/894 [==============================] - 2s 3ms/step - loss: 0.0381 - acc: 0.9899 - val_loss: 0.0382 - val_acc: 0.9844
Epoch 10/10
894/894 [==============================] - 2s 3ms/step - loss: 0.0095 - acc: 0.9978 - val_loss: 0.0957 - val_acc: 0.9688

<keras.callbacks.History at 0x21c820ea208>
```

Fig. 9.Training the model

We have LSTM layer with 200 memory cells. Adding more layers and cells can lead to more accuracy and gives good result. Finally the output layer with the sigmoid activation functions to predict the probability of review being positive. Now the model is created and set to train. After training the model for 10 epochs, this model achieves an accuracy of 98.44% on validation set and 95.93% on test set.

Accuracy is determined as the quantity of every right forecast separated by the all out number of the dataset. The best exactness is 1.0, while the most exceedingly awful is 0.0.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

After the training the model, it is evaluated with test set. The confusion matrix is generated for test set with actual labels and predicted labels.

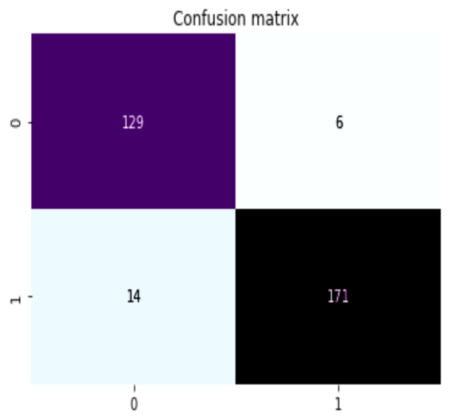

Fig. 10. Confusion matrix

IV. RESULTS AND DISCUSSIONS

In the following section presents the performance analysis of proposed model with traditional machine learning models like SVM, Naïve Bayes and Random Forest models. For the ease of operation GUI is designed for the proposed model.

A. Results and comparisons

The performance of the proposed model is compared based on accuracy obtained as performance metric. For the Support Vector Machine Classification model, the values gamma, C are tuned and comparison of various kernels is done. After tuning, the Poly Kernel seems to be the best fit and the degree used is 3. The above parameters on the SVM Classification model gave an accuracy of 85.54% on the Training dataset, 85.53% on the testing dataset. For the Naïve Bayes classification model, the priors are not needed to train the learning model. The Laplacian smoothing of 1.0 is applied to the Training dataset to estimate the value. The Multinomial Naïve Bayes gave a better training accuracy of 87.6% and a testing accuracy of 82.79%.

The Random Forest classifier gave training accuracy of 92.34% and testing accuracy of 91.2%.The training accuracy of the RNN model that 97.6% and testing accuracy 95.6%. Below figure shows the testing of model with sample data.

text : " i feel so lucky to have found this used phone to us not used hard at all phone on line from someone who upgraded and sold this one my son liked his old one that finally fell apart after years and didnt want an upgrade thank you seller we really appreciate it your honesty re said used phonei recommend this seller very highly would but from them again"
pos_acc 98.34 %
neg_acc 78.16 %
Result: positive

Fig. 11. Testing the model with sample data

As by the above plot we came to know that Recurrent Neural Network gives good accuracy over all the traditional machine learning algorithms such as Random Forests, SVM and Naïve Bayes.

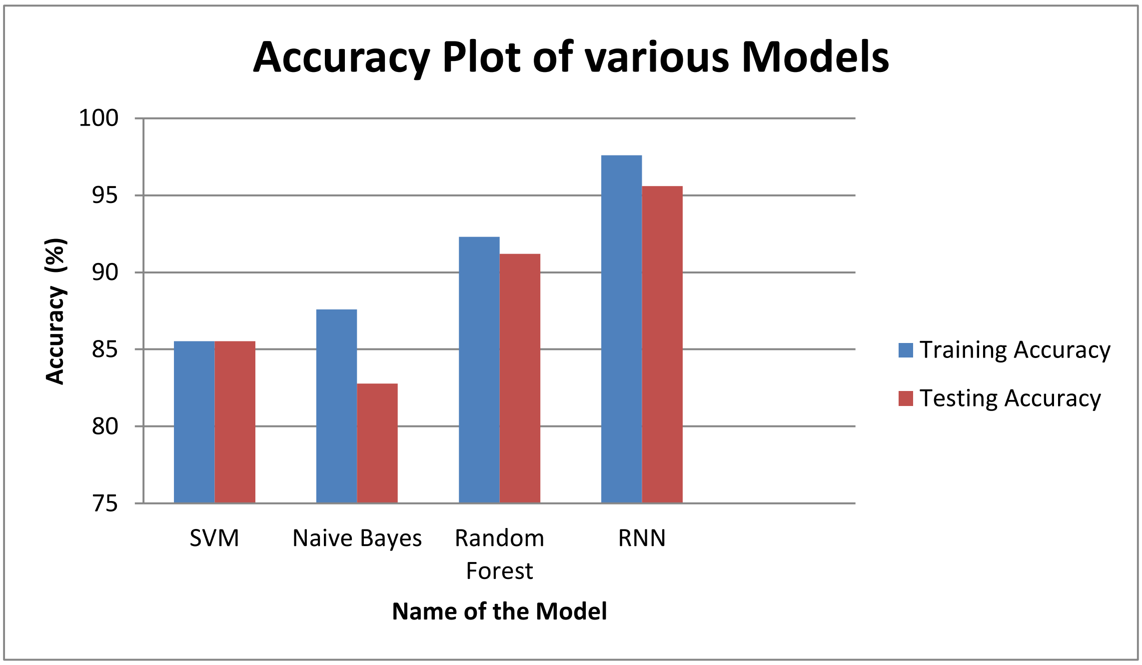

Fig. 12. plot for accuracy comparison of each classification models

B. Graphical User Interface Design

The GUI part of this application is completely designed using Visual Studio IDE in python. The user interface are included with all the basic requirements of the client such as the visualization of the data, number of words classified under different classes like positive, negative and neutral.

C. Output Screens

This is the main activity that takes the input's which are the product name and review. After providing inputs, click on "Analyse" button. This application uses the RNN model as this gives more accuracy over others.

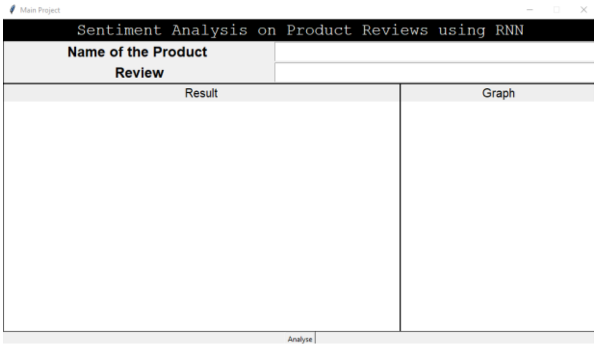

Fig. 13. Main Layout of the Application

After clicking "analyse" the review and the product name is sent to the RNN model to analyse the review. The result can be viewed in the result window shown in the UI Layout. This Result window contains all the important data such as the review, Tokenized text, vectorised text, Polarity and the overall sentiment.

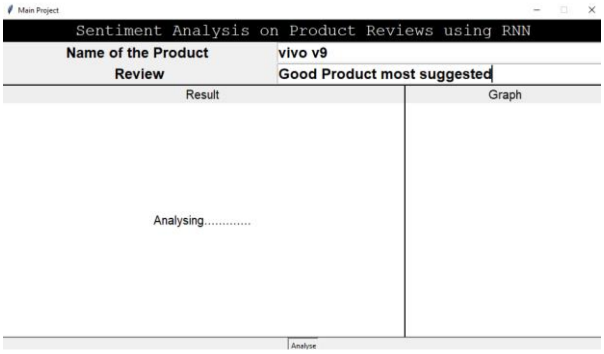

Fig. 14. Analyse Window

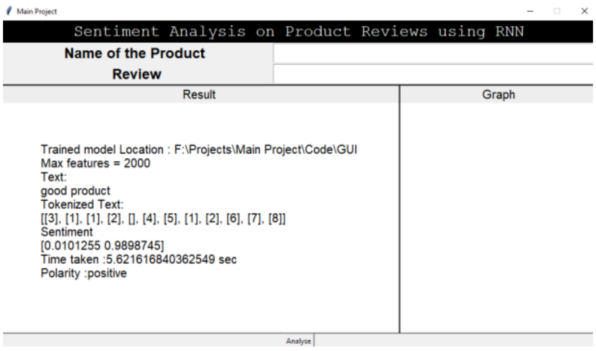

Fig. 15. Result of the review

VI. CONCLUSION

Sentiment analysis on text data is performed using traditional machine learning models and RNN model. The machine learning models are Naïve Bayes, random forest and SVM are used to classify the Amazon Reviews: Unlocked Mobile Phones dataset. The accuracies in classifications are 85.53%, 87.6%, 91.2% respectively.These methods are successful at predicting sentiment on topics in mobile reviews on a small scale using three different approaches Naïve Bayes, Random Forest, Support Vector Machine and also gained a lot of information in machine learning. Finally we have applied Recurrent Neural Network on the same mobile dataset and have had good result with accuracy of 95.6%. This work compares the accuracies of all the four methods we built and prove that the Recurrent Neural Network model gives much more accuracy than the traditional machine learning algorithm with less loss. A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

REFERENCES

1. Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002.
2. B. Pang, et al., "Thumbs up?: sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, pp. 79-86.
3. S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in Proceedings of the 20th international conference on Computational Linguistics, 2004, p. 1367.
4. http://www.ebizmba.com/articles/social-networking-websites
5. T. Wilson, et al., "Recognizing contextual polarity in phrase-level sentiment analysis," in Proceedings of the conference on human language technology and empirical methods in natural language processing, 2005, pp. 347-354.
6. https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones
7. Yuling Chen and Zhi Zhang. "Research on text sentiment analysis based on CNNs and SVM", 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA).
8. Neha Raghuvanshi and Prof. J.M. Patil, "A Brief Review on Sentiment Analysis", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016
9. Qiongxia Huang, Xianghan Zheng, Riqing Chen, and Zhenxin Dong, "Deep Sentiment Representation Based on CNN and LSTM", in 2017 International Conference on Green Informatics.
10. Vikas Malik. And Amit Kumar, "Analysis of Twitter Data Using Deep Learning Approach: LSTM," in International Journal on Recent and Innovation Trends in Computing and Communication .
11. Abdullah Aziz Sharfuddin, Md Nafis Tihami and Md Saiful Islam, " A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification" in International Conference on Bangla Speech and Language Processing.
12. Perikos, Isidoros, and Ioannis Hatzilygeroudis. "Aspect based sentiment analysis in social media with classifier ensembles." In 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), pp. 273-278. IEEE, 2017.
13. Zhou, Jie, Jimmy Xiangji Huang, Qin Chen, Qinmin Vivian Hu, Tingting Wang, and Liang He. "Deep Learning for Aspect-Level Sentiment Classification: Survey, Vision and Challenges." IEEE Access (2019).
14. Fachrina, Zulva, and Dwi H. Widyantoro. "Aspect-sentiment classification in opinion mining using the combination of rule-based and machine learning." In 2017 International Conference on Data and Software Engineering (ICoDSE), pp. 1-6. IEEE, 2017.
15. Huma Parveen, and shikha pandey, "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm" in IEEE 2016
16. K. M. Azharul Hasan, Mir Shankar Sabuj, and Zakia Afrin, "Opinion mining using Naïve Bayes" in 2015 IEEE conference.
17. Nurulhuda Zainuddin and Ali Selamat, "Sentiment analysis using Support Vector Machine" in IEEE 2014
18. Xi Ouyang, Pan Zhou, Cheng Hua Li, and Lijun Liu, "Sentiment Analysis Using Convolutional Neural Network" in 2015 IEEE.

AUTHORS PROFILE

Lakshmidevi.N, completed Bachelor of Technology from JNTU, Kakinada in 2009 and completed Master of Technology with Computer Science Engineering specialization from JNTU, Kakinada in 2013. She is working as an Assistant Professor in the department of CSE, GMRIT-Rajam, India and pursuing PhD(Part-time) in Sentiment Analysis from Centurion University, Parlakhemundi, India.

Dr. M. Vamsi Krishna, is currently working as Professor of CSE, Principal, Chaitanya Institute of Science & Technology, Kakinada. His area of research includes Medical Image processing, Artificial Intelligence, Data Mining, Machine Learning, Computer Networks. He has a total of 20 years of experience in teaching and has published various papers in national and international journals. He pursued his Ph.D from Centurion University, M.Tech (AI&R) from Andhra University, M.Tech(CS) from Allahabad University, MCA from IGNOU.

Dr. S.S.Nayak, is currently working as Professor, Centurion University, parlakhemundi, Odissa, India. He has more than 20 years of research experience in DSP and Computer science engineering field. Currently he is guiding more than 10 PhD students. He held so many administrative positions like Dean (R&D), Principal, etc..