# Comparative evaluation of deep learning and machine learning techniques for sentiment analysis of electronic product review data

*Archana* Nagelli[1,2], *B*. Saleena[1,*]and *B*.Prakash[1]

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India.
[2] Department of CSE, SreeNidhi Institute of Science & Technology, Hyderabad, India.

**Abstract:** The primary thoughts, perceptions, attitudes, feedback, and even emotions expressed by people on social networking and e-commerce sites are the primary focus of sentiment analysis also referred to as opinion mining. It provides meaningful information to various stakeholders and customers in influencing their next move. However, the biggest challenge is the extraction of relevant information from the tremendous data. Machine learning and deep learning techniques have obtained remarkable success in exemplifying and classifying information. Machine learning works with the binary classification of information, whereas deep learning provides automatic feature detection. A study was carried out to extract the relevant information from the Amazon reviews dataset of electronics products. The Naïve Bayes, support vector machine, decision tree, convolution neural network, long short term memory, recursive neural networks, and recurrent neural networks were used on the dataset after applying different data preprocessing. To evaluate the performance of various machine learning and deep learning techniques, frameworks, F1 score, precision, recall as well as, accuracy was used. The results suggest that deep learning techniques have outperformed the machine learning techniques, and RNN shows the highest accuracy among all the techniques.

**Keywords:** Sentiment analysis, Deep Learning, Machine learning, NB, SVM, LSTM, RNN, CNN, RecNN

## 1. Introduction

In the modern world, individuals use social media to exchange ideas, knowledge, and experiences; they transform their way of life in accordance with online interfaces; and they are influenced by the opinions and reviews of any updates, products, or services. Such

*Corresponding Author: saleena.b@vit.ac.in

sentiments or opinions are classified into positive, negative or neutral or even fine-grained attributes [1,2]. The identifying, categorising and analysing of perceptions of several entities like objects, amenities, personalities, companies, organisations, events and topics received from multidisciplinary field are called as Sentiment analysis. 'Opinion Mining' is another term used for sentimental analysis based on its role in inspecting sentiments towards an object [3]. Two terms named 'Extremity' and 'Subjectivity' can be investigated as sentiment analysis parts. Subjectivity alludes to the person's convictions, perspectives, or individual suppositions, while extremity essentially alludes to the suppositions communicated in the wording of positive, negative, or nonpartisan. Supposition examination covers the extent of dealing with sentence level, archive level, and sub-sentence level.

Essentially, there are three major classification type sentiment analysis tasks, which is Document-level classification, Sentence-level classification and aspect-level classification and related sub-tasks classification such as multi-domain or multimodal sentiments [4,5]. In general, if the number of lexicons in the text is greater, and those lexicons are from the positive list, then it will be labeled positive. Thus, before the classification of sentiment, the feelings' lexicon dictionary has to be made. Some of the scholars used SentiWords, MPQA lexicons, and SentiWordNet to analyze sentiment in their research works refer [6,7,8]. Digital models use the concept of labelled documents to get a computer to learn out an outcome. If a condition label (positive or negative) or a rank or any other regular attribute may be labelled, then it is essential to provide that, the performance of the model depends on the algorithm one uses to label the data, the accuracy of the label and however many of them that is used in the sentiment analysis [9]. [10] Liu et al. proposed a strategy to sort feelings utilizing an enormous word reference of normal sense information and semantic models. Sentiment classification can be performed at the word, angle, express, idea, sentence, articulation, and archive level [11]. The investigation is principally founded on archive level opinion characterization [12].

SVM and Naïve Bayes are two commonly and effectively incorporated classification techniques when handling sentiment analysis [13,14]. In recent years, sentiment classification using deep-learning algorithms also pass greatly beyond the mere lexicon-based or other machine-based models of sentiment classification using specific algorithms like CNNs or RNNs [15,16]. Though, the deep learning classification models are very efficient and the above architecture is perfect for variable datasets and provinces respectively; so it is a difficult task for the practitioners to decide the right deep learning architecture for their datasets. Various forms of research on sentiment analysis on machine learning and deep learning algorithm have been conducted and are now trending. An advanced level of neural network design utilizes the creation of a convolution neural system (CNN) and recurrent neural system (RNN) and joints them together for the thoughtful examination of short messages. Specifically, the pooling procedure on adjoining words can hold the neighborhood highlights and their consecutive relations in a sentence. The CNN can perform the unbiased classification of the dataset with the application of the max-pooling layer

Additionally, RNN can gain proficiency with the long-haul conditions and the positional connection of highlights just as the entire sentence's worldwide highlights [17]. Various deep learning architectures have been suggested in several researches such as Auto-encoders; for CNNs, [18] and for RNNs [17,19]. CNN, LSTM and other such deeper learning methods are advantageous with the cut of false text cataloguing rate and are closer to correct [20,21]. The above mentioned methods was firstly implemented in the small text categorization [22,23] also in the process of movie feedback reviewing [14,22]. LSTM is a profound technique for evaluating natural language processing, as it can store long term memory and utilize it to assess variation in characteristics of different features. Blitzer et al.[24] applied the opinion-oriented analysis on the Amazon dataset containing appraisals/criticisms, specific for books, electronic products, and homecare products from the list of 25 products. They used auto-encoder (marginalized Stacked Denoising Auto-encoder (mSDA)) and decoder for classification, and observed that analysis generate the high accuracy of 87.69%. Zhang et al [25] applied the cell phone reviews labelled dataset containing an equal set of positive and negative reviews. The author classified the dataset based on feature extraction. At the same time, Bansal and Srivastava [26] collected a dataset of 40,000 reviews of various mobile phones. The author applied the data balancing techniques to select an equal number of positive and negative responses, applied data reduction techniques to remove the stop words. The author used NB, SVM, RF, and logistic regression and concluded the better performance of random forest with high accuracy than other techniques. Azarang and Kehtarnavaz [27] proposed a comparative evaluation of pioneering approaches used in speech desensitising. The first experiment was performed on the public domain IEEE dataset with four simple objective measures using the Short-Time Objective Intelligibility (STOI) [28] and the Perceptual Evaluation of Speech Quality (PESQE) [29], Segmented Signal-to-Noise Ratio (SSNR) [30] and log spectral distance. It is evident from the literature that the different techniques' performance depends upon the type of the dataset and features available in the dataset. Each technique has certain advantages that can be used while analyzing and obtain high accuracy. This study is carried out to identify the applicability of various ML and DL techniques for sentiment analysis performed on the Amazon dataset. The Amazon electronic product review dataset is selected for the study, and different techniques are compared on various evaluation measures for the identified dataset.

## 2. Methodology

The methodology adopted to classify the electronic product reviews is presented in this section to perform the sentimental analysis, as shown in figure 1. The Amazon product review dataset is chosen for this study. The dataset contains the reviews from 1996 to 2018. The total dataset includes 233 million reviews; the electronic products selected for sentiment analysis contain 20.99 million reviews [31]. The original data was available in JSON file format, converted to a CSV file for further processing.

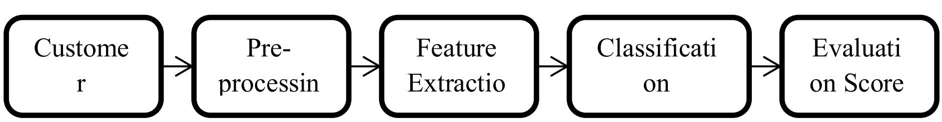


**Fig.1** Workflow adopted for the study

### 2.1. The Analysis of input data, called Pre-processing and feature extraction:

The data pre-processing was started with filtering out the data from data frames and storing the filtered data into another data frame. The dataset is divided to apply supervised machine learning algorithms based on vote value. The vote of fewer than three stars was considered negative sentiment, and above three was considered a positive response; however, three stars were regarded as neutral sentiment. Reviews were then examined on a language basis

to rectify the spelling errors, followed by the lowercase conversion of letters in words. The stemming and lemmatisation of the reviews are performed to get the exact form and meaning of the words. The stop-words (about, again, and, an many, just, etc.) were removed from the dataset to make it crisper. All the words are converted into number and counter for their multiple appearances, and other root level token changes were performed before splitting the dataset into specific sets. The dataset's fragmentation was then made into three sets as training (70%), validation (15%) and remaining 15% data for testing.
In different natural language processing operations, TF and TF-IDF are standard feature extraction techniques. They describe a vocabulary for a given dataset as a collection of unique terms, which ignore the semantic and syntactic similarities between these terms. TF-IDF is a weighting regimen to weigh the words that often occur in a given dataset by giving minimum weight. The more often an object will appear, the less its weight, and consequently the weight of the word, which is produced by it, will be. In contrast to several occurrences, the same word was used through the whole text, and the Inverse document frequency, IDF, is inversely proportional to the words. It indicates how a particular word is document particular holds importance.

The deep learning techniques are applied after performing pre-processing techniques on the dataset that includes the conversion of text into lowercase, symbols such as [/(){}\[\]\|@,;] are replaced with the space in the reviews. The symbols matched the suitable strings in [^0-9a-z #+_] from text and substituted the black space. The reviews are tokenised and padded in tokenised sequences before splitting the dataset for training and testing.

## 2.2. Machine Learning Approaches

The Machine Learning (ML) approach is appropriate to handle complex issues in wide applications for decision support, decision making, by utilizing pattern of data variation [32,33]. Artificial intelligence techniques train the classifier from truly stamped data. The ML approaches use the linguistic characteristics to enhance the efficiency of the model by using the classifier to learn the characteristics of dataset [34]. The general methodology to perform the machine learning approaches is shown in figure 2. The NB, SVM, and Decision Tree (DT) algorithms are used in this study to identify their applicability for the sentiment analysis. The NB algorithm performs data classification considering the Bayes Theorem for figuring and restrictive probabilities. In straightforward terms, an NB classifier accept that the closeness of a various features in a class, which is irrelevant to other component, to measure the regularity of components

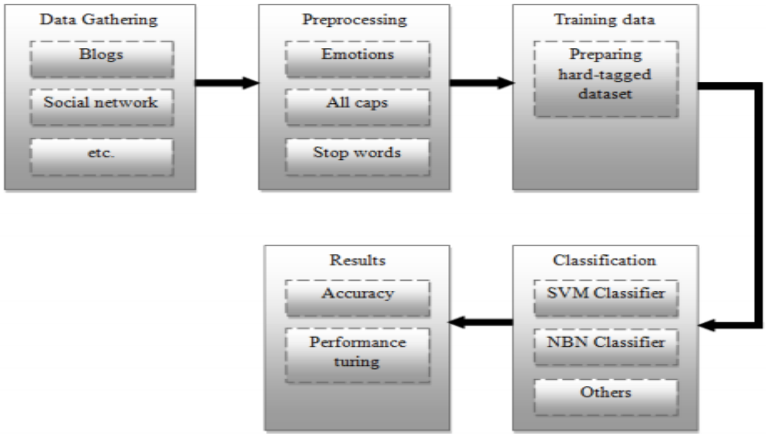


**Fig.2** General Structure for ML approaches

SVM is a popular AI technique utilized to arrange and classify multi-class data [35]. The objective of SVM is to develop an algorithm which can forecast the target assessment of dataset in the testing set, on the basis of characteristics of dataset. The particularity of the support vector machine technique is that the aim is to find a hyperplane in N-dimensional space (N - the number of features) that uncontestably defines the data centers. The decision tree (DT) algorithm has a spot with the gathering of directed learning counts. As opposed to other oversaw learning counts, the DT calculation can be used to deal with backsliding and gathering issues. Using such DT is to create the readiness model that is capable of predicting the class or assessment of the target variable by considering the indirect decision standards that were activated from before data (training data).

## 2.3. Deep Learning Approaches

The deep learning approaches allow learning of data through multiple layers to describe the characteristics of dataset by superimposing various layers. These strategies have improvised the forefront in talk affirmation, visual article affirmation, object distinguishing proof, and various spaces, for instance, cure divulgence and genomics. Deep learning demonstrates an incredible structure in immense educational lists through the back-spread count, depicting the machine's role in altering its inside limits to enroll the depiction in the individual layer from the depiction in the former layer. The deep learning convolutional nets have acknowledged forward jumps in planning pictures, video, talk, and sound; however, tedious ones have excelled light on progressive data, for instance, text and talk [36]. The deep learning approached employed in this research are Convolution Neural Network (CNN), Long Short-Term Memory (LSTM), Recursive Neural Networks (RecNN), and Recurrent Neural Networks (RNN).

CNN is an outstandingly picking up design animated by the trademark visual insight arrangement of the living creatures. Uses of CNN, including picture grouping, article or text identification and acknowledgement, object following, present assessment, visual saliency discovery, activity acknowledgement, scene naming, discourse, and standard language processing, have increased exponentially [37]. The three types of layers include convolutional, pooling and Complete associated. The Complete associated layers as illustrated in the CNN framework and related architecture in the figure 3 below are fully stacked [38].

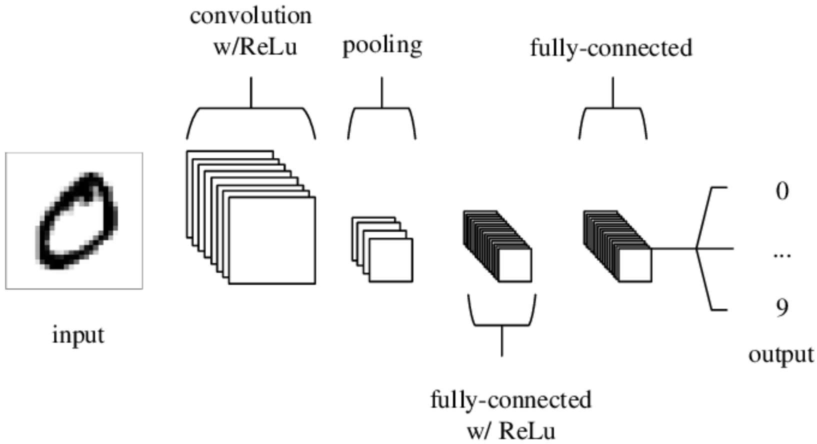


**Fig.3** A basic CNN architecture (O'Shea and Nash, 2015)

Mostly RNNs are combined with LSTM and gated recurrent unit (GRU) layers and their memory states. Some of the problems which relate to the sequential data include identifying a pattern of variation in data, predicting sales of this product, forecast on the speech pattern and behaviour. Since LSTM networks provide a solution to such problems, they tend to memorize and recall data variation for long time [39]. The RNNs technique work on the ground with two essential properties, a) distributed shrouded express that permits them to store a great deal of data about the past productively, and b) Non-direct elements that permit them to refresh their concealed state in muddled ways.

RecNN process organized information in various structures. For instance, common language sentences have various lengths and structures with equal importance. A model working on sentences ought to sum up to a wide range of their various structures. Boundary sharing permits recursive neural systems to sum up various information types not seen before preparing information [40]. RecNN are used to learn the deep structure of information by catching data from successions and time arrangement information.

## 2.4. Performance Evaluation Metrics

The model’s performance is then measured using the parameters which include; accuracy, precision, recall, and F1 score. The accuracy provides the information about the correct prediction of the count of samples. Accuracy is calculated as the ratio between true positive and all positive responses multiplied by 100. Recall defines the proportion of actually positive responses from the positive labelled responses.

Thereby for determining the balance of Precision, recall and F-measure F1-score is chosen from between them. Equations used in the process are as follows where TP, TN, FP, FN, TPP, and TAP stands for, true positive, true negative, false positive, false negative, true predictive positive, and true actual positive respectively.

$$\mathrm{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} = \frac{\mathrm{TP}}{\mathrm{TPP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{TAP}}$$

$$\text{F1 Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

.

## 3. Results and Discussion

The electronic product review dataset is categorised into three (positive, negative, and neutral) classes based on the star rating. The rating below the three stars is marked as negative, more than three stars are defined as positive, and three stars are referred to as the neutral response. The dataset contains more than 20 million reviews, and experiments were performed on the healthy unbalanced data. Table 1 describes the experimental results of a comparison between machine and deep learning techniques by applying pre-processed and feature extraction techniques on the Amazon electronic product review dataset for sentiment analysis. All the techniques are evaluated based on accuracy, precision, recall, and F1 score.

The entire data set is thereby split into training set, test set and validation set. After applying the pre-processing techniques such as filtering data from data frames, division of sentiment based on vote values, stemming, lemmatization and tokenization, the dataset is split for training and validation, 80% and 10%, and for testing, 20% of the dataset is used. Naïve Bayes, SVM, and Decision Tree are exploited for analysis as a machine learning technique. Whereas CNN, LSTM, RecNN, and RNN are the deep learning techniques used in the analysis, as shown in Table 1. The dataset is tokenized, and padded the tokenized sequences before applying to the deep learning algorithms. The embedding matrix is created for the tokenized sequences to estimate the fitting of deep learning models. From the comparative analysis, it was found that among the machine learning techniques, SVM outperforms the other algorithms and generates an accuracy of 87.07% and the least accuracy obtained with a Decision Tree of82.05%. The SVM uses the different kernels to classify the non-linear dataset whereas, the decision tree derives hyper-rectangles from the dataset to classify the dataset. The SVM also outperform the Naïve Bayes, as the features are mutually dependent on each other, and an extensive training dataset was used for the classification, whereas Naïve Bayes work more efficiently on a smaller dataset.

Nevertheless, RNN is performing satisfactory among other techniques regarding the performance evaluating parameters as accuracy 0.88%, precision 0.66, recall 0.64 and F1-score 0.65. After RNN technique, LSTM, CNN and SVM techniques are almost having the same performance on the Amazon reviews for electronics dataset, as shown in Table 1. The CNN uses the multiple convolution and pooling layers to convolute the input data through different kernels and generate different output. Whereas RNN and LSTM (variant of RNN), applies input in a sequence across time. The time-dependent input space captures the information more efficiently and produces high accuracy in training the dataset.

**Table 1** Comparison between Machine and Deep learning technique for sentiment analysis on amazon reviews for electronics products

| | **Algorithm** | **Accuracy (%)** | **Precision** | **Recall** | **F1-Score** |
|---|---|---|---|---|---|
| Machine Learning | Naïve Bayes | 84.47 | 0.565 | 0.340 | 0.319 |

| | | | | | |
|---|---|---|---|---|---|
| (ML) | SVM | 87.07 | 0.651 | 0.474 | 0.487 |
| | Decision Tree | 82.05 | 0.540 | 0.520 | 0.530 |
| Deep Learning (DL) | CNN | 87.81 | 0.650 | 0.630 | 0.640 |
| | LSTM | 87.20 | 0.580 | 0.560 | 0.540 |
| | RNN | 88.00 | 0.660 | 0.640 | 0.650 |
| | RecNN | 85.35 | 0.540 | 0.540 | 0.520 |

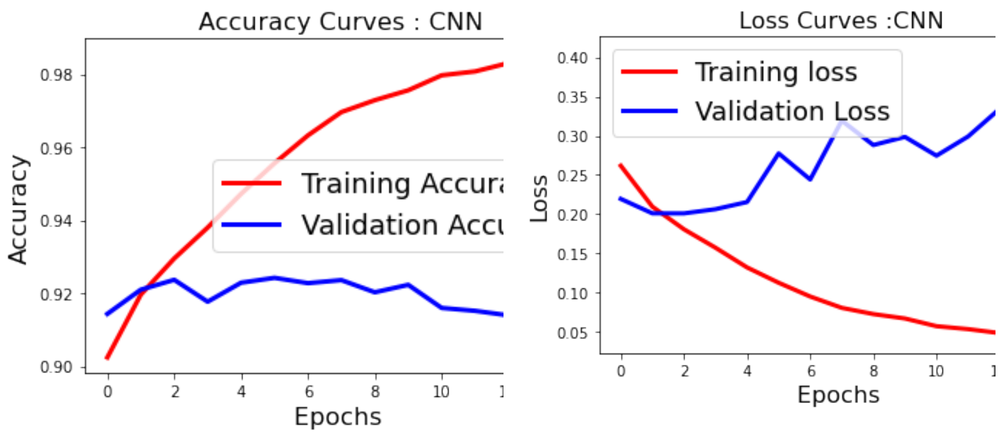


**Fig. 4.** Accuracy and Loss curve between Training and Validation for CNN

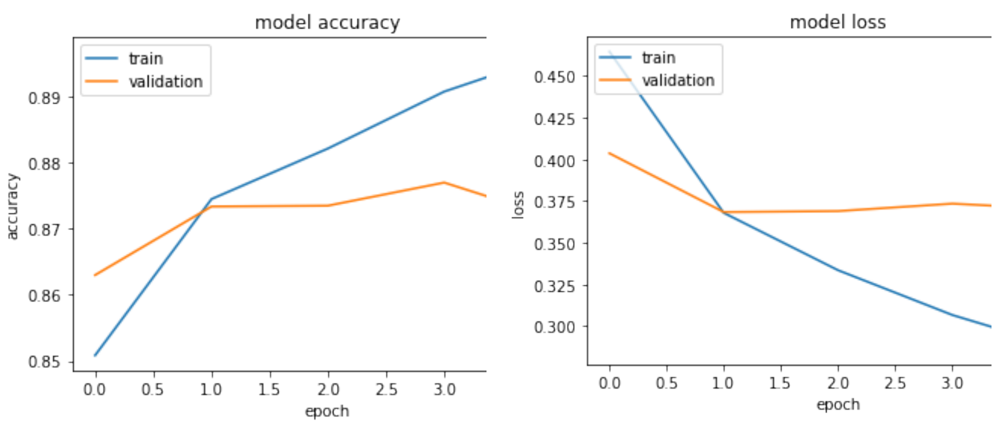


**Fig. 5.** Accuracy and Loss curve between Training and Validation for LSTM.

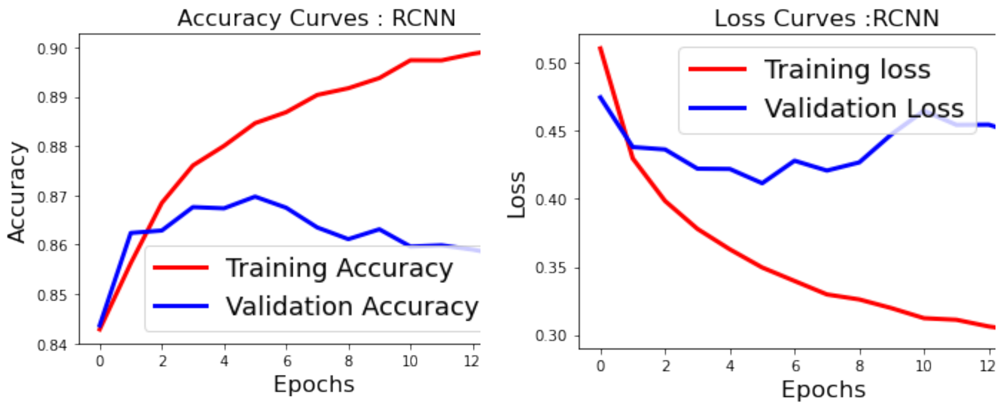


**Fig. 6.**Accuracy and Loss curve between Training and Validation for RCNN.

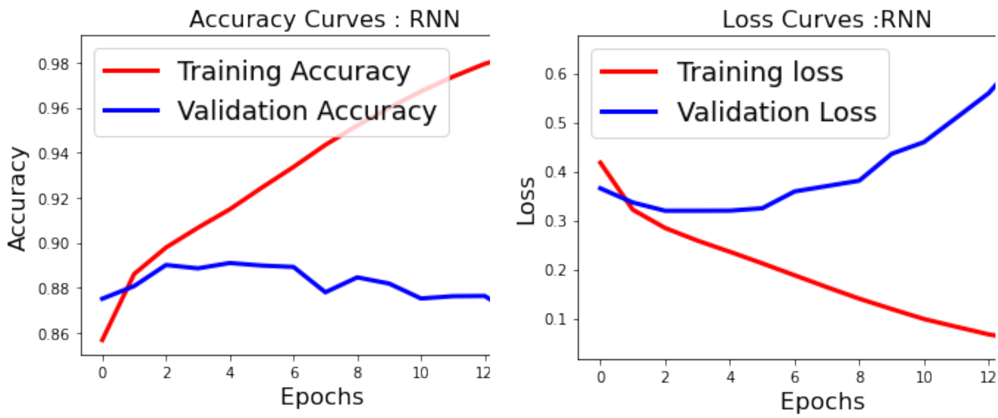


**Fig. 7.** Accuracy and Loss curve between Training and Validation for RNN.

Figure 4 depicts the training vs validation accuracy and loss with Epochs for the CNN technique. Similarly, Figures 5, 6, and 7 illustrate the training vs validation accuracy and loss with epochs for LSTM, RecNN, and RNN technique, respectively. The training accuracy of all the technique increases gradually. The CNN RecNN and RNN give the highest accuracy at 14 epochs, whereas LSTM gives the highest accuracy at epoch four. A ROC bend that depicts a recipient working trademark bend of a paired classifier framework is a graphical plot that depicts the analytic sector capacity measure of the paired classifier framework as a function of the segregation limit. The analysis of ROC curve presented below reveals that the true positive of the model predicted by SVM is considerably higher than Naïve Bayes and Decision Tree. Figure 8 represents the SVM, NB, and decision tree's ROC curve to validate the obtained results.

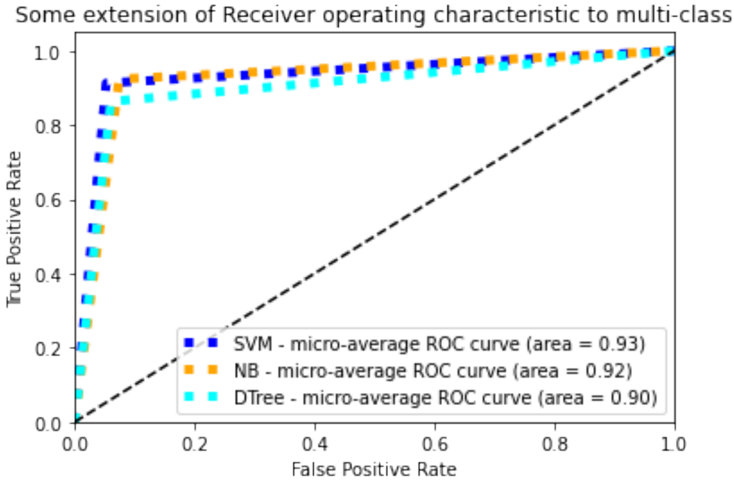


**Fig.8**. True positive Rate vs False Positive Rate

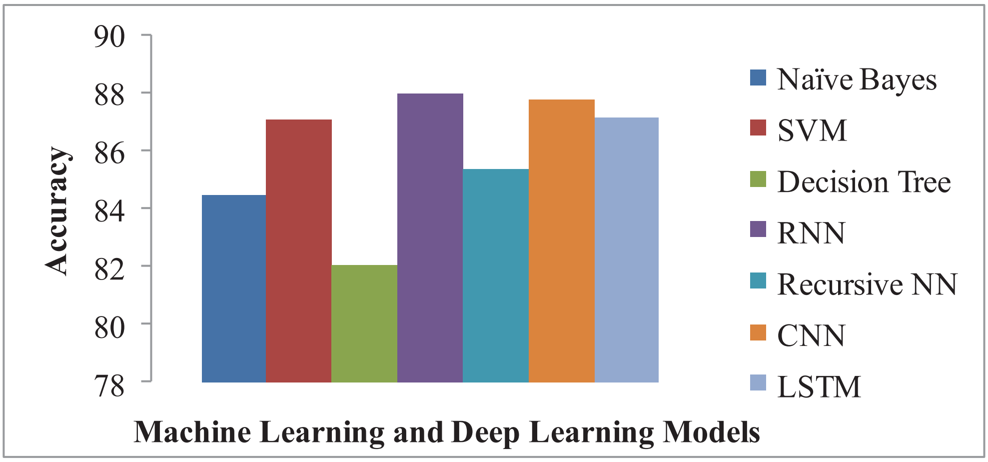


**Fig.9.** Comparative Analysis based on Accuracy of various Machine Learning and Deep Learning Models

From the Fig 9, it is evident that the accuracy of RNN is the best and the accuracy of the decision tree is the lowest. Since RNN and CNN show very high accuracy but out of these two techniques, RNN outperforms CNN because of its ability to recognize pattern across time which CNN performs spatial recognition. It is observed from the analysis that both RNN and CNN overtake the other learning approaches based on their application of significant features extraction in comparison to the traditional handcrafted features. However, in basic CNN architecture, the building blocks used are filters/kernels, which help extract the apt features by performing the convolution operation. RNN architecture has a recurrent connection, whereas the feedback loops presences confirm the capture of the sequential information from the input text data, which makes both a suitable choice for the used text dataset.

## 4. Conclusion

Sentiment analysis is a noticeable field assessed through machine and deep learning techniques applied to examining, distinguishing sentimentalities, assertiveness, sensations and views about several entities like public opinions on various products, services, political opinion, performance of companies, new product launched by companies etc. Different machine and deep learning techniques are utilized to foresee the estimation that is typically considered as the first guidance and decision-making tool for the potential buyers. This paper evaluated the of Amazon reviews for electronics based on various parameters like accuracy, precision, recall, F1 and scores using various machine and deep learning models. The SVM, NB and DT were used among machine learning, and CNN, RNN, RecNN and LSTM were used among deep learning techniques. With the RNN algorithm, we get higher accuracy values for Amazon reviews than the other algorithms, followed by CNN and LSTM. Whereas in Naïve Bayes, the Recall and F1 Score of Amazon reviews values are more diminutive. In the future, it is possible to compare the various dataset using the integration of machine learning and deep learning to enhance the model performance as well as its resistance to errors.

## References

1. Prabowo, R., &Thelwall, M. Sentiment analysis: A combined approach. Journal of Informetrics, 3(2), 143-157 (2009).
2. Medhat, W., Hassan, A., &Korashy, H. Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113 (2014).
3. Ullah, M. A., Marium, S. M., Begum, S. A., &Dipa, N. S. An algorithm and method for sentiment analysis using the text and emoticon. ICT Express (2020).
4. Mejova, Y. Sentiment analysis: An overview. University of Iowa, Computer Science Department (2009).
5. Gatti, L., Guerini, M., &Turchi, M.SentiWords: Deriving a high precision and high coverage lexicon for sentiment analysis. IEEE Transactions on Affective Computing, 7(4), 409-421. (2015).
6. Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., &Pantic, M. A survey of multimodal sentiment analysis. Image and Vision Computing, 65, 3-14 (2017).
7. Khoo, C. S., &Johnkhan, S. B. Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. Journal of Information Science, 44(4), 491-511 (2018).
8. Baccianella, S., Esuli, A., &Sebastiani, F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In LREC (Vol. 10, No. 2010, pp. 2200-2204) (2010).
9. Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. Sentiment classification of Internet restaurant reviews written in Cantonese. Expert Systems with Applications, 38(6), 7674-7682 (2011).
10. Liu, H., Lieberman, H., &Selker, T. A model of textual affect sensing using real-world knowledge. In Proceedings of the 8th international conference on Intelligent user interfaces (pp. 125-132) (2003).
11. Ravi, K., & Ravi, V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46(2015).
12. Ghosh, R., Ravi, K., & Ravi, V. A novel deep learning architecture for sentiment classification. In 2016 3rd International Conference on Recent Advances in Information Technology (RAIT) (pp. 511-516). IEEE . (2016)

13. Abbasi, A., Chen, H., Thoms, S., & Fu, T. Affect analysis of web forums and blogs using correlation ensembles. IEEE transactions on knowledge and data engineering, 20(9), 1168-1180 (2008).
14. Pang, B., Lee, L., &Vaithyanathan, S.Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070 (2002).
15. Heikal, M., Torki, M., & El-Makky, N. Sentiment analysis of Arabic Tweets using deep learning. Procedia Computer Science, 142, 114-122 (2018).
16. Chakraborty, K., Bhattacharyya, S., Bag, R., &Hassanien, A. E. Comparative sentiment analysis on a set of movie reviews using deep learning approach. In International Conference on Advanced Machine Learning Technologies and Applications (pp. 311-318). Springer, Cham (2018).
17. Wang, J., Fu, J., Xu, Y., & Mei, T. Beyond Object Recognition: Visual Sentiment Analysis with Deep Coupled Adjective and Noun Neural Networks. In IJCAI (pp. 3484-3490) (2016).
18. Liao, S., Wang, J., Yu, R., Sato, K., & Cheng, Z. CNN for situations understanding based on sentiment analysis of twitter data. Procedia computer science, 111, 376-381(2017).
19. Rong, W., Nie, Y., Ouyang, Y., Peng, B., &Xiong, Z. Auto-encoder based bagging architecture for sentiment analysis. Journal of Visual Languages & Computing, 25(6), 840-849 (2014).
20. Bhargava, R., Sharma, Y., & Sharma, S. Sentiment analysis for mixed script indic sentences. In 2016 International conference on advances in computing, communications and informatics (ICACCI) (pp. 524-529). IEEE (2016)
21. Kilimci, Z.H., & Akyokus, S. Deep learning-and word embedding-based heterogeneous classifier ensembles for text classification. Complexity. (2018).
22. Mandelbaum, A., & Shalev, A. Word embeddings and their use in sentence classification tasks. arXiv preprint arXiv:1610.08229 (2016).
23. Patil, S., Gune, A., & Nene, M. Convolutional neural networks for text categorisation with latent semantic analysis. International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS). 10.1109/ICECDS.2017.8390217, IEEE (2017).
24. Blitzer, J., Dredze, M., & Pereira, F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In Proceedings of the 45th annual meeting of the association of computational linguistics (pp. 440-447) (2007).
25. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253 (2018).
26. Bansal, B., & Srivastava, S. (2018). Sentiment classification of online consumer reviews using word vector representations. Procedia computer science, 132, 1147-1153.
27. Azarang, A., & Kehtarnavaz, N. A Review of Multi-Objective Deep Learning Speech Denoising Methods. Speech Communication (2020).
28. Jensen, J., & Taal, C. H. An algorithm for predicting the intelligibility of speech masked by modulated noise maskers. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(11), 2009-2022. (2016).
29. Santos, J. F., & Falk, T. H. Speech dereverberation with context-aware recurrent neural networks. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(7), 1236-1246 (2018).
30. Hershey, J. R., Chen, Z., Le Roux, J., & Watanabe, S. Deep clustering: Discriminative embeddings for segmentation and separation. In *2016 IEEE*

*International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 31-35). IEEE (2016).
31. Amazon, (2018). Amazon electronic product review dataset. https://nijianmo.github.io/amazon/index.html
32. Sebastiani, F. Machine learning in automated text categorisation. ACM computing surveys (CSUR), 34(1), 1-47(2002).
33. Veale, M., &Binns, R. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, *4*(2), 2053951717743530 (2017).
34. Sadia, A., Khan, F., & Bashir, F. An Overview of Lexicon-Based Approach For Sentiment Analysis (2018).
35. Cristianini, N., &Shawe-Taylor, J. An introduction to support vector machines and other kernel-based learning methods. Cambridge University press (2000).
36. Collaboration, S. I. Machine Learning and Health Care Disparities in Dermatology (2018).
37. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., & Chen, T. Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377 (2018).
38. O'Shea, K., & Nash, R. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458 (2015).
39. Hochreiter, S., & Schmidhuber, J. Long short-term memory. Neural computation, 9(8), 1735-1780 (1997).
40. Setiono, R., Baesens, B., &Mues, C. Recursive neural network rule extraction for data with mixed attributes. IEEE Transactions on Neural Networks, 19(2), 299-307(2008).