Sentiment Analysis with Deep Learning RNN–LSTM framework for Unstructured Text

Dr. Kusum Mehta[1], Juveria[2], Dr. Veena Mittal[3]
[1,2]Department of CSE, SEST, Jamia Hamdard University
[3]Department of CSE, GCET, Greater Noida
E-mail:kusumprerak@gmail.com,juveriakhan88@gmail.com, veena.mittal06@gmail.com

Abstract - In recent times,immense information is increasing on social media data.. Sentiment Analysis (SA) has proven to be a valuable tool in this field. It is a burgeoning technology focused on the Natural Learning Process (NLP), tapping into the demands of customers. However, SA faces many challenges like NLP and Word Sense Disambiguation. To overcome these challenges, Deep Neural Networks (DNN) like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) are used. There are primarily two study fields of this; Lexicon-based and ML-based, where each of the fields has further subdivisions. The introduction of Deep Learning (DL), which learns multiple layers of features and generates a result, has also enriched Machine Learning (ML).In this paper, various studies of ML algorithms like Naive Bayes', Support Vector machines (SVM) are compared with DL techniques like CNN and RNN. Feature extraction also plays crucial role in result prediction. For these keywords, we have used Term Frequency-Inverse Document Frequency (TF-IDF) and Word Embedding approaches. Although Naive Bayes' and SVM demonstrated satisfactory results, Vectorization and Tokenization for the TF-IDF method were not successful. Sentence tokenization and training with DL techniques have better performance. The RNN with an extra cell called the Long Short-Term Memory (LSTM) displays 89 percent reliability, better than other traditional models that fall in the range of 84%-86%.

Keywords-Sentiment Analysis; MachineLearning; DeepLearning; TF-IDF; Embedding of Words; Recurrent Neural Networks; Long Short-Term Memory.

I. INTRODUCTION

The rapidly developed web tool has brought revolutionary change for the growth of social media data. E-Commerce has become a big platform because it is used for many reasons like shopping, sharing valuable information, entertainment, household, and company things. To analyze, public opinions have become very necessary at this time because the behaviour and activities of people are influenced from one to another. Collecting unstructured data give meaningful information and provide a chance for new services for people in the business and government arena. Customers' feedbacks help both in improving business activities and in planning to increase profit.

The enormous growths of social networks generate very complex and intermingled information. There are theoretical and technical challenges. For solving the complexity, many studies are focused and have made powerful models in many areas like financing, marketing, forecasting, and health, among others. ML is applicable in many problems like NLP, information retrieval, computer vision, and robotics, to name a few. It has gained momentum because of its automatic ability to infer hierarchical structure used for regression, classification, and prediction by Weston et al.[1]. The Sentiment Analysis is a versatile technology with outstanding potential.

The various forms of DL neural networks are changing global communication, such as Convolutionary Neural Networks (CNN), Recurrent Neural Networks (RNN), Artificial Neural Networks (ANN). At the heart of the DL, revolutions are these various forms of neural networks, powering apps such as unmanned aerial vehicles, self-driving cars, speech recognition, etc. The feature vectors are input for the DL algorithms (CNN, DNN, and RNN). Hasan and Mahmood used the CNN and RNN model, which can overcome the shortcomings of small text [2]. Liu et al. considered three important criterias that are necessary that is informative quality reviews, it must be readable, and subjective [3]. Higher readability gives a higher quality of satisfaction. The reliability becomes affected when size of data changes

The rest of this article is as follows:

The ML and DL methods from the literature surveys are used for emotion detection. Different models have performed experimental comparison processes. The analysis that is proposed is highlighted based on the selected pre-available and accessible data sets. The privacy of staff is considered and data sets are carefully used to conduct the experiments and cover disparate subjects from different sources. Finally, as shown in Table-5, Precision, Recall, and F1-scores are measured, concluding that Long Short-Term Memory(LSTM) shows 89% reliability, which is higher than other conventional models, which are in the range of 84%-86%.

II. RELATED WORK

Zhang and Zheng discussed ML for Sentiment Analysis and used TF-IDF to calculate weights of words for analysis [6]. Bhavitha et al. proposed a comparative study of lexicon and hybrid approach and CNN, DNN, RNN for sentiment polarity and feature extraction on social media data[7]. Zhang et al. used 3 different CNN models for extracting features. Sensitive features were extracted by sentiment and semantic embedding [8]. Schmitt et al. used CNN and LSTM on seminal datasets. Aspect sentiment analysis was used for classification [9]. Jangid et al.used CNN, RNN, and LSTM for finance Tweets forimproving the deep learning technique efficiency [10]; Aspect- based analysis was done.

Jeong et al. used a topic modeling technique on social media data for the identification of products in customer review data[11]. Liu et al worked on movie reviews with CNN and LSTM showed the impact of textual features like word count, etc. [12]. Wu et al. used LSTM, BI-LSTM by comparing the study with the Hybrid approach on data sets of SemEVAL workshops for the classification of sentiments[13]. Kraus et al. used LSTM on movie review datasets for improving accuracy[14]. For the user-behavior information- feature extraction, CNN was used by Alharbi et al. from the Twitter analysis[15]. Pham et al. used analyzed travel review, and five aspect sentiments such as Value, Cleanliness, Service, Room, and Location were determined [16].

Gupta et al. combined semantic and sentiment features in the **Long Short- Term Memory(LSTM)** model.[17].Prithi et al. A recommender system in the cloud was used for analysis.[18]

Saleszirate et al. used feature-level sentiment analysis applied to diabetes ontology using SENTIWORD NET used DL method; the polarity- based method was used to increase accuracy [19]. A novel multilayer architecture was used by Qian et al. The same processes were used by many studies. Firstly, text features are extracted from different sources of data, and then are transferred into Word Embedding using the word2vec tool [20]. In this area, most methods are based on filtering information. They are classified into four categories as Content-based, Demographic-based, Collaborative, and Hybrid.

Content-based used characteristics of items and users' profile. The Demographic method exploits the age, gender, and nationality information of the user. The hybrid approach picks up the advantages of any item and information extracted from social media (actions, behaviour), etc. Explicit data is directly provided by the user and implicit data comes from the behaviour and action of the user.

Shoham proposed a Hybrid recommendation system that took advantage of content i.e., Collaborative filtering. Implicit user feedback is obtained by Sentiment Analysis[21]. Wang et al. used a hybrid approach for analysis of movie reviews for improving preliminary recommendation lists by combining the two methods as collaborative filtering and content-based method [22]. Singh et al. used sentiment classifier on movie reviews [23]. Stai et al. applied a framework named social recommendation [24]. Its main aim was to create of enriched multimedia of the users.

Difference between ML and DL

The difference between ML and DL are shown pictorially in Fig.1 and listed in Table-1.

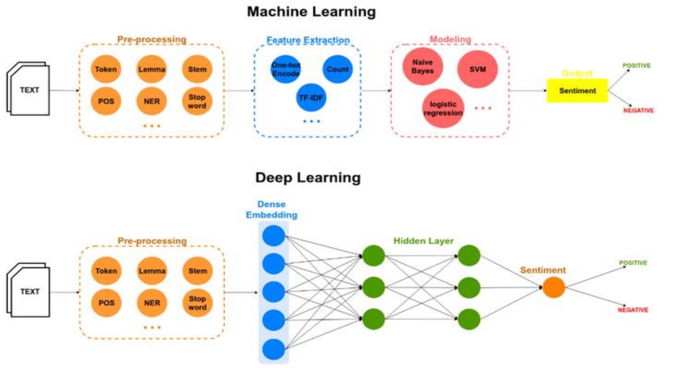

Fig.1: ML vs. DL

Architecture of DNN Vs. CNN

The difference between the architectures of DNN and CNN are shown pictorially in Fig.2 and listed in Table-2.

Table 2:DNN Vs. CNN

Deep Neural N/W(DNN)	Convolutional Neural N/W(CNN)
The DNN architecture uses mathematically sophisticated model data comprising three layers; 1. the input layer for feeding the data, second the hidden layer containing neurons, and third the output layer having one or many neurons. Data is fed after preprocessing.	CNN is a feed-forward NN. Convolutional layers and the pooling layers are used as input to the classification layer. Feature resolutions are reduced by pooling layer [4]. Classification is done by a fully connected layer.
Used in text feature extraction,etc.	Used in many areas like NLP, Computer Vision, Recommender system, etc. Mainly used for Unstructured Data having many layers.

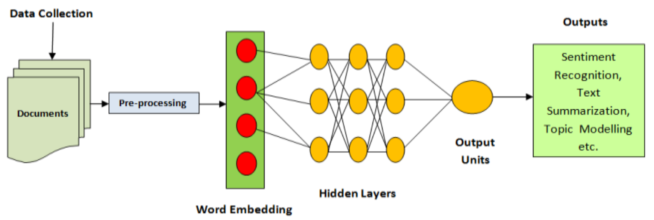

Fig. 2: DNN Architecture, Weston et al.(2008)[1]

Hubel and Wisel(1968) used cells ,which acts as a natural filter for analyzing the natural image. The combination of many layers is done as shown in Fig. 3.

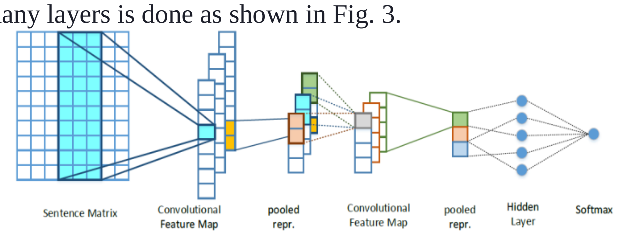

Fig. 3:CNN model in Sentiment Analysis,Deriu(2016)[29]

Basic features are learned by hidden layers and are fed to the output unit. Features are extracted from input data and every time this feature is incremented. The two convolutional layers that are on the left side have 64 and 32 filters used for training different features.
Next, the convolutional layers have 16 and 8 filters. After the first two and last two layers there is the max-pooling layer, that is used for reducing complexity and prevents overfitting of data. Last, is a fully connected layer called the output layer, which reduces the vector of height 8 to 1.Two classes are predicted(positive or negative).

Architecture of RNN Vs. LSTM

The difference between the architectures of RNN andLong Short- Term Memory(LSTM) are shown pictorially in Fig.4, Fig.5, respectively ,and listed in Table-3.

Table-3:RNN Vs. LSTM

Recurrent Neural Networks(RNN)	**Long Short- Term Memory(LSTM)**
A neural network whose neuron connections make a direct cycle.	It is a typical type of RNN; RNN with one cell extra called LSTM.
It typically captures internal memory; captures previous information and reuses it by applying it to the next element in a sequential manner Hockieter and schminder[5].	Data is processed for embedding matrix as shown in Fig 4:
Things are remembered for a small duration of time. If lots of words are fed in, pieces of information are lost.	LSTM uses long memory as an input to the hidden layer.

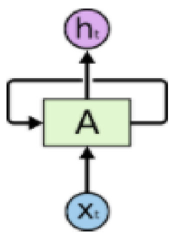

Fig. 4: RNN have loops,Norah Saleh Alghamdi[30]

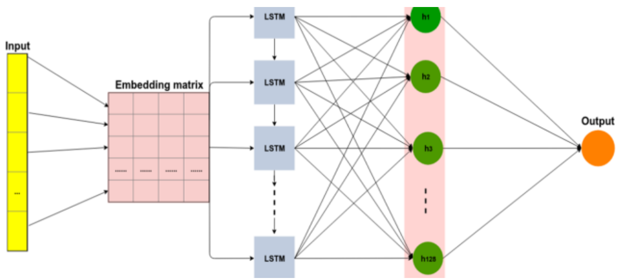

Fig. 5: LSTM Model,Hochreiter and chmidhuber(1997)[5]

Sentiment Analysis Steps:

Sentiment Analysis done by DL and ML require the following two steps:

1. Clean data before inducing it to the classification model.
2. Tweets usually have white spaces, punctuations, RT and stop words, etc., which can be eradicated by using libraries like Beautiful Soap.

Clean tweets can be split into words individually and are transformed into base form by lemmatization. After this, they are converted into the vector by word embedding. Then two approaches TF-IDF and word embedding are used for features and modeling languages in which each word turns into a vector of real views. The Word2vec is commonly used for word embedding. The Scikit- learn library is used for computing TF-IDF. Intensity and Polarity are reused for the scoring of Sentiment Analysis.

III. RESEARCH METHOD

For Sentiment Analysis, five types of data sets are used: IMDB, Music Reviews, Sentiment 140, Tweets SemEval, Kaggle Data Sets
As self-generated data labeling is a difficult task, the proposed work depends on selected data sets that are pre-available and accessible. Personnel privacy is kept into account; Data sets are used carefully from different sources for performing experiments and covering disparate topics. Data set size is also taken into consideration as Big Data offers more possibilities. Dataset description is given below:
Tweets SemEval: Data sets containing geopolitical entities about 70000 onsite tweets are taken [25].
IMDB: Movie reviews containing comments of people.
A total of 40000 samples are divided into two classes; positive or negative[26].
Music Reviews: Datasets containing user's comments obtained from the department of computer science of John's Hopkin University [27].
Sentiments 140: From Stanford University containing Tweets of products labeled with polarity 0 is considered as **negative, 1 as** positive[28].
Kaggle Data Sets: Data Sets were collected by searching the WebPages of Amazon.com.[31]
The research is focused on the comparison of various methods.
The experiment is done with all studies. TF-IDF and Word Embedding are used with a model in two stages:
In the 1st stage, which is the preprocessing stage, data is cleaned and feature-extraction is done and the 2nd stage is the training stage.
Cleaning means the removal of words not having information.
1. **Tweet Cleaning**: RTS,@,# and links
2. **Lemmatization:** Removal of stop words, Conversion of text to lower case, Removing punctuation; Extra white space is eliminated.
3. **Decoding of HTML** to general and dropping a column, this is not used for analysis (ID, Dates, Strings, Users, Query, etc.
After cleaning datasets, sentiments are separated into words individually. Words are converted into a real number of vectors by lemmatization.

These feature vectors are input for the DL algorithms(CNN, DNN, and RNN). Information goes iteratively as the RNN is trained, thereby rendering the network unstable by accumulating error ingredients. The weight value became very high causing an overflow and not emitting any values. It was the main drawback that was solved in LSTM. Owing to forgetting, recalling, and updating data, it showed its importance. The model construction is carried out with the Keras library comprising four layers:

- **Embedding layer**: It reduces the size of the input.
- **Drop out layer:** Prevents from overfitting.
- **LSTM layer:** Output is sequenced rather than a single value output to the underneath LSTM layer

Dense layer: for converting the output to binary.

IV. RESULTS AND DISCUSSION

Algorithm:

Precondition: Input is given as word sequence; the word is encoded as a word vector by an embedded method.

Input: *The TF-IDF and Word Embedding techniques combine before feeding into the model*.

Output: A dimension vector that is sent for pooling and relevant features are obtained from the feature set.

The stepwise algorithm is as follows:

Step1. embedding words

Step2. receiving vector as input

Step3. words that are passed to the embedding layer go to the LSTM cell and then to the output layer for prediction. The single unit works as an output layer.

Data preprocessing was made by fixing the size of a sample of 200.

The short review padded with zero in front and longer to 200. The encoded labels are 1 for positive and 0 for negative.

According to this process, data comes in nice shape and is split into training, validation, and test sets. The result is shown in Table-4:

Table-4 Results of Data Split

Datasets	**Accuracy**	**Precision**	**Recall**	**F1 score**
TweetSemEval	0.88	0.73	0.86	0.86
IMDB	0.89	0.88	0.87	0.88
Music Review	0.89	0.88	0.88	0.86
Kaggle DataSets	0.89	0.88	0.86	0,86
Sentiments 140	0.88	0.87	0.87	0.86

The Accuracy Model was fitted for checking. The Word2vec is used to train word vectors. After comparing the performance of different trained models, Precision, Recall, and F1-Scores are shown in Table-5.

Table-5: Performance Comparison of different trained models

ML Algorithms	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Naive Bayes	0.84	0.84	0.84	0.84
SVM	0.83	0.83	0.82	0.83
DNN	0.85	0.84	0.82	0.84
CNN	0.86	0.85	0.84	0.85
LSTM	0.89	0.90	0.88	0.90

V. CONCLUSION

In this paper, various methods of Sentiment Analysis are calculated by traditional models like Naive Bayes' and SVM Vs. Deep Learning (DL) model. The TF-IDF and Word Embedding techniques combine for the transformation of input before feeding into the model.

Experiments have been done on various data sets of different topics related to topics and reviews. Precision, Recall, and F1-scores are calculated as shown in Table-5:

LSTM is a specific type of RNN with an extra cell. Although, Naive Bayes and SVM also showed results satisfactory, however, vectorization and tokenization were not efficient for the TF-IDF approach. The sentence tokenization and training with DL methods give better results.

VI. FUTURE SCOPE

The purpose of comparing different techniques related to different studies guide researchers for achieving good results. Future studies will be focused on a hybrid approach.

REFERENCES

[1]. Collobert, R. and Weston, J., 2008, July, A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning* (pp. 160-167).

[2]. Hassan, A. and Mahmood, A., 2017, April. Deep learning approach for sentiment analysis of short texts. In *2017 3rd international conference on control, automation and robotics (ICCAR)* (pp. 705-710), IEEE.

[3]. Liu, J., Chang, W.C., Wu, Y. and Yang, Y., 2017, August, Deep learning for extreme multi-label text classification,In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 115-124).

[4]. Hubel, D.H. and Wiesel, T.N., 1959, Receptive fields of single neurons in the cat's striate cortex. *The Journal of physiology*,*148*(3),pp.574-591.

[5]. Hochreiter, S. and Schmidhuber, J., 1997, LSTM can solve hard long time lag problems, In *Advances in neuralinformation processing systems* (pp. 473-479).

[6]. Zhang, X. and Zheng, X., 2016, July, Comparison of

text sentiment analysis based on machine learnin, *2016, 15th International Symposium on Parallel and Distributed Computing (ISPDC)* (pp. 230-233), IEEE.

[7]. Bhavitha, B.K., Rodrigues, A.P. and Chiplunkar, N.N., 2017, Comparative study of machine learning techniques in sentimental analysis. In *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)* ,(pp. 216-221). IEEE.

[8]. Zhang, Z., Zou, Y. and Gan, C., 2018, Textual sentiment analysis via three different attention convolutional neural networks and cross-modality consistent regression, *Neurocomputing*, *275*, pp.1407-1415.

[9]. Schmitt, M., Steinheber, S., Schreiber, K. and Roth, B., 2018, Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks, *arXiv preprint arXiv:1808.09238*.

[10]. Jangid, H., Singhal, S., Shah, R.R. and Zimmermann, R., 2018, Aspect-based financial sentiment analysis using deep learning. In *Companion Proceedings of the Web Conference*,pp. 1961-1966.

[11]. Jeong, B., Yoon, J. and Lee, J.M., 2019, Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis, *International Journal of Information Management*, *48*, pp.280-290.

[12]. Wu, C., Wu, F., Wu, S., Yuan, Z., Liu, J. and Huang, Y., 2019, Semi-supervised dimensional sentiment analysis with variational autoencoder, *Knowledge-Based Systems*, *165*, pp.30-39.

[13]. Do, H.H., Prasad, P.W.C., Maag, A. and Alsadoon, A., 2019, Deep learning for aspect-based sentiment analysis: a comparative review, *Expert Systems with Applications*, *118*, pp.272-299.

[14]. Kraus, M. and Feuerriegel, S., 2019, Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees. *Expert Systems with Applications*, *118*, pp.65-79.

[15]. Alharbi, A.S.M. and de Doncker, E., 2019, Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. *Cognitive Systems Research*, *54*, pp.50-61.

[16]. Pham, D.H. and Le, A.C., 2018, Learning multiple layers of knowledge representation for aspect based sentiment analysis. *Data & Knowledge Engineering*, *114*, pp.26-39.

[17]. Gupta, U., Chatterjee, A., Srikanth, R. and Agrawal, P., 2017. A sentiment-and-semantics-based approach for emotion detection in textual conversations. *arXiv preprint arXiv:1707.06996*.

[18]. Preethi, G., Krishna, P.V., Obaidat, M.S., Saritha, V. and Yenduri, S., 2017, Application of deep learning to sentiment analysis for recommender system on cloud, *International Conference on Computer, Information and Telecommunication Systems (CITS)* (pp. 93-97). IEEE.

[19]. Salas-Zárate, M.D.P., Medina-Moreira, J., Lagos-Ortiz, K., Luna-Aveiga, H., Rodriguez-Garcia, M.A. and Valencia-Garcia, R., 2017, Sentiment analysis on tweets about diabetes: an aspect-level approach. *Computational and mathematical methods in medicine.*

[20]. Qian, J., Niu, Z. and Shi, C., 2018, Sentiment Analysis Model on Weather Related Tweets with Deep Neural Network, In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing* ,pp. 31-35.

[21]. Shoham, Y., 1997, Combining content-based and collaborative recommendation, *Communications of the ACM*.

[22]. Wang, Y., Wang, M. and Xu, W., 2018, A sentiment-enhanced hybrid recommender system for movie recommendation: a big data analytics framework. *Wireless CommunicationsandMobile Computing*.

[23]. Singh, V.K., Mukherjee, M. and Mehta, G.K., 2011, Combining collaborative filtering and sentiment classification for improved movie recommendations, *International Workshop on Multi-disciplinary Trends in Artificial Intelligence* pp. 38-50, Springer, Berlin, Heidelberg.

[24]. Stai, E., Kafetzoglou, S., Tsiropoulou, E.E. and Papavassiliou, S., 2018, A holistic approach for personalization, relevance feedback & recommendation in enriched multimedia content. *Multimedia Tools and Applications*, *77*(1), pp.283-326.

[25]. http://alt.qcri.org/semeval2017: Sentiment Analysis in Twitter

[26]. [26] http://www.cs.cornell.edu/people/pabo/movie-review-data: movie-review data for use in sentiment-analysis experiments.

[27]. Blitzer, J., Dredze, M. and Pereira, F., 2007, June. Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification, *Proceedings of the 45th annual meeting of the association of computational linguistics* ,pp. 440-447.

[28]. http://help.sentiment140.com/site-functionality

[29]. Deriu, J.M. and Cieliebak, M., 2016, Sentiment analysis using convolutional neural networks with multi-task training and distant supervision on Italian tweets, Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, Napoli, Italy, 5-7,Italian Journal of Computational Linguistics.

[30]. Al Surayyi, W.I., Alghamdi, N.S. and Abraham, A., Deep Learning with Word Embedding Modeling for a Sentiment Analysis of Online Reviews in International Journal of Computer Information Systems and Industrial Management Applications.

[31]. https://www.kaggle.com/bittlingmayeramazonreviews#train.ft.txt.bz2

Dr. Kusum Mehta is working as Assistant Professor in Department of Computer Science and Engineering School of Engineering Science and Technology at Jamia Hamdard University.Her Research areas include sentiment analysis,machine learning,artificial intelligence,deep learning.She has !6 years of experience and also reviewed many papers also.

Ms. Juveria is an Assistant Professor in the Department of Computer Science and Engineering, School of Engineering Science and Technology at Jamia Hamdard University. Her research focuses on Data Mining, Machine Learning and Artificial Intelligence. Ms. Juveria holds a M.Sc. in Computer Science from Jamia Hamdard University (Gold Medalist). She has eleven years of teaching experience and her expertise extends to data analysis, software development, programming languages etc.

Dr Veena Mittal is working as professor in Galgotias College of Engineering and Technology,Greater Noida.She is having more than 16 years of teaching experience.She has many research papers to her credit including joournals of higher repute.She is also a reviewer in many journals including Inderscience,Willey etc.