PalArch's Journal of Archaeology of Egypt / Egyptology

# SENTIMENT ANALYSIS OF USER REVIEW TEXT THROUGH CNN AND LSTM METHODS

**Harinder Kaur**

Assistant Professor

Department of English

SSM College, Dinanagar, Punjab, India

**Harinder Kaur, Sentiment Analysis of User Review Text through CNN and LSTM Methods- Palarch's Journal Of Archaeology Of Egypt/Egyptology 17(12), ISSN 1567-214x**

## ABSTRACT

Sentiment analysis, a process of natural language processing, is gaining popularity these days due to hike in the number of Internet users. The Internet users put up their opinions in the form of reviews regarding some contexts like products, services, government, politics, medicine, and entertainment world to name a few. The evaluation of correct sentiments out of web text is one of the main focuses of business houses or large organizations. Several machine learning approaches are used to address the issue of sentiment prediction from raw web text. This paper contributes to the existing methods by proposing a combination of convolutional neural networks and long-short-term-memory network for evaluation of appropriate sentiments. Two open source datasets are extracted in the pre-processing phase of proposed model. The major contribution of this research work comes from the pre-processing phase of data where a novel zero-padding method is used for normalization of word features before configuring neural networks. The proposed system outperforms the baseline classifiers almost 13% higher than the best performed baseline system for the first dataset and 3% improvement in accuracy is observed for the second dataset.

***KEYWORDS:***Sentiment Analysis, Machine Learning, Convolution Neural Networks, Long Short Term Memory Network, User Reviews.

## 1. Introduction

Opinion extraction or sentiment analysis is the way toward utilizing natural language processing, text investigation, and insights to dissect client estimation. The best organizations comprehend the assessment of their clients—what individuals are stating, how they're stating it, and what they mean. Client feeling can be found in tweets, remarks, audits, or different spots where individuals notice your text. Conclusion Analysis is the space of understanding these feelings with programming, and it's an unquestionable requirement comprehend for designers and business pioneers in a cutting edge working environment (Chen, Lee, & Chen, 2020).

AI has been seeing a great development in overcoming any issues between the capacities of people and machines. Specialists and researchers the same, take a shot at various parts of the field to cause astonishing things to occur. One of numerous such territories is the space of Computer based text processing(Rehman, Malik, Raza, & Ali, 2019).

The plan for this field is to empower machines to see the world as people do, see it along these lines and even utilize the information for a large number of assignments, for example, word sense disambiguation, text Analysis and Classification, paraphrase detection, Recommendation Systems, Natural Language Processing, and so on. The headway in Computer based text processing with Deep Learning have been built and idealized with time, basically more than one specific calculation a Convolutional Neural Network(Lavanya & Parvathavarthini, 2019).

Similarly as with numerous different fields, propels in profound learning have brought feeling examination into the frontal area of bleeding edge calculations. Today we utilize common language handling, insights, and text investigation to separate, and distinguish the feeling of words into positive, negative, or unbiased classes(Oussous, Benjelloun, Lahcen, & Belfkih, 2019).

### 1.1. Convolutional Neural Network

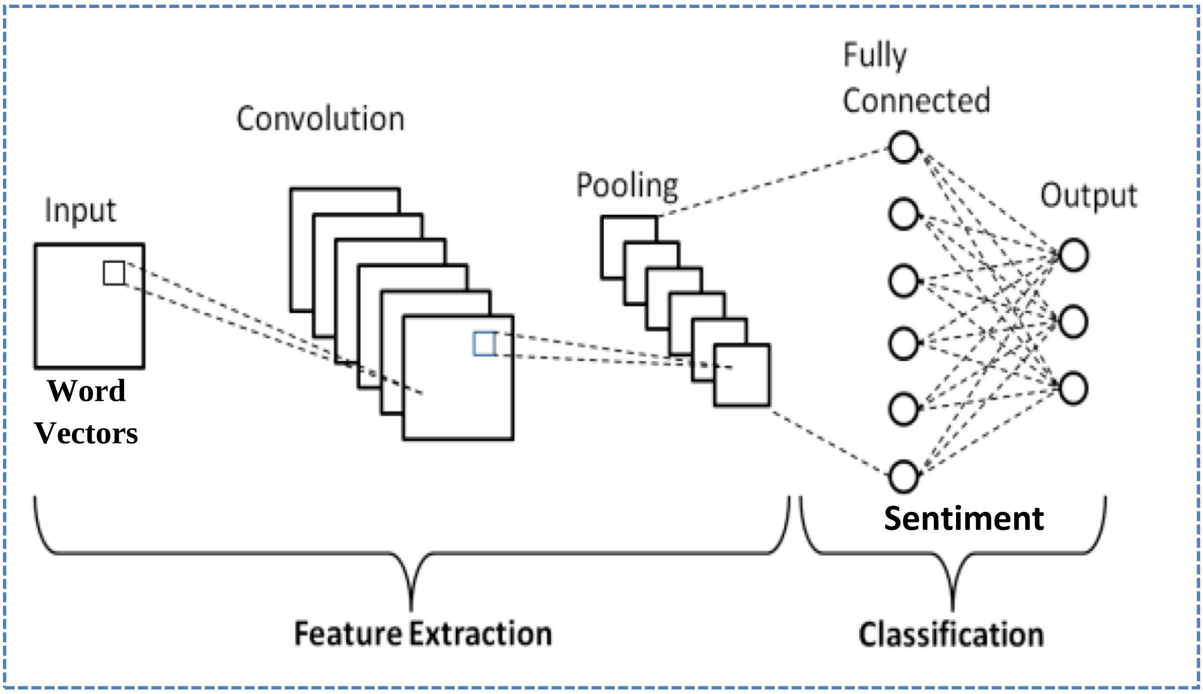


**Figure 1: Architecture for CNN**

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning calculation which can take in textual information, dole out significance (learnable loads and predispositions) to different sentiments and target words in the text and have the option to separate one from the other. The pre-preparing required in a ConvNet is a lot of lower when contrasted with other arrangement calculations. While in crude techniques channels are hand-built, with enough training, ConvNets can gain proficiency with these channels/attributes(Salim, 2019).

The design of a ConvNet is practically equivalent to that of the network example of Neurons in the Human Brain and was roused by the association of the Visual Cortex. Singular neurons react to improvements just in a limited area of the visual field known as the Receptive Field. An assortment of such fields cover to cover the whole visual territory(Lotfi & Akbarzadeh-T, 2013).

## 1.2. Long Short Term Memory Network (LSTM)

LSTM 's and GRU's were made as the answer for momentary memory. They have inner instruments considered gates that can control the progression of data. These gates can realize which information in an arrangement is essential to keep or discard. Thus, it can pass important data down the long chain of successions to make expectations. Practically all cutting edge results dependent on repetitive neural systems are accomplished with these two systems. LSTM's and GRU's can be found in discourse acknowledgment, synthesis of speech, and text generation. You can even utilize them to create subtitles for recordings(Ororbia, Mikolov, & Reitter, 2017).

Alright, so why LSTM's and GRU's are acceptable at handling long successions? Let us talk about this with natural clarifications and delineations and maintain a strategic distance from however much math as could reasonably be expected(Ororbia et al., 2017).

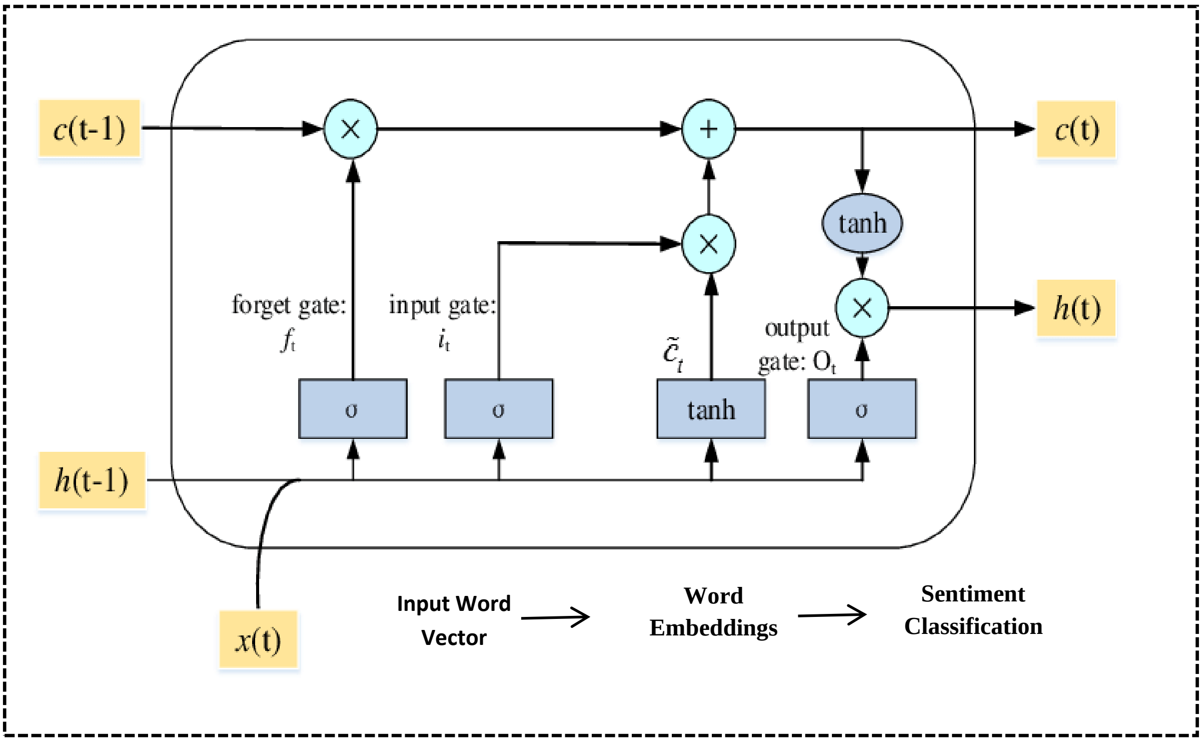


**Figure 2: Architecture for LSTM network**

Alright, let's beginning with a psychological test performed while purchasing something from supermarket. Suppose you're taking user reviews at surveys online to decide whether you need to purchase wheat or rice grains. You'll previously peruse the survey at that point decide whether somebody thought it was acceptable or on the off chance that it was terrible. At the point when you read the survey, your mind subliminally just recollects significant watchwords. You get words like "astounding" and "completely adjusted breakfast". You couldn't care less much for words like "this", "gave", "all", "should", and so forth. On the off chance that a companion asks you the following day what the survey stated, you presumably wouldn't recall it in exactly the same words. You may recall the central matters however like "will purchase once more". In case you're a ton like me, different words will blur away from memory. What's more, that is basically what a LSTM or GRU does. It can figure out how to keep just significant data to make expectations, and overlook non applicable information. For this situation, the words you recalled made you judge that it were acceptable(Ruthven, Buchanan, & Jardine, 2018).

## 2. Related Work

Most of the machine-learning methods are proved important for sentiment analysis of web text by achieving significant results so far. Almost all machine-learning frameworks started deploying neural network based deep learning models for the same. The feature extraction through advanced neural networks is gearing into its next turn in feature engineering. The newly adopted convolutional neural networks and LSTM networks perform excellently on some benchmark datasets. The following are some recent contributions (refer Table 1) in the field of sentiment analysis through deep learning constructs:

| Reference | Description of Work | Findings |
|---|---|---|
| (Jin, Yang, & Liu, 2019) | Prediction of stock's closing points using sentiment analysis and LSTM networks | Best performed proposed model yields 70.56% of accuracy in sentiment prediction |
| (Basiri, Abdar, Cifci, Nemati, & Acharya, 2020) | Fusion of machine learning and deep learning methods used for sentiment classification of user text | 85.98% for 3CNN |
| (Ji & Wu, 2020) | Proposed an architecture for sentiment prediction using rhetoric and discourse vector LSTM network models | 82.53% for D-LSTM proposed network |
| (Ghorbani, Bahaghighat, Xin, & Özen, 2020) | Coined a deep learning based method for sentiment analysis using cloud computing | LSTM sandwiched between convolutional layers yields 89.02% of accuracy. |
| (Mahmood et al., 2020) | Sentiment prediction from Roman Urdu language using deep learning method acquired from recurrent convolutional neural networks | 65.20% for RCNN model |
| (Grissette & Nfaoui, 2020) | Extracted dubbed vocabulary related to drugs from Twitter and utilized n-gram based word embedding model to train convolutional neural networks | Proposed system was 81% accurate in predicting sentiments from Twitter posts |
| (Chen et al., 2020) | Deep learning deployed on social media data to extract sentiments of Taiwan's biggest online forum | 88.41% for Bi-LSTM using Tanh activation function |
| (Li, Rzepka, Ptaszynski, & Araki, 2020) | A novel humour detection method for sentiment analysis of social media text | 89.79% of precision for proposed method |
| (Shah, Yan, Shah, & Mamirkulova, 2019) | Service quality evaluation and patient's opinion mining by deploying deep learning classifiers | Around 15 % improvement from existing methods |
| (García-Díaz, | Sentiment analysis of case studies of | Proposed approach |

| | | |
|---|---|---|
| Cánovas-García, & Valencia-García, 2020) | infectious diseases in Latin America based on ontology driven approach | outperforms among existing baseline methods |

**Table 1: Summary of literature in the latest submission on sentiment analysis**

## 3. Proposed Methodology:

Several web scrapers are available these days to scrape web text, Python's bs4 (beautiful soup version 4) is one of the finest web scraper. The pre-processing phase of this research work involved scraping of user text through bs4 library. The scraped text is stored in a text file followed by labelling of sentiments (positive or negative) through sentiment lexicons of Python 3.6. The already stored emotion words are compared with the user reviews, those sentences which are either containing similar words (synsets) or the same emotion words are being labelled with the respective sentiment. The labelling is manually validated through random tests from arbitrarily chosen sample sentences. Then the sentence length is normalized using zero padding method where all the sentences of a sample get fixed length property. The following step will act on the feature extraction part. The important features include word sense, word length, emotion, word similarity, slang words, stop words, nouns, articles and foreign words to name a few. These features are found by using feature extraction module of Python's NLTK. The following step does the training of convolutional neural network and LSTM network on training part of dataset. The testing and validation part only done after the successive training loop of network models through iteratively fed samples of training set. Lastly, the prediction of sentiment followed by trained model.

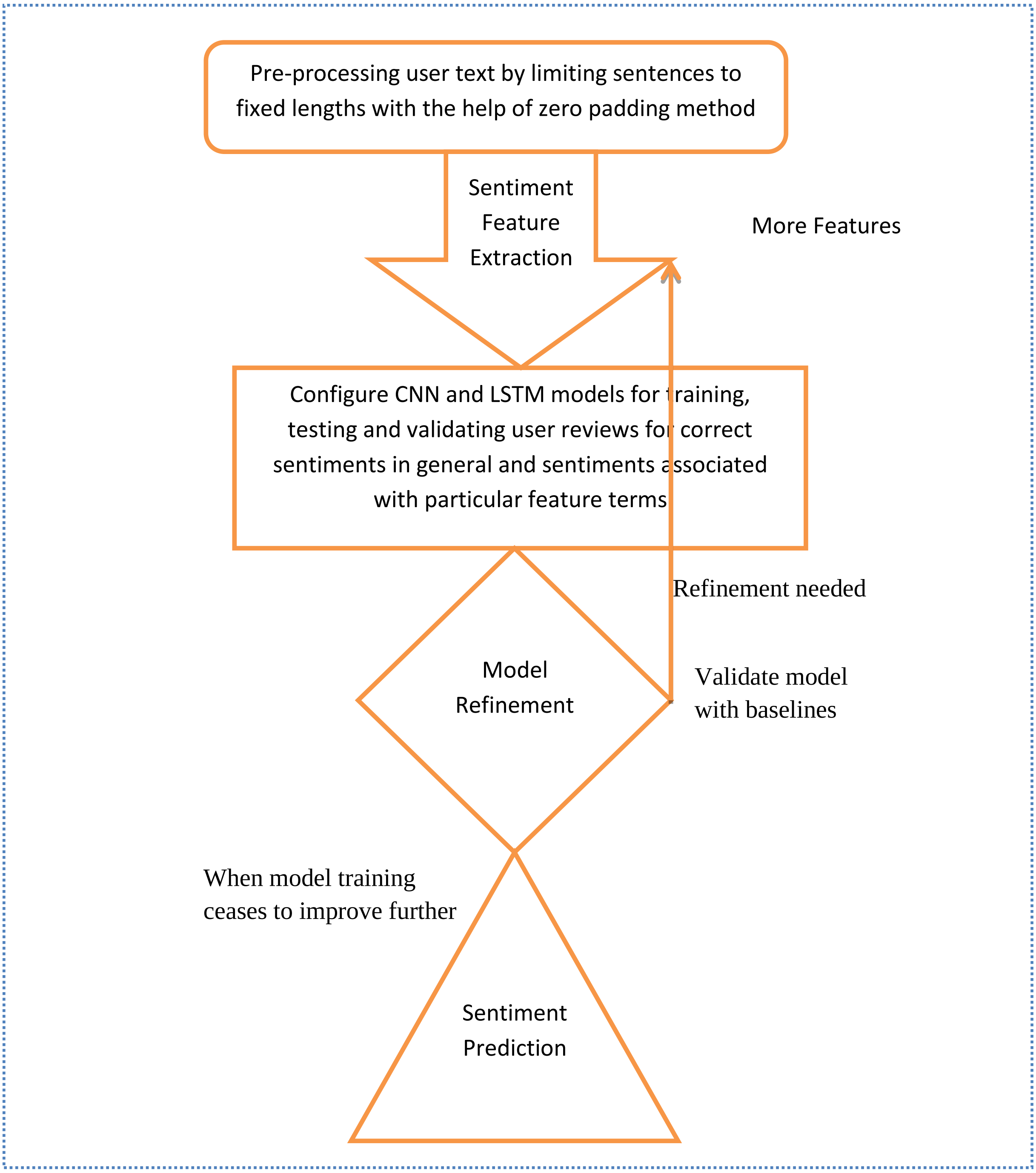


**Figure 3: Proposed methodology for sentiment prediction using CNN and LSTM**

## 4. Dataset Collection

There are abundant sources of free web text over the Internet these days. In the present day and age, brands of every kind imaginable have significant communications with clients, leads, and even rivalry on informal communities like Facebook, Twitter, and Instagram. Most promoting divisions are as of now tuned into online notices similar to volume – they measure more babble as more brand mindfulness. These days, be that as it may, organizations need to search for more profound bits of knowledge. By utilizing estimation examination via web-based networking media, we can get amazing bits of knowledge into the nature of discussion that is occurring around a brand.

### 4.1. Statistics of Wikipedia dataset

| Sentiment Label | Training | Test | Dev | Total |
|---|---|---|---|---|
| **1 (Extremely Positive)** | 2253 | 117 | 188 | 2558 |
| **2 (Positive)** | 1453 | 89 | 121 | 1663 |
| **3 (Negative)** | 835 | 21 | 35 | 891 |
| **4 (Extremely Negative)** | 1120 | 74 | 65 | 1259 |
| **Total** | 5661 | 301 | 409 | 6371 |

**Table 1: Statistics of Wikipedia dataset taken for the month of January, 2020. Labelling is done through Python's NLTK library. These numeric figures represents number of sentences used for training set, test set and development sets respectively.**

This research work considers user reviews from two freely available repositories viz., Wikipedia dataset for the month of January, 2020 and latest IMDB movie reviews. These two datasets are highly accessible and structured around the pre-defined contexts. This is why these datasets are good to be exploited for sentiment prediction. Around six thousand reviews from Wikipedia dataset and 18 thousand reviews from IMDB dataset are taken for the sentiment prediction task. The statistics of these two datasets are given in the following two tables (refer Table 1 and Table 2).

### 4.2. Statistics of IMDB Movie Reviews

| Sentiment Label | Training | Test | Dev | Total |
|---|---|---|---|---|
| **1 (Negative)** | 7253 | 1217 | 150 | 8620 |
| **2 (Positive)** | 8497 | 1489 | 160 | 10,146 |
| **Total** | 15,750 | 2706 | 310 | 18,766 |

**Table 2: Statistics of movie reviews taken from IMDB through web scraping using Python's bs4 library**

## 5. Proposed Algorithm:

A sentiment prediction task is typically displayed as an classification problem, whereby a classifier is taken care of a text and returns a classification, for example positive, negative, or impartial. In the training procedure, our model figures out how to relate a specific information (for example a text) to the relating yield (tag) in view of the test samples utilized for training. The element extractor moves the content contribution to a component vector. Sets of highlight vectors and labels (for example positive, negative, or unbiased) are taken care of into the machine learning to create a model.

The proposed algorithm generates word embeddings using word2vec, gloVe and libSVM models to train the CNN and LSTM classifiers. The zero-padding step will provide the normalization to the sentence length and word length features. In the expectation procedure (step 6), the element extractor is utilized to change concealed content contributions to highlight word-features. These element features are then taken care of into the model, which creates anticipated labels (once more, positive, negative, or nonpartisan).

*Step 1: Given a sentence or message S = {w1, w2, w3, ………,wn}*

*Step 2: Transform n words into feature vector a =(a1,a2,a3,………an)*

*Step 3: Generate word embeddings wi using word2vec, gloVe and libSVM.*

*Step 4: Remove articles, prepositions, and stop words that carry no sentiment.*

*Step 5: Python Feature Extraction module extracts word features using pre-defined lexicons.*

*Step 6: Perform zero padding to maintain the word length feature.*

$$N = \frac{Senti\_word_i}{\left|Senti\_word_j\right| \sum Senti\_word_k} \times p$$

*Step 7: Train CNN and LSTM models using observed features.*

*Step 8: Test the unlabelled sentences for sentiments.*

The initial phase in an AI text classifier is to change the content extraction or text vectorization, and the old style approach has been pack of-words or sack of-ngrams with their recurrence.

The new element-extraction methods have been applied dependent on word embeddings (word2vec, gloVe and libSVM). This sort of portrayals makes it workable for words with comparable significance to have a comparable portrayal, which can improve the exhibition of classifiers.

## 6. Results and Observations

The following Table 3 shows the user reviews taken from IMDB dataset. The proposed model classified these reviews as positive or negative based on the degree of polarity they carry in their respective sentiment lexicons.

| User Review | Sentiment Words |
| --- | --- |
| **To my surprise, I enjoyed this movie!**<br>valtierraitza 6 January 2019<br>Went in without watching trailers or reading any background info on it. My friends really wanted to see so I tagged along. With horror movies, I never get my hopes up as it's hard to find a good one these days. I surprisingly really enjoyed the concept of this movie and would consider it more of a thriller. It never bored me and it wasn't gory like the saw films (which I never enjoyed). I liked the characters and their backgrounds.. it was nice to see how they all tied in together. It was a pretty good movie and I'd recommend.<br>253 out of 351 found this helpful. Was this review helpful? Sign in to vote.<br>Permalink | Surprise, enjoyed, horror, thriller, liked, nice, pretty, good, recommend |
| **It's scary that some people are giving this 10/10**<br>thailerrhyne 13 January 2019<br>Don't get me wrong, the people who are saying "1 star, it's visual vomit" are being dramatic. But this movie is not good. It's beginning and half way point are somewhat decent. Plot wise I mean. Acting and visual effects are lacking in many areas. But the second half till the end is horribly executed. The fact that it sequel baits you hard at the end is also a little insulting. Save your money. Do a real escape room instead.<br>245 out of 413 found this helpful. Was this review helpful? Sign in to vote.<br>Permalink | Scary, wrong, visual vomit, dramatic, not-good, lacking, insulting, little, escape, instead |

| | |
|---|---|
| **Boring,bland,mediocre performances at best**<br>nick_papadai 10 January 2019<br><br>This was supposed to be a mystery thriller. I guess like the series Saw or The Cube/ The exam. It is about some guys that are invited "randomly" to an escape room and they try to figure the way out and the reasons behind it. The story is very predictable which is sad to say the least, the acting is pretty bad as well and I'm not saying this from an elitist viewpoint because the actors aren't A tier, it's very below average. The viewer can't really search for clues in the movie which is kinda disappointing as well.<br><br>There is a twist somewhere but it isn't worth the time. If you are like me and expect a Saw tier movie or actual mystery thriller, you will get very very disappointed from this. As a big fan of the genre, I had to see it but I wish I hadnt<br><br>40 out of 76 found this helpful. Was this review helpful? Sign in to vote.<br>Permalink | Boring, bland, mediocre, mystery, escape, sad, predictable, pretty bad, below average, disappointing, mystry, hadnt |

**Table 3: Examples of user reviews and sentiment words found using proposed model**

## 6.1. Validation against baselines for Wikipedia dataset

The Table 4 below depicts the accuracy of classification for our proposed model validated against the other three baselines for the first dataset. The experimental results show that proposed model outperforms the baseline classifiers almost 13% higher than the best performed baseline system for CNN.

| **Classifier** | **Word vector method** | **Accuracy (%age)** |
|---|---|---|
| Logistic Regression | LibSVM | 39.72 |
| Support Vector Machine | word2vec | 39.90 |
| CNN | GloVe | 46.40 |
| CNN+LSTM (Proposed) | Word embeddings | **59.27** |

**Table 4: Validation of classification accuracies against baselines for first dataset**

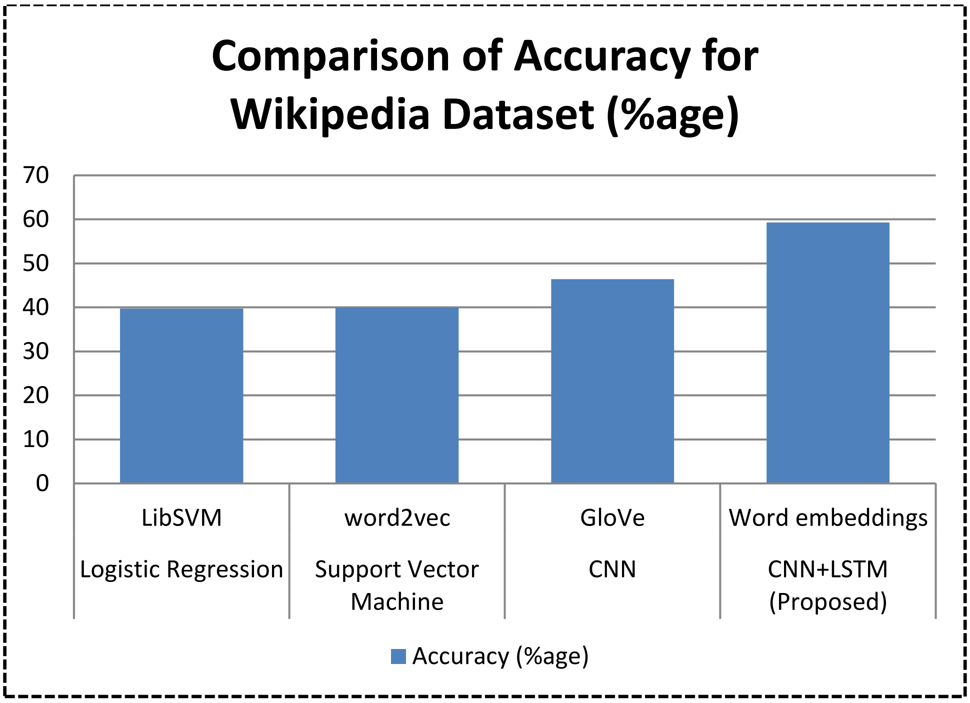


**Figure 4: Accuracy validation of proposed model for first dataset**

### 6.2. Validation of accuracies for IMDB dataset

The proposed model outweighs the best performed baseline by almost 3% in terms of accuracy of classification. The proposed model is benefitted in the pre-processing phase due to normalization performed through zero-padding process. Also the parallel combination of CNN and LSTM networks helped in correctly classifying the user reviews. The following Table 5 shows the validation of proposed model against the baseline methods.

| Classifier | Word vector method | Accuracy (%age) |
|---|---|---|
| Logistic Regression | LibSVM | 57.10 |
| Support Vector Machine | word2vec | 78.54 |
| RNN | word2vec | 92.51 |
| CNN | GloVe | 88.77 |
| CNN+LSTM (Proposed) | Word embeddings | **95.01** |

**Table 5: Validation of classification accuracies against baselines for second dataset**

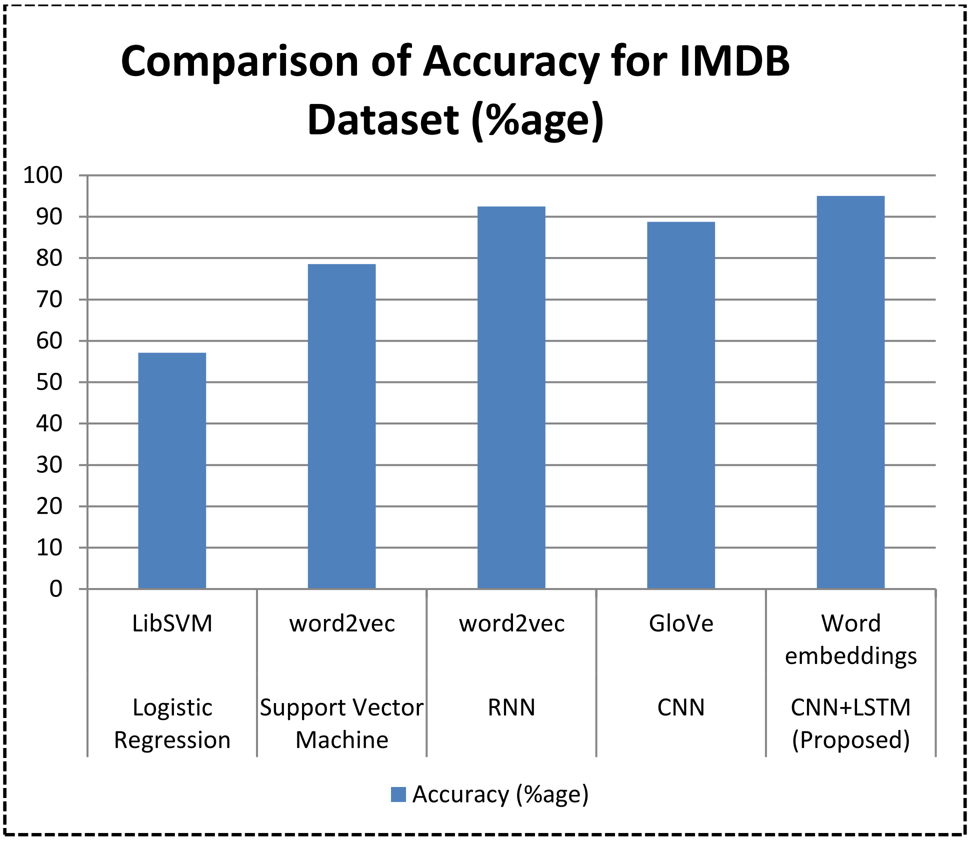


**Figure 4: Accuracy validation of proposed model for second dataset**

## 7. Conclusion

The proposed model outweighs the best performed baseline by almost 3% in terms of accuracy of classification for the IMDB dataset. The experimental results show that proposed model outperforms the baseline classifiers almost 13% higher than the best performed baseline system for CNN. However, there are certain shortcomings which are neglected while data pre-processing. The very first issue is the challenge of emojies processing, because emojies in the user reviews carry real sentiments. Since our pre-processing phase removes all the special characters like [☺] and [☹]. The second real challenge was time factor; some of the initial reviews are of opposite polarity sometimes, whereas the following reviews reverse the polarity by adding some certain logics and introducing the wave of reviews of same polarity. Thirdly, the identities of authors of user reviews are not considered in the pre-processing phase. It is seen most of the times that some fake reviewers inject negativity in order to win the market confidence in case of product reviews. The future scope of this work will address these issues by fortifying the pre-processing phase of proposed model. Addressing these issues while in the dataset evaluation phase may enhance the productivity and accuracy of sentiment classifier.

## References

Basiri, M. E., Abdar, M., Cifci, M. A., Nemati, S., & Acharya, U. R. (2020). A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. *Knowledge-Based Systems*, *198*, 105949. https://doi.org/10.1016/j.knosys.2020.105949

Chen, L. C., Lee, C. M., & Chen, M. Y. (2020). Exploration of social media for sentiment analysis using deep learning. *Soft Computing*, *24*(11), 8187–8197. https://doi.org/10.1007/s00500-019-04402-8

CABALLERO, ARLENE R., JASMIN D. NIGUIDULA, and JONATHAN M. CABALLERO. "TWITTER FEEDS SENTIMENT ANALYSIS AND VISUALIZATION." International Journal of Educational Science and Research (IJESR)7.4, Aug 2017, 31-40

García-Díaz, J. A., Cánovas-García, M., & Valencia-García, R. (2020). Ontology-driven aspect-based sentiment analysis classification: An infodemiological case study regarding infectious diseases in Latin America. Future Generation Computer Systems, 112, 641–657. https://doi.org/https://doi.org/10.1016/j.future.2020.06.019

Ghorbani, M., Bahaghighat, M., Xin, Q., & Özen, F. (2020). ConvLSTMConv network: a deep learning approach for sentiment analysis in cloud computing. Journal of Cloud Computing, 9(1), 1–12. https://doi.org/10.1186/s13677-020-00162-1

Grissette, H., & Nfaoui, E. H. (2020). Enhancing convolution-based sentiment extractor via dubbed N-gram embedding-related drug vocabulary. Network Modeling Analysis in Health Informatics and Bioinformatics, 9(1). https://doi.org/10.1007/s13721-020-00248-5

Ji, C., & Wu, H. (2020). Cascade architecture with rhetoric long short-term memory for complex sentence sentiment analysis. Neurocomputing, 405, 161–172. https://doi.org/10.1016/j.neucom.2020.04.055

Jin, Z., Yang, Y., & Liu, Y. (2019). Stock closing price prediction based on sentiment analysis and LSTM. Neural Computing and Applications, 32(13), 9713–9729. https://doi.org/10.1007/s00521-019-04504-2

Lavanya, S. K., & Parvathavarthini, B. (2019). Co-extraction of feature sentiment and context terms for context-sensitive feature-based sentiment classification using attentive-LSTM. Applied Mathematics and Information Sciences, 13(5), 749–758. https://doi.org/10.18576/amis/130507

Li, D., Rzepka, R., Ptaszynski, M., & Araki, K. (2020). HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. Information Processing and Management, 57(6). https://doi.org/10.1016/j.ipm.2020.102290

Lotfi, E., & Akbarzadeh-T, M. R. (2013). Brain emotional learning-based pattern recognizer. Cybernetics and Systems, 44(5), 402–421. https://doi.org/10.1080/01969722.2013.789652

Mahmood, Z., Safder, I., Nawab, R. M. A., Bukhari, F., Nawaz, R., Alfakeeh, A. S., … Hassan, S. U. (2020). Deep sentiments in Roman Urdu text using Recurrent Convolutional Neural Network model. Information Processing and Management, 57(4), 102233. https://doi.org/10.1016/j.ipm.2020.102233

Mahmood, Q. A. S. I. M., and K. H. U. R. S. H. I. D. Ahmad. "Portrayal of the US in leading Pakistani newspapers: An analysis." International Journal of Humanities and Social Sciences (IJHSS) 2.2 (2013): 21-29.

Madhavi Lakshmi, P., and P. Siva Pratap. "HR Analytics-a Strategic Approach to HR Effectiveness." International Journal of Human Resource Management and Research (IJHRMR) ISSN (P) (2016): 2249-6874.

Mulkalwar, Anurag, and Kavita Kelkar. "Sentence Level Sentiment Classification Using HMM with the help of Part of Speech Tagging." International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) (2014).

Ororbia, A. G., Mikolov, T., & Reitter, D. (2017, March 26). Learning simpler language models with the differential state framework. Neural Computation. https://doi.org/10.1162/NECO_a_01017

Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2019). ASA: A framework for Arabic sentiment analysis. Journal of Information Science. https://doi.org/10.1177/0165551519849516

Rehman, A. U., Malik, A. K., Raza, B., & Ali, W. (2019). A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis. Multimedia Tools and Applications, 78(18), 26597–26613. https://doi.org/10.1007/s11042-019-07788-7

Ruthven, I., Buchanan, S., & Jardine, C. (2018). Isolated, overwhelmed, and worried: Young first-time mothers asking for information and support online. Journal of the Association for Information Science and Technology, 69(9), 1073–1083. https://doi.org/10.1002/asi.24037

Salim, N. (2019). Deep learning approaches for big data analysis. In

International Conference on Electrical Engineering, Computer Science and Informatics (EECSI) (pp. 1–2). Institute of Advanced Engineering and Science. https://doi.org/10.23919/EECSI48112.2019.8977075

Shah, A. M., Yan, X., Shah, S. A. A., & Mamirkulova, G. (2019). Mining patient opinion to evaluate the service quality in healthcare: a deep-learning approach. Journal of Ambient Intelligence and Humanized Computing, 11(7), 2925–2942. https://doi.org/10.1007/s12652-019-01434-8

Sulphey, M. M. "The utility of Q-methodology in Human Resource Management research." International Journal of Human Resources Management 3.3 (2014): 15-26.