

# Sentimental Classification of News Headlines using Recurrent Neural Network



S. Prakashini, D. Vijayakumar

**Abstract:** Sentiment analysis combines the natural language processing task and analysis of the text that attempts to predict the sentiment of the text in terms of positive and negative comments. Nowadays, the tremendous volume of news originated via different webpages, and it is feasible to determine the opinion of particular news. This work tries to judge completely various machine learning techniques to classify the view of the news headlines. In this project, propose the appliance of Recurrent Neural Network with Long Short Term Memory Unit(LSTM), focus on seeking out similar news headlines, and predict the opinion of news headlines from numerous sources. The main objective is to classify the sentiment of news headlines from various sources using a recurrent neural network. Interestingly, the proposed attention mechanism performs better than the more complex attention mechanism on a held-out set of articles.

**Keywords:** Classification, Clustering, News headlines, Recurrent Neural Network.

## I. INTRODUCTION

The most generally used technique in natural language processing tasks is sentiment analysis. It is useful in cases such as customer surveys, reviews, news data, and social media comments where people express their opinions and feedback. It combines the analysis of text and Natural Language Processing task that attempts to predict the sentiment of text by analyzing a large number of documents. For example, "Good Morning, have a nice day" is a general positive text. In general, sentiment analysis target is to find out the thought of the person who speaks/write concerning specific topics. It is otherwise called as opinion mining. It can be done at the document level, the sentence level, and the aspect level. In the document level, the document has examined and also find the opinion of that whole document. In sentence level, every sentence has managed and analyzed to seek out the polarity of the sentence. In aspect level, it communicates the opinion itself. The existing system implements Naïve Bayes' algorithm to perform sentimental analysis. In this paper, use the Recurrent Neural Network to perform sentimental classification of news headlines to achieve high accuracy.

The rest of this paper composed as follows: Part II represents the review paper conducted in sentiment analysis of news articles. Part III describes the proposed methodology of sentiment classification. Part IV represents the outcome of the proposed method. Finally, Section V represents the conclusion of this research.

## II. RELATED WORK

Somanath et al. have performed a classification of sentiment on News Headlines in the US Newspaper about India. In this paper, machine learning techniques and semantic orientation approach used for sentiment classification. The results of the sentimental analysis of news headlines classified as positive, negative, and neutral sentiments. This opinion can assist the Indian government with ongoing sentiment analysis on news headlines in the US Newspaper related to India [2].

Islam et al. have conducted sentence structure and dynamic lexicon for opinion detection in online news. To find the polarity of the news, they used the sentence level and different dictionary with predefined positive and negative words. The process involved in news classification. 1) Online news article selection 2) Sentence extraction from news articles, 3) Seek for positive phrases in the extracted sentences, and finding their polarities of those sentences. 4) Integrating the polarities of all sentences to acquire the sentiment of a news article [1]. Taj et al. have led a dictionary-based methodology for the classification of news articles in terms of positive and negative sentiment. In this method, the opinions delivered by a word deduced on the grounds of the view the polarity of the individual words in the document that convey the sentiment of the text. Thus the sum of contradictions of every phrase in the sentence is the sentiment of a sentence. [3].

Apoorv Agarwal et al. has examined individual word in the features of news, whether it is a thing, action word, intensifier, descriptor, or some other grammatical form. In this paper, they find out the positive and negative scores of an individual word using sentiwordnet and classify the words using python library and also estimating the overall positive/negative effect in that news headlines [1].

## III. PROPOSED SYSTEM

The proposed framework depends on the Recurrent Neural Network of news headlines is used for sentiment classification of news headlines. It comprises several stages, such as the gathering of data, data cleaning, extraction of feature, clustering, and sentiment classification.

Revised Manuscript Received on April 30, 2020.

\* Correspondence Author

**S.Prakashini\***, Computer Science and Engineering, National Engineering College, Kovilpatti, India. Email: [182807@nec.edu.in](mailto:182807@nec.edu.in)

**D.Vijayakumar**, Computer Science, and Engineering, National Engineering College, Kovilpatti, India. Email: [dvk@nec.edu.in](mailto:dvk@nec.edu.in)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

# Sentimental Classification of News Headlines using Recurrent Neural Network

Data gathering involves obtaining news headlines from various sources such as Times of India, The Hindu, Hindustan Times, Indian Express, etc. After the collection of a dataset, the data will be cleaned using the NLP technique. Next, the useful features like term recurrence, grammatical feature (POS) data, sentiment word based on unique lexicon are recognized and extracted. Then clustering of data will be performed to seek out similar news headlines. Finally, apply the Recurrent Neural Network with long short term memory unit for classifying the sentiment of news headlines, which say positive or negative.

### Advantage:

- RNN technique is more efficient compared to other methods in prediction for Sentimental Classification.
- Reduce processing time.

### 3.1 Proposed System Flow

The following figure 3.1 shows the flow of sentimental classification using RNN

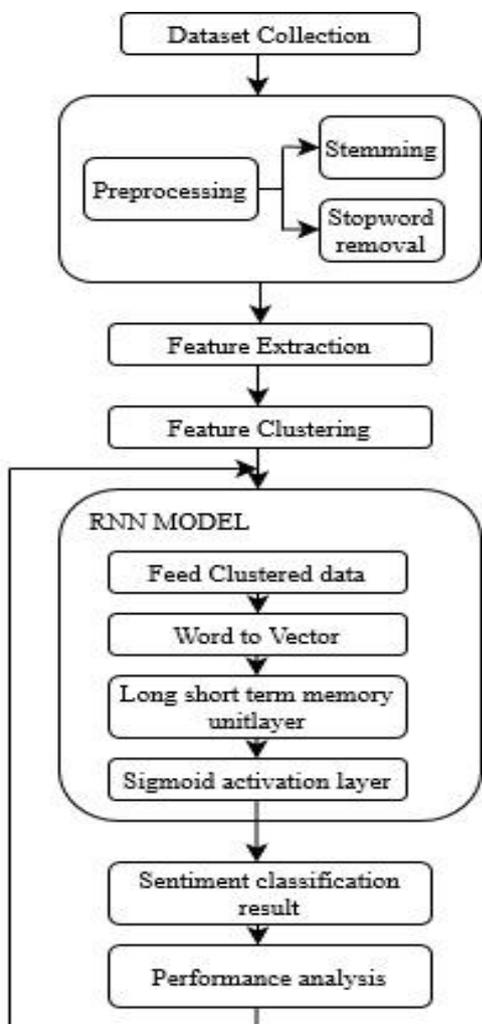


Fig 3.1 Flow of proposed system

The different modules of the analysis are explained below.

#### 3.1.1 Dataset Collection

News headlines are collected from various sources such as Times of India, Hindustan Times, The Hindu, etc. using a BeautifulSoup library in Python.

#### 3.1.2 Data Preprocessing

The collected news data will be preprocessed using the Python library.

##### 3.1.2.1 Tokenization

Tokenization can be done using `sent_tokenize`. In this process, the sentence is split into a list of words called as tokens, which comprises all alpha characters, and numerical characters only.

##### 3.1.2.2 Normalization

In this method, shifting all the tokens into either lower case or upper case because most of the news comprises both instances. The purpose of shifting all words into one format can be easily utilized for prediction.

##### 3.1.2.3 Stemming

Stemming is the process of changing all the tokens into their stem or root structure. Stemming is a simple and easy methodology that makes the feature extraction process more effortless. It can be done using a snowball stemmer. For example, the word "predicted" can be stemmed to "predict." It is used to diminish the inflection form.

##### 3.1.2.4 Stop word removal

Stop words are ubiquitous and high-recurrence words. This method is carried out by removing recurrently used stop words (prepositions, inappropriate words, special character, ASCII code), newline, extra white spaces, etc. to improve the performance of feature selection technique. It can be done using the stop word library.

#### 3.1.3 Feature Extraction

The method of feature extraction is used to reduce the number of resources required for processing without losing important or relevant information. It can also reduce the amount of redundant data for a given analysis. The resulting scores of Term frequency and TF-Inverse document frequency assigned to each word, are recognized and extracted.

#### 3.1.4 RNN Classification

In RNN classification, the first step is embedding of the word into the vector using the word2vector model. The output of word2vector is called the sequence matrix. Now the word vector as input for a recurrent neural network with Long Short Term Memory Unit (LSTM). There are several layers in RNN with LSTM: 1) LSTM defined by several layers. The addition of LSTM units makes it possible to find out the exact and accurate information; 2) Fully connected layer that maps the output of the LSTM layer to the intended output size; 3) Sigmoid Activation Layer that turns all output values into a value between 0 and 1. Finally, the Sigmoid output from the last time step is considered as the final output of this network.

#### 3.1.5 Performance Analysis

The performance analysis can be done by following steps

- Creation of Confusion matrix

- Calculate  $TP = TP/(TP+FN)$   
 $FP = FP/(FP+TN)$   
 $TN = TN/(FP+TN)$   
 $FN = FN/(FN+TP)$
- Calculate Precision =  $TP/(TP+FP)$   
Recall =  $TP/(TP+FN)$   
Accuracy =  $(TP+TN)/(TP+TN+FP+FN)$

Where TP = True Positive (The algorithm exactly estimate the positive class).

FP = False Positive (The algorithm estimate the inappropriate positive class).

TN = True Negative (The algorithm accurately estimates the negative class).

FN = False Negative (The algorithm estimate the inappropriate negative class).

- Recall indicates a suitable outcome, which is correctly classified by the model.
- Precision indicates the relevant outcome.
- Accuracy indicates the exact or appropriate outcome.

#### IV. RESULTS AND DISCUSSION

The Following fig 4.1 shows that the News headlines are collected from various sources like The Hindu, Hindustan Times, Indian Express, etc.

```
https://timesofindia.indiatimes.com/india
https://indianexpress.com/section/india
https://www.hindustantimes.com/india-news
https://news.google.com/news/headlines?hl=en-IN&gl=IN&ned=in
https://www.thehindu.com/news/national
https://www.thehindu.com/news/national
https://www.deccanchronicle.com/nation
https://mumbaimirror.indiatimes.com
https://economictimes.indiatimes.com/news/politics-nation
```

Fig 4.1 Various sources for collecting news headlines

The Following fig 4.2 shows the loading of data in CSV form using the panda's library. The dataset contains several entities/fields such as date, news headlines, and the source.

	Headline	Datasource	Date
0	Train services resume in Kashmir	Times Of India	2019-11-12
1	Sonia Gandhi speaks to Sharad Pawar, authorise...	Times Of India	2019-11-12
2	Congress leader Shivakumar hospitalised due to...	Times Of India	2019-11-12
3	Let us resolve to follow Guru Nanak's ideals o...	Times Of India	2019-11-12
4	LJP to contest 50 seats in Jharkhand: Chirag P...	Times Of India	2019-11-12
5	Maharashtra govt formation: Will speak to Cong...	Times Of India	2019-11-12
6	Maharashtra logjam: Senior Cong leaders defer...	Times Of India	2019-11-12
7	Will not form government in Maharashtra: BJP	Times Of India	2019-11-12
8	Ram Janmbhoomi-Babri Masjid: SC to pronounce v...	Times Of India	2019-11-12
9	Sidhu granted clearance for Kartarpur ceremony	Times Of India	2019-11-12
10	Yogi govt sacks 7 'corrupt' officials	Times Of India	2019-11-12

Fig 4.2 Loading of data using panda's library

The following figure 4.3 shows that the Similarities of News Headlines.

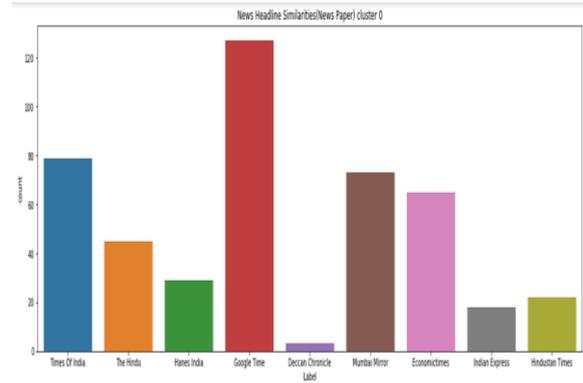


Fig 4.3 News Headlines Similarities

The following fig 4.4 shows the visual representation of positive and negative sentiment.

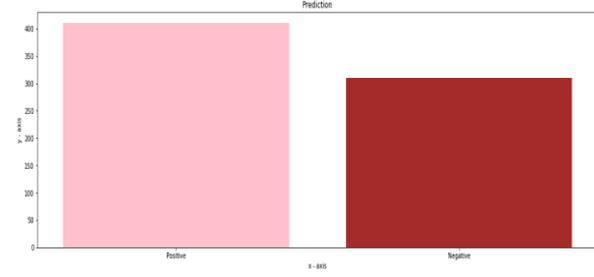


Fig 4.4 Visualization of positive and negative sentiment

#### V. CONCLUSION

In this project, the sentiment classification of news headlines in terms of positive, negative, and neutral using Recurrent Neural Network has been conducted. The datasets were scraped from the different sources using python library. The datasets undergo the preprocessing stages, feature extraction, and find the news headlines similarities. Then apply the recurrent neural network for sentiment classification. At last, the performance analysis can be done by estimate the accuracy of the classification algorithm.

#### REFERENCES

1. Apoorv Agarwal and Geeta Sikka, "Opinion Mining of News Headlines using SentiWordNet," IEEE 2016.
2. Somanath S. Chavan, "Sentiment Classification of News Headlines on India in the US Newspaper: Semantic orientation approach vs. Machine Learning," M.Sc Research project, Data analytics, 2019.
3. Soonth Taj, Baby Bakhtawer Shaikh, "Sentiment Analysis of News Articles: A lexicon-based approach, International Conference on Computing Mathematics and Technology, 2019
4. B.Meyer, M.Bikdash and X.Dai, " Fine-grained financial news sentiment analysis," IEEE 2017.
5. Ankita Gupta, Jyothika Pruthi, Neha Sahu, "Sentimental analysis of Tweets using Machine Learning approach," International Journal of Computer Science and Mobile Computing, Vol 6, Issue 4, April 2017.
6. Xing Fang and Justin Zhan, "Sentiment analysis using Product review data," Springer 2015.
7. J.Lei, Y.Rao, Q.Li, X.Quanand L.Wenyin, "Towards building a social emotion detection system for online news," Elsevier 2014.
8. Kranti Ghag, "SentiTFIDF- Sentiment Classification using Relative Term FrequencyInverse Document Frequency," International Journal of Advanced Computer Science and Application 2014.

9. Thirupathi Rao and Sai Balakrishna Allamsetty, "Implementation of Sentiment analysis on Twitter data, International Journal of Pure and Applied Mathematics, Volume 116 No 5, 2017.
10. Shuhaida Mohamed Shuhidan and SaidatulRahah Hamidi, "Sentiment Analysis for Financial News headlines using Machine Learning Algorithm" Springer 2018.
11. Islam, Ashraf, Abir, "Polarity Detection of Online News Articles based on sentence structure and dynamic dictionary, IEEE 2017.
12. Kiran, Shrinivas, Doddi, Dr.Mrs.Y.V.Haribhakta, Dr.Parag, Kulkarni, "Sentiment Classification of News Articles," "International Journal of Computer Science and Information Technology" Vol 5, 2014.
13. Rajkumar, S.Jagdale, S.N.Deshmukh "Document Level Sentiment Analysis from News Articles," IEEE 2017.
14. V.Kharde and P.Sonawane, "Sentiment analysis of twitter data: a survey of techniques," International Journal of Computer Applications, Vol 139, No 11, 2016.
15. Tan, Phang, W.S.Chin, K.O. Anthony, "Rule-based Sentiment analysis for financial news," International Conference on Systems, Man and Cybernetics, IEEE 2015.
16. Younis, E.M, "Sentiment analysis and Text mining for social media microblogs using open source tools" International Journal of Computer Applications 2015.
17. Raina, Prashant, "Sentiment analysis in News articles using Sentic Computing," IEEE 2013.
18. Doddi, Kiran S, "Sentiment Classification of News Article," Diss.College of Engineering Pune, 2014.
19. Padmaia S, "Comparing and Evaluating the Sentiment on Newspaper articles," "Science and Information Conference, IEEE 2014.
20. Haseena Rahmath P, Tanvir Ahmad, "Fuzzy based Sentiment analysis of product reviews using Machine Learning Techniques," International Journal of Computer Applications, Volume 99, No 17, 2014.

## AUTHORS PROFILE



**Prakashini S** is a PG Scholar in the Department of Computer Science and Engineering at National Engineering College, Tamil Nadu, India. She did her BE in Computer Science and Engineering from Dr. Sivanthi Aditanar College of Engineering, Tamil Nadu during 2018.



**Vijayakumar.D** is working as an Assistant Professor (Sr. Grade) in the Department of Computer Science and Engineering at National Engineering College, Tamil Nadu, India. He completed BE (CSE) from Manonmanium Sundaranar University and MS (By research) at Anna University Chennai. He has published seven International Journals and three National Journals.