IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 9, September 2020

DOI 10.17148/IJIREEICE.2020.8905

# Machine Learning Based News Validation

**Reetodeep Hazra[1], Megha Banerjee[1], Judhajit Sanyal[2]**

Student, Department of Electronics and Communication Engineering, Techno International New Town, Kolkata, India[1]

Assistant Professor, Department of Electronics and Communication Engineering,

Techno International New Town, Kolkata, India[2]

**Abstract:** The use of machine learning has become widespread in recent years, especially due to the impact of media content on the general population. In particular, the validation of truth in the context of news has become a critical necessity, due to its ready availability from verified and unverified sources, and its ability to influence people majorly. The present work outlines a LSTM (long short-term memory) based approach to news validation. The results obtained by the model are presented in terms of the training data set and contrasted with the results obtained from the test data set. Appreciable accuracy is achieved through the model, seen through the corresponding loss curves and confusion matrix.

**Keywords:** Machine Learning, LSTM, News Validation, Loss Curves, Confusion Matrix.

## I. INTRODUCTION

In recent years, the focus of a significant amount of research has been on the application of Machine Learning (ML) techniques in gauging social interaction as well as the impact of media on society. The multitude of ML techniques available allow for in-depth analysis of different types of data, among which techniques such as Naïve Bayesian Analysis, Spline Regression, Support Vector Machines (SVM) and neural network-based methods are quite popular. A model which has recently shown promise in the field of sentiment analysis and truth validation is Long-Short Term Memory (LSTM) based neural network. The work presented here outlines the performance of an LSTM network in fake news detection.

The paper is arranged in the following manner. Section II presents a survey on the different approaches employed by researchers. Section III presents the proposed model. The results obtained by application of the proposed model are outlined and the appropriate discussions are presented, in Section IV. Section V concludes the paper.

## II. LITERATURE SURVEY

Efforts have been made by researchers in recent times to map user sentiment with respect to different types of media, with one recent crowdsourcing based approach presenting a user approval estimation model for different types of web series from minimal data [1]. In a similar manner user behaviour prediction in terms of attrition probability of employees of a company has been illustrated in [2], using Naïve Bayesian estimation. A regression spline-based estimation technique of customer spending score has been presented in [3].

In recent years, one of the greatest challenges of the populace has been the identification of news items as fake. The spread of fake news and hoaxes in recent years also has its roots in the socio-political unrest sweeping over the world today, and hence many scholars have considered this a critical problem that needs to be addressed immediately. Some of the approaches to fake news detection in recent times have relied on SVM analysis for natural language processing [4], Naïve Bayes classification of news items for validation [5], and machine and user based multi-validation model [6]. One of the most interesting approaches in recent times has been the application of social and content-based models for fake news detection, as resented in [7], where real-world testing of the model has provided promising results.

## III. PROPOSED MODEL

The stacked LSTM network has a visible layer of input, three hidden layers and an output layer with Rectified linear circuit (ReLu) activation unit that predicts single value or binary classification. An LSTM layer requires three-dimensional input and by default delivers a two-dimensional output. In this classifier, the primary hidden layer is an embedding layer with a input sequence length of 300. The input dimension characterizes the size of the vocabulary in the text information. Output dimension was set as 100. It defines the size of the output vectors from this layer for each word. Output of the embedding layer was a 2-dimensional vector with one embedding for each word in the input sequence of

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 9, September 2020

DOI 10.17148/IJIREEICE.2020.8905

word. The default activation function sigmoid is utilized for the LSTM blocks. The classifier was trained for 10 epochs with a batch size of 256 with adam optimizer. The operational information of the network is given in Table 1 below.

Table 1. LSTM Network Information

| Layer | Type | Output shape | Units |
|---|---|---|---|
| 1 | Input | (1,300, 100) | - |
| 2 | Embedding | (1, 300, 128) | 128 |
| 3 | LSTM 1 | (1, 64) | 64 |
| 4 | LSTM 2 | (1, 32) | 32 |
| 5 | Sigmoid | 1 | - |

The distribution of news categories on which the model was applied is shown in Figure 1 below.

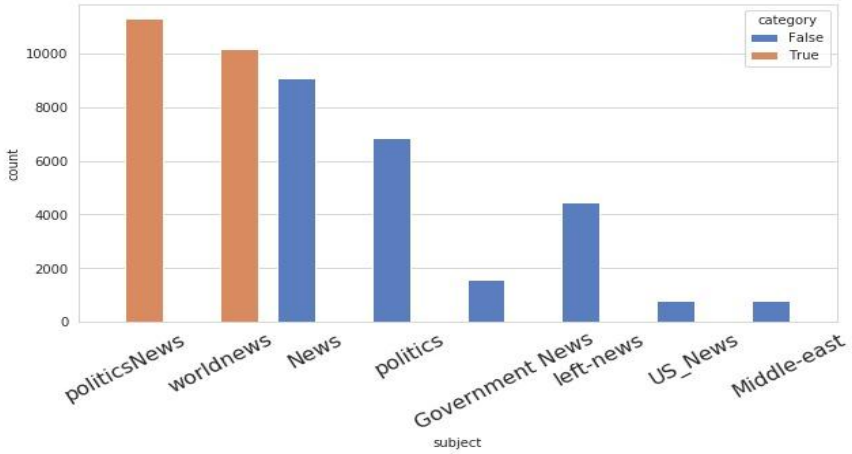


Figure 1. News Categories

The corresponding distributions of news items (true and fake) for the training dataset are shown in Figure 2.

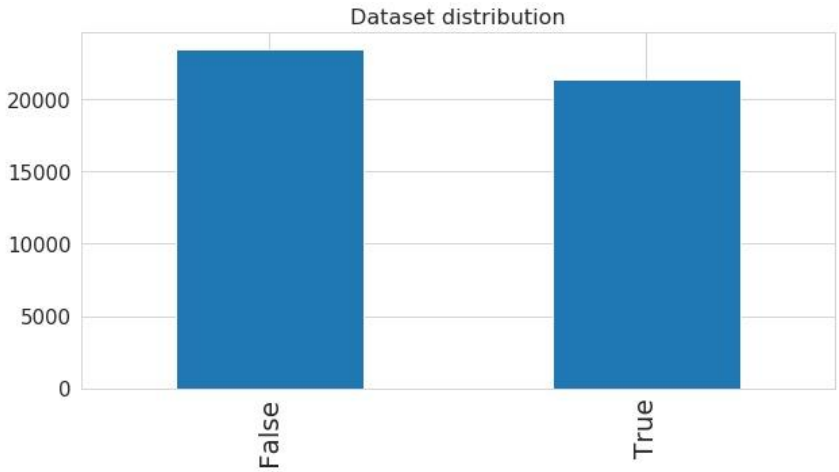


Figure 2. Dataset Distribution

## IV. RESULTS AND DISCUSSIONS

The LSTM model is trained and tested on the dataset selected. The corresponding loss curves and accuracy are shown in Figure 3 and Figure 4, respectively.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 9, September 2020

DOI 10.17148/IJIREEICE.2020.8905

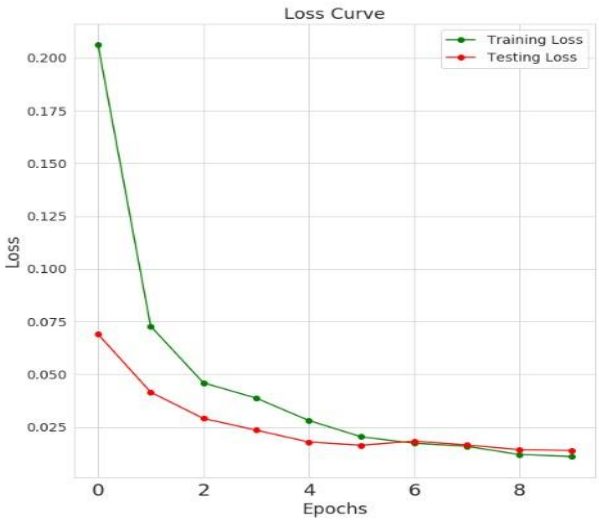


Figure 3. Loss Curves

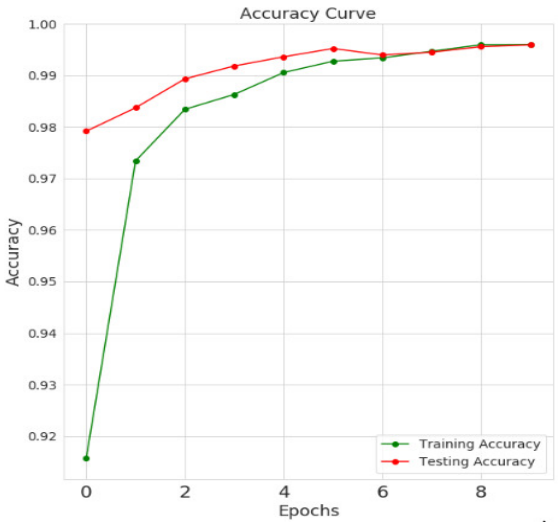


Figure 4. Accuracy

The confusion matrix for the testing dataset is presented in Figure 5. The confusion matrix shows a high degree of accuracy for the proposed LSTM model.


Figure 5. Confusion Matrix

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 8, Issue 9, September 2020

DOI 10.17148/IJIREEICE.2020.8905

From the output obtained, it can clearly be seen that the proposed model is effective in distinguishing fake news from real news, to an acceptable degree.

## V. CONCLUSION

As an extension to the present model, the authors intend to investigate the application of Kohonen maps for reducing high dimensionality often observed in news datasets. Another alternative is to investigate the applicability of Convolutional Neural Networks (CNNs) for news validation.

## ACKNOWLEDGMENT

The authors acknowledge Techno International New Town for providing the facilities and resources without which this research would not have been possible.

## REFERENCES

[1]. A. N. Ray and J. Sanyal, "Media Content Regulation using Crowdsourcing," 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-4.
[2]. A. N. Ray and J. Sanyal, "Machine Learning Based Attrition Prediction," 2019 Global Conference for Advancement in Technology (GCAT), BANGALURU, India, 2019, pp. 1-4.
[3]. P. Sharma, A. Chakraborty and J. Sanyal, "Machine Learning based Prediction of Customer Spending Score," 2019 Global Conference for Advancement in Technology (GCAT), BANGALURU, India, 2019, pp. 1-4.
[4]. A. Jain, A. Shakya, H. Khatter and A. K. Gupta, "A smart System for Fake News Detection Using Machine Learning," 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), GHAZIABAD, India, 2019, pp. 1-4.
[5]. A. Jain and A. Kasbe, "Fake News Detection," 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, 2018, pp. 1-5.
[6]. S. Khan, T. Khan, C. Prasad, A. Khatri and I. Khan, "Intelligent News Aggregator and Validator," 2019 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 2019, pp. 1-5.
[7]. M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro and L. de Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals," 2018 22nd Conference of Open Innovations Association (FRUCT), Jyvaskyla, 2018, pp. 272-279.