NIET Journal of Engineering & Technology (NIETJET)
Volume 10, Issue summer 2022 ISSN: 2229-5828 (Print)

Fake News Detection using Deep Learning

Mansi Pandey [1], Mayank Kumar[1]
Dhananjay Singh[2], Anshuman Singh[2], Pavan Kumar Shukla[2], Vinod M. Kapse[2]
[1]Student, Department of Electronics and Communication Engineering, Noida Institute of Engineering and Technology, Greater Noida
[2]Faculty, Department of Electronics and Communication Engineering, Noida Institute of Engineering and Technology, Greater Noida

Abstract: *Detection of fake news based on deep learning techniques is a major issue used to mislead people. For the experiments, several types of datasets, models, and methodologies have been used to detect fake news. Also, most of the datasets contain text id, tweets id, and user-based id and user-based features. To get the proper results and accuracy various models like CNN (Convolution neural network), DEEP CNN, and LSTM (Long short-term memory) are used*

Keywords: Deep Learning, Neural Network, long short-term memory

1. Introduction

There is a need to find a way to detect this fake news. Machine learning classifiers are used for a variety of purposes and can also be used to detect fake news. The classifier is first trained on a dataset called the training dataset. These classifiers can then automatically detect fake news. The main goal of this research work is to detect misleading messages that follow classical text-based taxonomies. A machine learning model should be prepared to classify messages into two broad categories: "true" and "fake" messages. A text zone must be specified during implementation. News or URLs can be placed in a text zone to check authenticity. The proposed machine learning model provides message correctness. As a result, the user is informed of the nature of the message and the accuracy percentage of the new data. The simple meaning of fake news is to absorb misleading information. Today, fake news spreads like water, and people share such information without confirming it. This is often done to promote or impose a particular idea and often involves a political agenda. Media companies need the ability to attract viewers to their websites to generate online advertising revenue. Therefore, you need to be aware of fake news. So that, we can't get trapped in any malicious or any kind of activity which is harmful to us or for society as well as for nature also.

2. RELATED WORK AND METHODOLOGY

In recent years, active research has been conducted to detect fake news. However, most of the work in this area focuses on the idea of investigating and publishing hoaxes on social media, the main distribution channel. Examples of the above are [5] or [6], where features examine the probability that a given message is false. Use classic machine learning techniques such as likes, followers, and posts (ranking trees, SVMs, etc.).

Some approximations [7] provided the best results, 93% were found by Clickbait News. Other works in collaboration with users who share messages, for example [8]. Content sharing is a way to stop content sharing to mitigate potential spoofing. The general trend is to analyze the spread of hoaxes, but other variations focus on analyzing hoaxes. News content starts to appear. [9] Thus, in addition to the user's ability to share messages, text. The difference between fake news. On the other hand, in [10] the statements of the articles are examined to identify false facts.Fabula.ai, recently acquired by Twitter, uses the latest deep learning algorithms. How to achieve this, considering both news content and features extracted from social networks. The AUC result was 0.93. [4] compares the performance of different algorithms (classical and deep learning) and new algorithms. In the categories "true" and "false" we received results with an accuracy of 95%. Finally, we propose and test a method based on a convolutional neural network using only message content [13]. Fake news with headlines and headline images returned 92% accurate results.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

Machine learning is a branch of computer science that improves the efficiency of a system repeatedly performing tasks using data rather than being explicitly programmed by a programmer. The two of machine learning are shown in Fig 1. Also, understand the differences between three machine learning methods:
Supervised learning, Unsupervised learning, and Reinforcement learning.

A. Supervised learning

It is a part of machine learning in which the machines are trained with the help of already labeled data, on that basis that result is predicted by the machines. Labeled data means some input data that is already tagged with the correct output. A supervised algorithm aims to find a mapping function to map the input variable with the output variable. The real implementation of supervised learning is in Risk assessment, Image classification, fraud detection, etc.

B. Unsupervised learning

Unsupervised data is the method in which the machines are not supervised using a training dataset. In it, the model itself identifies the hidden patterns and insights from the data. Unsupervised learning cannot be directly applied because we have the input data but no corresponding output data. The purpose of unsupervised learning is to find the underlying structure of the dataset, and the group that data according to dissimilarities represent that dataset in a compressed format.

C. Reinforcement learning

Reinforcement learning differs from supervised learning in a way that in supervised learning the answer key is in the training data so the model is only trained on correct answers whereas in reinforcement learning there are no answers and the reinforcement decides. Must be run to complete the task set. Without a training dataset, you will necessarily have to learn from experience.

Deep learning is the advanced version of machine learning, the machine learning algorithms, have a finite capacity to learn and understand things no matter the amount of data they acquire, and deep learning has higher performance with access to data. They get more experience through deep learning, it's then going to busing for many specific work/tasks like detecting diseases, driving cars, identifying faults, and so on. Deep learning learns from the complex structures or patterns that are available in the data. For binary classification, dataset of fake and real news is used such as Multinomial Naïve Bayes and, Passive Aggressive Classifier. Two datasets that are provided are the test dataset and the training dataset. Test data and train datasets are later matched to know the accuracy which is found using Binary classification. In this way, fake or real news is determined. By delivering maximum accuracy, fabricated news is then identified. Common English words are useful for data pruning.

LSTM Based Architectures The LSTM block consists of several gates that are responsible for maintaining the state of the hidden block. This eliminates the problem of leaky gradients and allows more information to be stored. Go beyond simple repeating units. Ambiguous words such as bank, mouse, and book indicate that modeling requires context. A network diagram is shown in Figure 2. As you can see, it consists of two different merged branches. One is responsible for the title of the article, the other for the content. Later in the network, these statements are combined into one of two possible categories of] true or false.

Convolution Based Architectures CNN has become one of the most successful areas of object detection and classification in deep learning, gaining widespread recognition and popularity in many fields, including medicine. CNN can be defined as a neural network system derived from a biological model of a visual system. Here, the individual's neurons respond to overlapping areas and are reliably inserted into the imaging system. It works on the principle of discrimination and applies similar neurons to the front layer region to achieve translational symmetry. This process changes the CNN to position, size, orientation, etc. Protect it from changes. CNN's structure allows you to customize hierarchical functionality. Usually, there are several hidden layers, and most are convolutional layers. A seven-layer CNN architecture is shown in Fig 3. These hierarchies are followed by an activation or pooling layer. This CNN architecture first tells about the role of the convolutional layer, which is used to extract the various features from the input images. In this layer mathematical operation of convolution is performed. The convolutional layer in CNN passes the result to the next layer once applying the convolutional operation in the input. This layer is so beneficial that it ensures the spatial relationship between the pixels is intact. Second, comes the Pooling layer which is used to decrease the size of the convolved feature map to reduce the computational costs. It summarizes the features generated by a convolutional layer. If we talk about the Fully connected (FC) layer, it consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. In this, the input image from the previous layers is flattened and fed to the FC layer. The reason two layers are connected is that two fully connected layers will perform better than a single

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

connected layer. These layers in CNN reduce human supervision. Usually, overfitting in the training dataset is caused by which a negative impact on the model's performance arises when all the features are connected to the FC layer. For overcoming this situation, a dropout layer is utilized wherein a few neurons are dropped from the neural network during the training process which results in a size reduction of the model. Lastly comes the role of Activation Function which is one of the most important parts of the CNN model. Mainly they are used to learn and approximate any kind of continuous and complex relationship between variables of the network. It adds non-linearity to the network. The use of the activation function is to determine which neuron should be activated or not. It decides whether the input to the work is important or not to predict using mathematical operations.

3. INDENTATIONS AND EQUATIONS

The result proposed by the model is shown in the section. Also, it's a comparison with other generation models. Model performance is evaluated using the confusion matrix in terms of various performance metrics such as accuracy, recall, F1 score, specificity, and sensitivity. All these indicators are defined as:

Accuracy: Accuracy is defined as the correctness of the predictions. It is calculated as the ratio of the number of correct predictions to the total number of samples.

$$\text{Accuracy} = \frac{\text{Correct predictions}}{\text{Total number of samples}} \qquad (1)$$

Precision: It refers to the positive prediction done by the model of any specific task. The more the precision more the correct output the user will get.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \qquad (2)$$

Recall: It calculates the number of positives labeled or captured by our model by labeling it as a true positive. It is expressed by the below formula:

$$\text{Recall Rate} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \qquad (3)$$

F1 score: The F1 score is defined as the harmonic mean between precision and recall. It is used as a statistical measure to rate the performance of the model. It is defined as follows:

$$\text{F1 score} = 2 \times \frac{\text{Precision} * \text{Recall Rate}}{\text{Precision} + \text{Recall Rate}} \qquad (4)$$

Specificity: Specificity is important for confirming or excluding the correctness percentage of the result during screening & is expressed as:

$$\text{Specificity} = \frac{TN}{TN + FP} * 100 \qquad (5)$$

Sensitivity: Sensitivity is defined as the ratio of true positives to the total number of images. Simply it tells how well the machine learning model can detect the positive instances. It is expressed as:

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} * 100 \qquad (6)$$

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

True Positive Rate (TPR) and False Positive Rate (FPR) can be calculated as follows:

$$\text{True Positive Rate} = \frac{\text{True Positive}}{\text{TP + FN}} * 100\% \quad (7)$$

$$\text{False Positive Rate} = \frac{\text{True positive}}{\text{TP + TN}} * 100\% \quad (8)$$

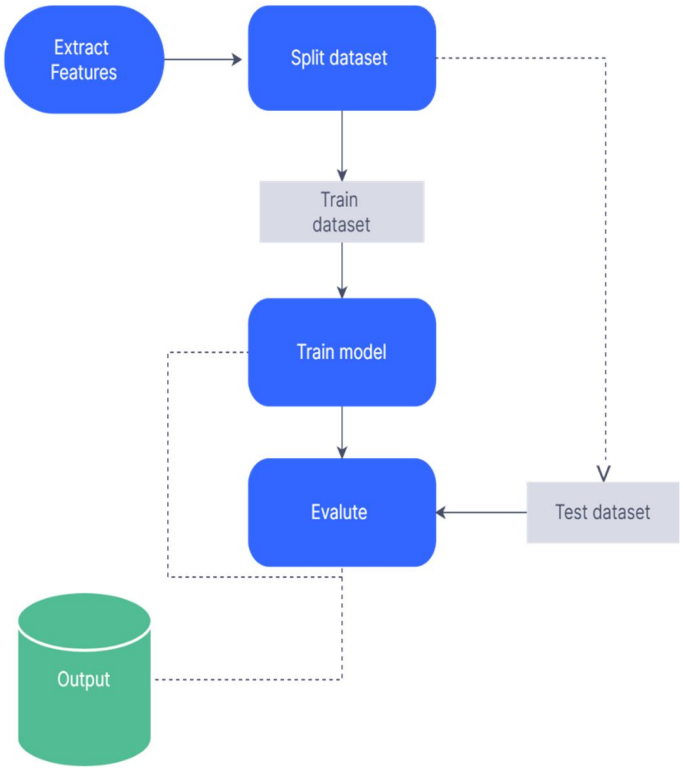

Fig 1: Workflow of Machine learning.

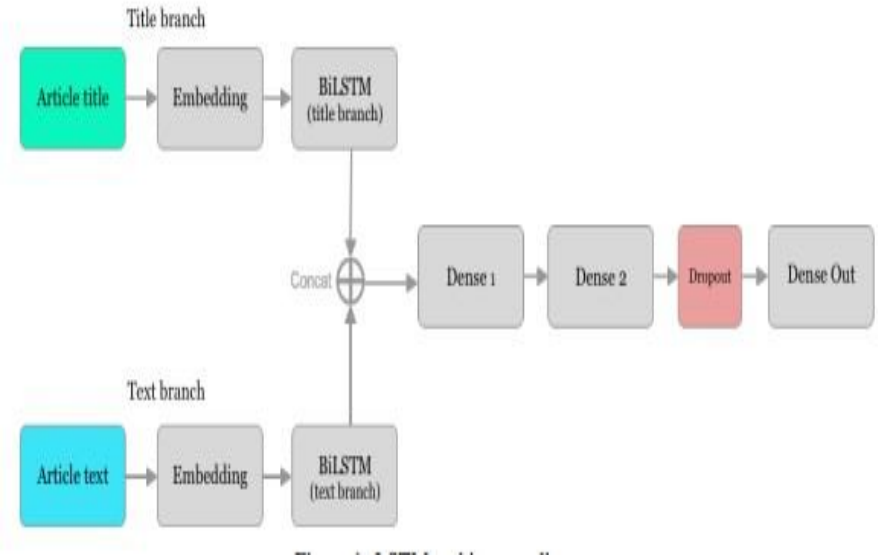

Fig 2: LSTM architechture workflow.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

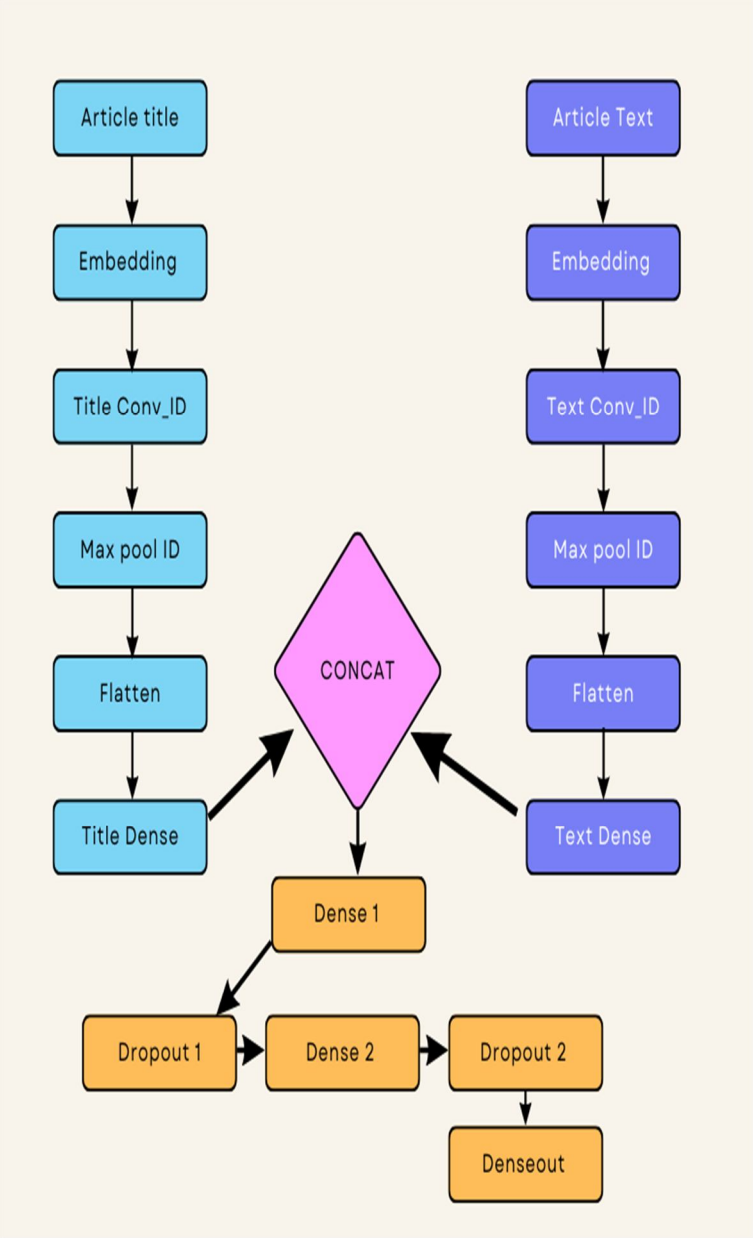

Fig 3: 7 Layer CNN architechture.

4. CONCLUSIONS

In this work, we have proposed two novel architectures applied to textual analysis for fake news detection. The experience gained in developing these models allows us to confirm that the use of deep learning models for this task has the potential to benefit a wide range of stakeholders, from social media companies to end users, to overcome the growing frustration of the Internet. The performance of the models solely depends on the architecture and the complexity of the model depending on the dataset and its size. The proposed model achieved satisfactory results with accuracy, precision, and recall rate of 98.1%, 99.3, and 98 respectively showing the result of detecting fake news.

REFERENCES

[1] Nguyen Vo and Kumin Lee. The rise of guardians: fact-checking url recommendation to combat fake news, 2018.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.

[2] Abhijan Chakraborty, Bhargavi Paranjape, Sourya Kakarla, and Niloy Ganguly. Stop clickbait: detecting and preventing clickbait in online news media. Pages 9–16, 08 2016
[3] Kyle Shaffer Svitlana Volkova. Separating facts from fiction: linguistic models to classify suspiciously and trusted news posts on twitter, 2018.
[4] S. Sasikala, M. Bharathi, B. R. Sowmiya. 2018, 'lung cancer detection and classification using deep CNN', international journal of innovative technology and exploring engineering,
[5] Federico Monti, Fabrizio Frasca, Davide Eymard, Damon Mannion, fake news detection on social media using geometric deep learning, 2019.
[6] Hands-on on machine learning with ticket learn and tensor flow by Geron Aurelien 2017.
[7] Machine Learning for dumbbies by John Mueller 2016 edition.

Publisher: Noida Institute of Engineering & Technology,
19, Knowledge Park-II, Institutional Area, Greater Noida (UP), India.