International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Fake News Identification and Stance Detection using CNN-LSTM

Himangi Srivastava[1], Prathmesh Bole[2], Radha Zade[3]

[1]*Computer Engineering, PES Modern College of Engineering, Pune*
[2]*Computer Engineering, PES Modern College of Engineering, Pune*
[3]*Computer Engineering, PES Modern College of Engineering, Pune*

Abstract – This research paper investigates the use of a combination of CNN and LSTM models for identifying fake news and stance detection. Fake news involves analyzing a piece of news article and checking whether the news contains fake or misleading information. Stance detection refers to a writers' emotion whether he is for or against a particular topic. Stance detection is a complex task which involves NLP and Machine Learning techniques. In the CNN-LSTM model, Firstly- Input layer receives the preprocessed text data which is then passed to the next layer. Secondly, CNN layer, which is used to extract important features from the text. Lastly, the LSTM layer which captures sequential structure of the text. The output of this layer is finally fed to the output layer for binary classification. The performance of the model is evaluated using several metrics like accuracy, precision, recall, and F1 score. Our experimental result demonstrates that the model is highly effective in detecting fake news using both CNN and LSTM. Overall, our research provides important insights into the problem and how the model has the potential to make a significant impact on major issues like "Fake News". It also has important implications for addressing the challenges of fake news.

Key Words: Fake News, Stance Detection, CNN, LSTM, NLP, Machine learning.

1.INTRODUCTION

In the era of social media and the Internet, News is generated in various ways. Earlier people used to generate news using traditional journalism which mainly involved gathering information, research, conducting interviews and seminars, extracting and analysing data to create original news content. But with the widespread of social-media, news is easily accessible and available to everyone, and it's hard to detect whether the news provided is authentic or fake.

Fake News is a significant challenge in the modern information age. People these days intentionally present false information through social media and it is presented in such a way that it would appear as true and authentic to the general public. Fake news has many severe consequences like undermining the trust of the media in front of the general public, inciting violence, spreading hate speech etc.

The first step towards detecting and combating the spread of fake news is identifying it.

For this, we have used the dataset which is related to the US elections 2020. The dataset contains national-level data for the elections. The model is trained on a dataset of news articles labelled as either "fake" or "real". This helps the model to learn to distinguish between the two types of articles. The second step is the stance detection method which is another critical aspect of addressing the fake news problem. Stance detection helps in finding whether the author is for, or against a particular topic or idea. In summary, it is important to address the issues of fake news identification and stance detection to ensure the integrity and authenticity of news articles or social-media posts.

In this Deep learning project, we are going to combine both CNN and LSTM models to identify fake news and check its stance towards a particular topic.

2. CNN and LSTM Models

Both Convolutional Neural Networks (CNN) and Long Short-term Memory (LSTM) are two popular types of Deep learning models commonly used in NLP tasks such as fake news identification as well as stance detection

CNNs are typically used for feature extraction processes. They are mainly used for images, but
can also be used for text classification.

LSTM, on the other hand, helps in capturing the sequential structure of the text and remembers important information from earlier parts of the text.

With the combination of both CNN and LSTM, the model can make the prediction more accurately and can also improve the performance of the model by leveraging the strengths of both types of neural networks.

Dataset used- The Dataset has been obtained from kaggle and it is related to the text-based news articles and social-media posts circulated during the US elections 2020. (fig 1)

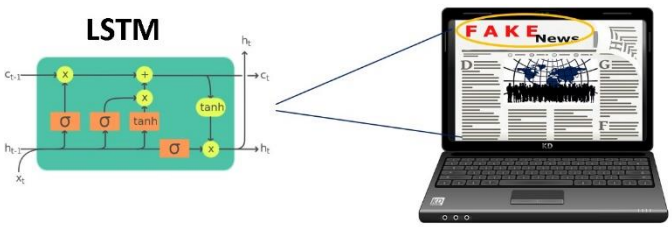

Figure -1: Figure 1

The Algorithmic approach used here involves several key steps-

1- Data Collection
2- Data Preprocessing
3- Feature Extraction
4- Model training
5- Model Evaluation

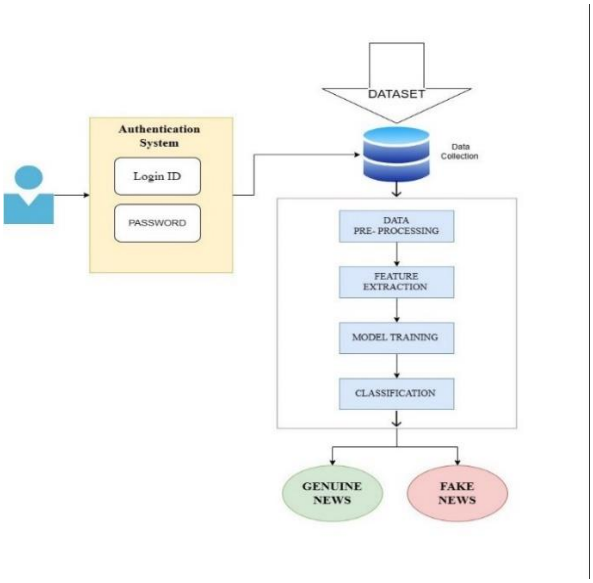

Figure 2 - System Architecture

Fig 2 above is the system architecture diagram of the model and the first step is Data acquisition i.e. we need to collect news articles and their corresponding labels (real or fake) from a reliable source. Second step is Data Preprocessing where we need to clean and pre-process the data by performing tasks by tokenization, stemming, and feature extraction. The third step is to train the CNN-LSTM model using the pre-processed data to predict the binary classification (real or fake) and the stance (positive, negative or neutral) towards a particular topic. The last step is developing an application that provides a user-friendly interface for users to input data and view the resultant output. The predicted result is in the form of binary classification which involves classifying data into one of the two categories. (0 or 1)

The CNN-LSTM model combines CNNs and LSTMs to analyze both the local and global context of the input data. CNN component extracts relevant features such as words or phrases while the LSTM helps in modelling the sequential dependencies between these features.

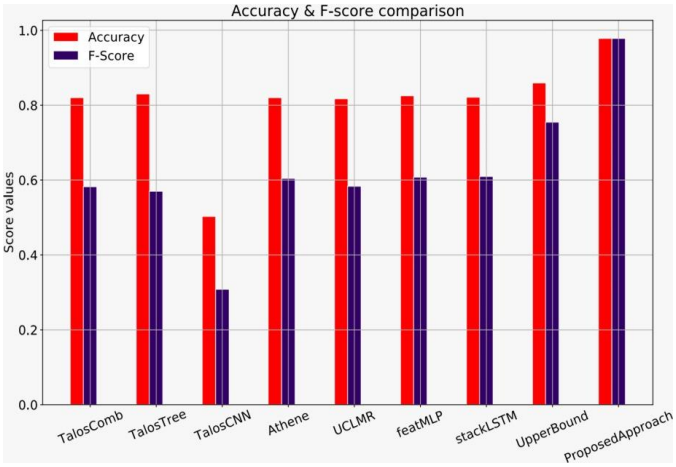

Figure 3- Accuracy and F-score comparison

The performance of the model is evaluated using several performance metrics such as accuracy, precision, F1 score etc. Accuracy is responsible for measuring the overall correctness of the models predictions, while precision and recall measures the model's ability to correctly identify positive and negative instances.
The F1 score is a harmonic mean of precision and recall and it is a commonly used performance metric in binary classification problems. Accuracy alone won't be enough to evaluate the performance of the model accurately. F1 score takes into account of both precision and accuracy and is helpful in providing a more balanced evaluation of the model's performance.
Fig 3 above depicts a bar graph which shows a Graphical representation between Accuracy and F-score.

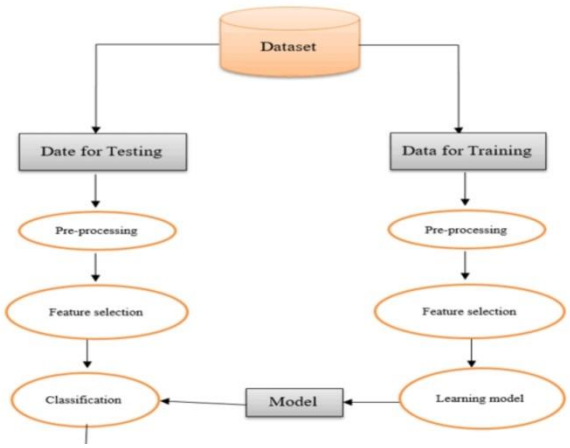

Figure 4- Data for testing and training

Training and testing the model is important to ensure that the model is accurate and effective in its predictions. The process of training and testing the model involves a combination of data preprocessing, feature selection, Model training, Model evaluation, classification etc.
It's important to train and test the model because it helps in improving model's accuracy, avoiding overfitting, and increasing the robustness of the model.

3. CNN-LSTM and State-of-the-art Methods

CNN and LSTM are considered more effective than state-of-the-art methods in this task due to their ability to capture both local and global features in text data.
For e.g.- In stance detection tasks, a CNN-LSTM model can determine the stance of particular topic by analyzing the local features of individual words and phrases and the global features of the entire article or social media post.
State-of-the-art methods are traditional methods which often rely on handcrafted features or pre-defined models. CNN-LSTM models are able to learn complex features and patterns directly from the data which makes them flexible as well as adaptable to different types of text data. They are also effective in identifying patterns and features that may be difficult for pre-defined patterns to capture.

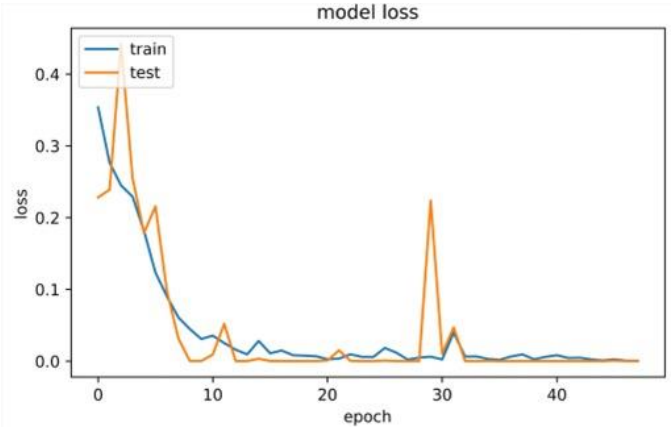

Figure 5a- Loss graph

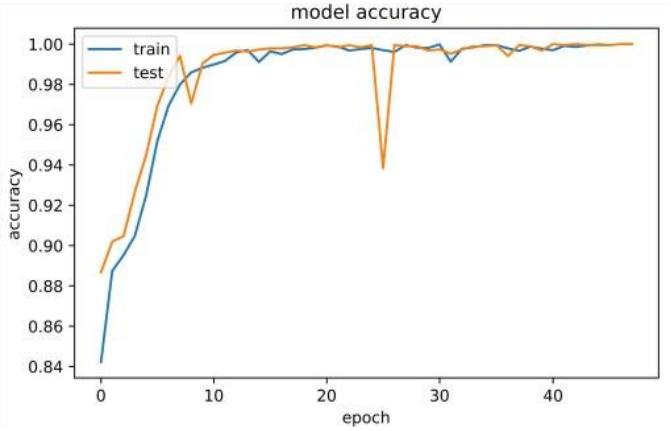

Figure 5b- Accuracy graph

Fig 5a represents the model loss and epoch graph representation.
Model loss and epoch are commonly used to evaluate the performance of the model during training. Model loss is responsible for measuring how well the model is able to predict the labels of the training data. During the training, the model tries to adjust the parameters to minimize its loss. The lower the loss, the better the models prediction and its performance.

Fig 5b represents the model accuracy and epoch graph representation.
Model accuracy is a measure of how accurately a model can predict its labels for the training data.
During the training, the model attempts to maximize its accuracy by adjusting the parameters. In the training process, the accuracy should increase because it is an indication that the model's performance is also increasing.
Therefore, both the epochs and model accuracy are important factors in evaluating the performance of the model. An ideal model should have an increasing accuracy curve and a decreased loss curve as the number of epoch increases.

3. CONCLUSION

In conclusion, this research paper demonstrates the use of combination of CNN-LSTM models and their effectiveness in improving the accuracy and performance of the model. By training the model on large datasets the model can provide high accuracy in distinguishing between the fake and authentic news articles. Overall, our research paper provides important insights into the problem and how the combination of both the models have the potential to make significant impact on issues like fake news, its challenges and how important it is to promote more responsible consumption of news in this digital era.

Overall, our research paper highlights the importance of using advanced Machine learning and Deep learning algorithms and techniques to address the challenges of issues like fake news and stance detection and to promote more responsible consumption of media and to make informed decision-making.

ACKNOWLEDGEMENT

We would like to express our gratitude to all those who have contributed to this research paper on fake news identification and stance detection using CNN-LSTM.

First and foremost, we would like to thank our guide **Prof. Mrs. Silkesha Thigale** for providing us with invaluable guidance and support throughout the research process. Their expertise and feedback have been instrumental in shaping the direction and scope of this project.

Lastly, we are grateful to our friends and family for their encouragement and understanding during the research process. Their support has been crucial in helping us stay motivated and focused throughout the project.

REFERENCES

1. T. Mihaylov, G. Georgiev, and P. Nakov, ‘‘Finding opinion manipulation trolls in news community forums,’’ in *Proc. 19th Conf. Comput.Natural Lang. Learn.*, Beijing, China, Jul. 2015, pp. 310–314. [Online]. Available: https://www.aclweb.org/anthology/K15-1032

2. S. Vosoughi, D. Roy, and S. Aral, ‘‘The spread of true and false news online,’’ *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.

3. Chopra, ‘‘Towards automatic identification of fake news: Headlinearticle stance detection with LSTM attention models,’’ Stanford Univ., Stanford, CA, USA, Tech. Rep., 2017.

4. K. Popat, S. Mukherjee, J. Strötgen, and G. Weikum, ‘‘Where the truth lies: Explaining the credibility of emerging claims on the Web and social media,’’ in *Proc. 26th Int. Conf. World Wide Web Companion*, Apr. 2017, pp. 1003–1012