IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.57
Volume 4, Issue 1, July 2024

A Complete System for Identifying False News Through Image Classification and Natural Language Processing

Sunit Jana, Rakhi Biswas, Rahul Kumar, Koushik Pal, Devanshu Dev, Pritam Mukherjee, Md Sufiyan Azam, Anirban Ghosal*
Department of Electronics & Communication Engineering
Guru Nanak Institute of Technology, Kolkata, India
*JIS College of Engineering , Kolkata

Abstract: *Confusion and false information are spread by fake news. This study employs machine learning and deep learning methods based on text and image data to identify bogus news. We classify news as real or fake using deep learning models and Natural Language Processing (NLP) techniques. Furthermore, the authenticity of images can be confirmed by image-based classification that makes use of pre-trained models. Our findings demonstrate the efficacy of integrating text and picture analysis, demonstrating encouraging accuracy in detecting bogus news*

Keywords: Natural language processing (NLP), text categorization, picture verification, machine learning, deep learning, hybrid detection system, and fake news detection

I. INTRODUCTION

The proliferation of fake news has become a serious problem due to the growing usage of social media and digital platforms. Misinformation or incorrect content masquerading as news is known as "fake news," and it has the power to sway public opinion, damage reputations, and cause confusion. Since incorrect information spreads so quickly, it is now essential to create automated systems that can effectively identify fake news.

Manual fact-checking is a time-consuming and labor-intensive part of traditional false news detection techniques. However, machine learning and deep learning models offer a scalable and automated method of detecting fake news thanks to developments in artificial intelligence (AI). These models are able to identify the validity of text and images by analyzing patterns in them.

In this study, we categorize news stories as true or fraudulent using a combination of deep learning and machine learning techniques. Deep learning models are used to detect fake news based on images, while Natural Language Processing (NLP) techniques are used to analyze the text data. Our strategy seeks to increase the precision of false news identification by combining text and image analysis.

II. MATERIALS AND METHODS

A. The dataset

We conducted our investigation using publicly accessible datasets that included both authentic and fraudulent news pieces. There are two primary parts to the dataset:

1. Dataset Based on Text:

• Two CSV files from the Fake and True News Dataset were used by us:
• True.csv, which contains actual news articles; • phony.csv, which contains phony news articles
• Title, text, subject, and publication date are among the columns found in every dataset.
• The dataset underwent preprocessing, which included handling missing values, eliminating duplicates, and cleaning the text (removing special characters, punctuation, and stopwords).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.57
Volume 4, Issue 1, July 2024

2. Dataset Based on Images:

• We gathered both authentic and fraudulent news photos from a variety of sources, such as:
• Google Images
• Fact-checking websites such as Politifact and Snopes.
• We used the inceptionv3 model to extract picture features and Google picture Search to validate authenticity in order to increase classification accuracy.

B. Text-Based Identification of False News

We take the following actions in order to categorize news stories according to their text:
1. Preparation
All text should be converted to lowercase.
• Eliminate digits, punctuation, and special characters.
The text should be tokenized into individual words.
Stop words like "the," "is," and "and" should be eliminated.
2. Extraction of Features
• To transform text into numerical vectors that represent word associations, we employ Word2Vec, a word embedding technique.
• In a continuous space, every word is represented by a dense vector.
3. Model Training To handle the text sequences, we employ an LSTM (Long Short-Term Memory) model, a kind of recurrent neural network (RNN).
• The LSTM model evaluates phrase patterns and contextual meaning to categorize an article as real or fraudulent.
• To improve accuracy, we use the Adam optimizer and binary cross-entropy loss when training the model on labeled data.

C. Detecting Fake News Using Images

Images that are edited or deceptive can occasionally be seen in fake news pieces. We use deep learning methods to evaluate images:
1. Extraction of Features
High-level features are extracted from photos using InceptionV3, a pre-trained convolutional neural network (CNN). The model analyzes images by performing normalization and scaling them to 299x299 pixels.
2. Classification To ascertain if an image is authentic or fraudulent, the features that have been retrieved are fed into a logistic regression classifier.
• We use Google Image search to confirm images that have been tagged as fraudulent.
3. Verification of Google Image Search
We use a reverse image search to increase accuracy; an image is more likely to be authentic if it shows up in several reliable news sources.
• The image is deemed fraudulent if it is not found in trustworthy sources or shows up in deceptive situations.

D. Evaluation of the Outcome

1. Accuracy: Indicates how accurately the model predicts outcomes overall.
2. Precision: Ascertains the proportion of the anticipated fake news items that were, in fact, fraudulent.
3. Recall: Evaluates the model's ability to recognize real fake news.
4. F1-Score: A balanced performance metric that combines precision and recall in a harmonic manner.

III. RESULTS AND DISCUSSION

A. Text-Based Identification of False News

The LSTM model did a good job of differentiating between authentic and fraudulent news stories after being trained on textual data. With a high F1-score of 98% and an accuracy of almost 98%, the model demonstrated a good trade-off between recall and precision.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.57
Volume 4, Issue 1, July 2024

	precision	recall	f1-score	support
0	0.99	0.97	0.98	5899
1	0.97	0.99	0.98	5326
accuracy			0.98	11225
macro avg	0.98	0.98	0.98	11225
weighted avg	0.98	0.98	0.98	11225

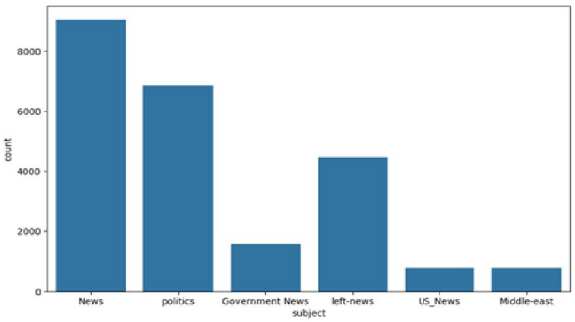

Fig-1: categories of news

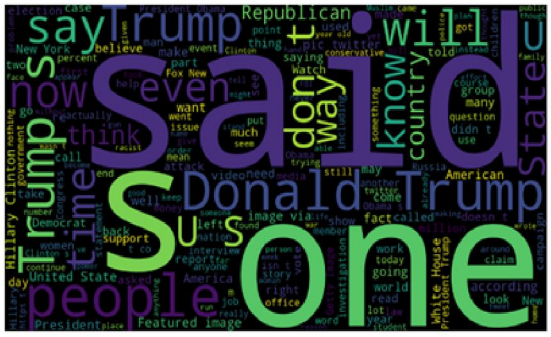

Fig-2: world cloud (visual representation of text data where the most frequently occurring words appear larger)

While the majority of fake and legitimate news pieces were accurately recognized, there were some misclassifications, according to the confusion matrix.

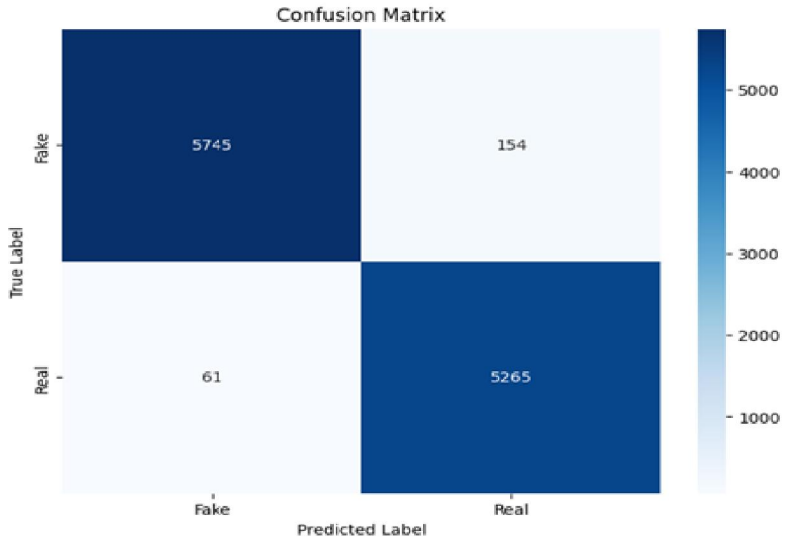

Fig-3: Confusion matrix

Key Findings:

• When language patterns and writing styles diverged considerably from authentic news, the model was able to identify phony news with high accuracy.

• Because they resembled actual news, humorous or ambiguous news items were sometimes misclassified.

• Longer articles improved categorization accuracy by fostering greater contextual awareness.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.57
Volume 4, Issue 1, July 2024

B. Detecting Fake News Using Images

The InceptionV3 model with logistic regression achieved 87% accuracy in classifying fake news based on images. The model was able to recognize conflicting visual features and photos that had been altered. However, there were issues with certain deepfake photos and deceptive actual photos that were used in the wrong situations.

Important Results:

• The model did a good job of identifying photos that had been manipulated, but it had trouble figuring out why real photos were used.

• By cross-referencing picture sources, the incorporation of Google picture Search for verification reduced misclassification.

• The system's ability to identify minute changes in deepfake-generated images was limited.

C. Difficulties and Restrictions

Even with high accuracy, there are still certain difficulties:

• Contextual Misinterpretation: In certain cases, when the language seemed reliable, the text model was unable to identify false information.

• Deepfake Images: The image model had trouble telling the difference between artificial intelligence-generated and real deepfake content.

• Dataset Bias: Because the models were trained using certain datasets, they might not be broadly applicable to all types of fake news.

IV. CONCLUSION

In this study, we showed how well deep learning and machine learning methods work for text and image-based false news detection. Our method obtained great accuracy in differentiating between fake and authentic news by using a pre-trained InceptionV3 model for picture verification and Natural Language Processing (NLP) for text analysis. While the image verification model assisted in identifying modified or deceptive photos, the LSTM-based text classifier successfully identified trends in bogus news stories.

The findings demonstrated that integrating text and picture analysis offers a more reliable method of detecting false news than depending just on one technique. There are still issues, though, such identifying context-based false information and dealing with adversarial instances, where fake news is made to evade detection algorithms.

To increase detection accuracy, future research can investigate multimodal deep learning models that incorporate text, visuals, and even source credibility. Adding more varied sources to the dataset can help improve the model's generalizability.

One promising weapon in the fight against false information is AI-driven fake news identification. It is a useful tool in the battle against the dissemination of misleading information in the digital era, even though it is not a perfect answer.

V. ACKNOWLEDGMENT

We would like to thank Koushik Pal, our teacher, for all of his help and support during the project. We are also grateful to GNIT for providing the platform and resources needed to carry out this study.

Furthermore, we would like to thank the open-source community for contributing the tools and datasets that enabled our study. We would especially want to thank the academics and engineers who helped create the computer vision, deep learning, and natural language processing (NLP) frameworks used in this study.

We also value the assistance provided by scholarly and internet sources that aided in our comprehension of different fake news detecting methods. Finally, we recognize the contribution of fact-checking groups, whose efforts continue to spur improvements in the identification of false information.

REFERENCES

[1]. Goodfellow, I., Bengio, Y., & Courville, A.: Deep Learning. MIT Press (2016).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.57
Volume 4, Issue 1, July 2024

[2]. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J.: Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781 (2013).

[3]. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. Rethinking the Inception Architecture for Computer Vision. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016).

[4]. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

[5]. Zhang, X., Wang, S., & Liu, H. (2018). Detecting Fake News with Content-Based and Social Context Features. Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI).