Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.4, May : 2024

FAKE DETECT: A DEEP LEARNING ENSEMBLE MODEL FOR FAKE NEWS DETECTION

Sujatha Karri, Venkata Pavani Boppuri, Mahalakshmi Venigalla, Satish Kedari Computer Science and Engineering, JNTUK University Andhra Loyola Institute of Engineering and Technology Vijayawada, Andhra Pradesh, India sujathakarri701@gmail.com boppurivenktapavani@gmail.com dmahalv@gmail.com sathishasree@gmail.com

Abstract— In an age characterized by an abundance of information, the proliferation of false information presents a considerable challenge to public discourse and the credibility of the media. This initiative seeks to create a resilient system for detecting fake news through the utilization of a comprehensive deep learning ensemble model. By harnessing the capabilities of various deep learning architectures, the ensemble model aims to improve the precision and dependability of identifying fake news

Keywords—: Fake News, Deep Learning

I. INTRODUCTION

In an era dominated by the rapid dissemination of information through diverse digital channels, the surge in fake news poses a pervasive threat to the integrity of public discourse. The ability to differentiate between genuine and deceptive news articles is essential for upholding trust in the media and facilitating well-informed decision-making in society. The project, named "Fake Detect," strives to confront this challenge by harnessing the potency of deep learning to construct an ensemble model designed for the detection of fake news.

Background: The advent of social media and online news platforms has democratized access to information, but it has also facilitated the widespread circulation of misinformation. Fake news, often cloaked in a veneer of truth, can wield influence over public opinion, shape narratives, and even impact political landscapes. Traditional fact-checking and verification methods struggle to keep pace with the sheer volume and speed at which information spreads online.

Motivation: The impetus behind "Fake Detect" stems from the pressing need to develop advanced and effective tools capable of discerning between authentic and fabricated news stories. Leveraging deep learning, particularly through the creation of an ensemble model, presents a promising avenue for elevating the accuracy and reliability of fake news detection. By amalgamating the strengths of various deep learning architectures, the project seeks to establish a robust system adaptable to the diverse and evolving nature of fake news.

Dataset Compilation: Assemble a diverse and properly labelled dataset of news articles, ensuring a range of topics, sources, and writing styles to facilitate comprehensive training and evaluation of the fake news detection model.

Model Implementation: Develop and train distinct deep learning models, encompassing recurrent neural networks (RNNs), long short-term memory networks (LSTMs), convolutional neural networks (CNNs), and transformer models. Each model will contribute a unique perspective to enhance the overall robustness of the fake news detection system.

Ensemble Model Creation: Formulate an intelligent ensemble model that judiciously combines the predictions of the individual deep learning models. This collaborative approach aims to leverage the

collective strengths of each model, enhancing the overall predictive performance of the fake news detection system.

Performance Assessment: Evaluate the effectiveness of the ensemble model using a comprehensive set of evaluation metrics. Conduct a thorough comparison with the individual models to gauge the improvement achieved through the collaborative ensemble approach.

II. LITERATURE SURVEY

The surge in digital media and the prevalence of online platforms have fuelled the spread of fake news, eroding the trustworthiness of information sources. The field of fake news detection has garnered substantial attention, with researchers exploring various methodologies, including traditional rule-based approaches and, more recently, harnessing the potency of deep learning. This literature survey offers an insight into pivotal studies and techniques within the fake news detection domain, emphasizing the role of deep learning ensemble models.

Rule-Based Approaches

Early attempts in fake news detection leaned heavily on rule-based methods, employing predefined criteria to identify deceptive content. While these approaches exhibited some success, they grappled with challenges in adapting to the dynamic nature of fake news and were constrained in handling nuanced linguistic patterns.

Individual Deep Learning Models

Recurrent Neural Networks (RNNs)

Numerous studies have delved into the utilization of Recurrent Neural Networks (RNNs) to model sequential data for fake news detection. Despite the promising results of RNNs in capturing temporal dependencies, challenges surfaced in handling long-range dependencies.

Long Short-Term Memory Networks (LSTMs)

LSTM networks, an evolution of RNNs designed to tackle the vanishing gradient problem, have been deployed to model sequential patterns in textual data. Research suggests enhanced performance compared to traditional RNNs, particularly in capturing contextual information.

Convolutional Neural Networks (CNNs)

Widely acknowledged for their triumph in image classification, Convolutional Neural Networks (CNNs) have been adapted for text-based tasks, including fake news detection. CNNs excel in capturing local features and have demonstrated effectiveness in identifying subtle linguistic cues indicative of misinformation.

Transformers

Transformer models, exemplified by BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), have revolutionized natural language processing tasks. These models, pre-trained on extensive corpora, exhibit a profound understanding of contextual information and have been fine-tuned for fake news detection with considerable success.

III. METHODOLOGY

Data Collection: Identify and compile a diverse dataset of news articles labelled as "fake" or "real." Preprocess the text data, incorporating tokenization, stemming, and the removal of stop words.

Feature Extraction: Transform the preprocessed text data into numerical features using techniques such as TF-IDF or word embeddings.

Individual Models: Implement and train various deep learning models, encompassing RNNs, LSTMs, CNNs, and Transformers.Fine-tune hyperparameters for optimal performance.

Ensemble Model: Develop an ensemble model that aggregates predictions from individual models. Experiment with ensemble techniques such as averaging, weighted averaging, and majority voting.

Evaluation: Assess the ensemble model on a distinct testing dataset, utilizing metrics such as accuracy, precision, recall, and F1 score.

Averaging Techniques Ensemble models, which amalgamate predictions from multiple models, have gained widespread acceptance. Averaging techniques involve merging output probabilities, resulting in a more steady and dependable prediction by lessening the impact of individual model biases.

Weighted Averaging Weighted averaging assigns varying weights to the predictions of individual models based on their performance. This adaptive approach enables the ensemble to assign more influence to models demonstrating higher accuracy in specific contexts.

Majority Voting Majority voting involves merging predictions based on the most prevalent classification among individual models. This uncomplicated approach often proves effective, particularly when dealing with diverse model architectures.

Evaluation Metrics Studies commonly utilize metrics such as accuracy, precision, recall, and F1 score to evaluate the performance of fake news detection models. Additionally, some researchers underscore the importance of interpretability and transparency in the decision-making process of the model.

IV. RESULTS & DISCUSSION

Upload Fake News Dataset: This module facilitates the uploading of datasets to the application. Once uploaded, the application reads all news and generates a WORD cloud graph.

Dataset Preprocessing: This module involves essential preprocessing steps. It removes stop words, applies stemming and lemmatization, and then transforms all textual news into numeric vectors. This conversion is achieved by computing the average frequency of each word.

Run Existing SVM Algorithm: In this module, the processed numeric vector is split into training and testing sets. The application employs 80% of the dataset for training the Support Vector Machine (SVM) model, and the remaining 20% is used for testing. The trained model is then applied to the test data to calculate prediction accuracy.

Run Proposed DL-BILSTM-GRU Algorithm: Similar to the SVM module, this section involves splitting the processed numeric vector into training and testing sets (80% and 20%, respectively). The application utilizes the Proposed DL-BILSTM-GRU Algorithm to train a model, and the trained model is then evaluated on the test data to calculate prediction accuracy.

Comparison Graph: This module generates an accuracy comparison graph between the existing SVM algorithm and the proposed DL-BILSTM-GRU algorithm. The graph visualizes the performance disparity between the two models.

Fake News Prediction from Test Data: In this module, users can input test news, and the Proposed DL-BILSTMGRU Algorithm analyses the provided data, predicting whether it is fake or real based on the trained model.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.4, May : 2024

V. RESULT ON PROPOSED SYSTEM

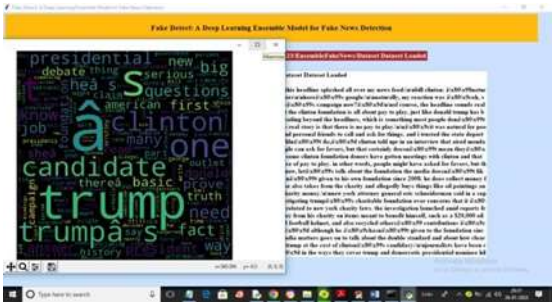

Fig.1.Word Cloud

In the text area above, the complete dataset of textual news is loaded, and the word cloud graph visually highlights words with higher frequencies by bolding them. To proceed, close the current graph, and then click the 'Dataset Preprocessing' button. This action will initiate the removal of stop words, application of stemming and lemmatization, and the conversion of all text data into a numeric vector. The resulting output will be displayed below.

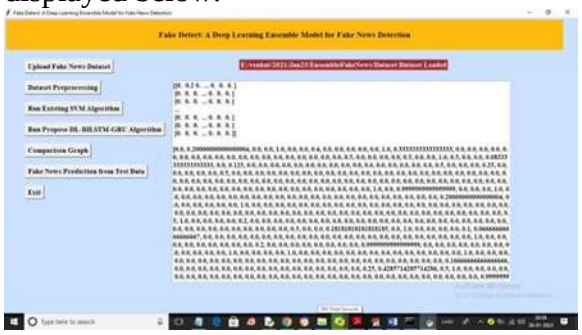

Fig.2.Numeric Vector

In the preceding screen, the entire dataset has been transformed into a numeric vector by replacing each word with its average frequency. To continue, click on the 'Run Existing SVM Algorithm' button. This action will trigger the training of the Support Vector Machine (SVM), and the resulting output will be displayed below.

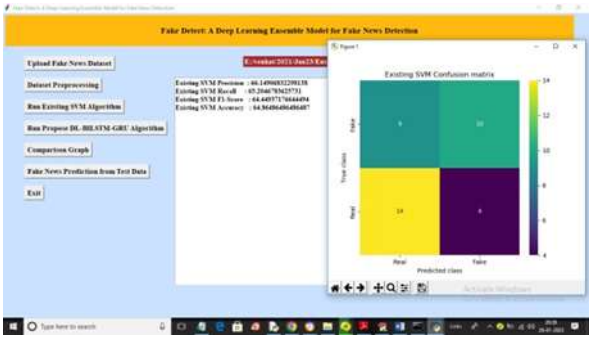

Fig.3.Confusion Matrix Graph

In the previous screen, the proposed algorithm achieved an accuracy of 94%. The confusion matrix graph shows that the blue boxes contain the count of incorrect predictions (only 2), while the green and yellow boxes in the diagonal represent correct predictions. To proceed, close the current graph and action will allow you to upload test data and obtain the corresponding output.

.

Fig.4.Prediction

In the screen above, following the '=' symbol, you can observe the TEST news. After the arrow symbol (=>), the predicted output is displayed, indicating whether the news is classified as 'Fake' or 'Real'.

VI.CONCLUSION

The literature survey underscores the progression of fake news detection methodologies, tracing their evolution from rule-based approaches to the current prominence of deep learning models. The survey places a particular emphasis on ensemble techniques. Drawing inspiration from the insights gleaned in existing literature, our approach in undertaking the "Fake Detect" project will be informed by the successes and challenges outlined. This informed strategy aims to guide the development of a cutting-edge deep learning ensemble model for the purpose of fake news detection.

VII. FUTURE WORK:

While considerable strides have been made, challenges persist in the realm of fake news detection. These challenges include the imperative for more extensive and diverse datasets, the mitigation of biases in model predictions, and the enhancement of interpretability in complex ensemble models. Future research endeavours may explore novel architectures and innovative techniques to fortify the robustness of detection systems in response to evolving strategies employed in the dissemination of misinformation.

VIII. REFERENCES

[1] 2021, https://datareportal.com/social-media-users Global Social Media Overview.
[2] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” Journal of Economic
[3] Perspectives, vol. 31, no. 2, pp. 211–236, 2017.
[4] M. G. Sherry Girgis and E. amer, “Deep learning algorithms for detecting fake news in online text,” in Proceedings of the ICCES, pp. 93–97, Cairo, Egypt, July 2018.
[5] W. Y. Wang, ““Liar, liar pants on fire”: a new benchmark dataset for fake news detection,” in Proceedings of the Annu. Meet. Assoc. Comput. Linguist, pp. 422–426, Vancouver, Canada, July 2017.
[6] 2020, https://www.kaggle.com/mrisdal/fake-news Fake or Real News.
[7] Y. Khan, M. T. I. Khondaker, A. Iqbal, and S. Afroz,
“A benchmark study on machine learning methods for fake news detection,” pp. 1–14, 2019, https://arxiv.org/abs/1905.04749.
[8] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, “Fake news detection through multi-perspective speaker profiles,” in Proceedings of the Eighth Int. Jt. Conf. Nat. Lang, pp. 252–256, Taipei, Taiwan, November 2017.

[9] F. A. Ozbay and B. Alatas, “A novel approach for detection of fake news on social media using metaheuristic optimization algorithms,” Elektron. IrElektrotechnika, vol. 25, no. 4, pp. 62–67, 2019.
[10] Y. Fang, J. Gao, C. Huang, H. Peng, and R. Wu, “Self multi-head attention-based convolutional neural networks for fake news detection,” PLoS One, vol. 14, no. 9, pp. 1–14, 2019.
[11] C. Janze and M. Risius, “Automatic detection of fake news on social media platforms,” in Proceedings of the 21st Pacific-Asia Conference on Information Systems, pp. 261– 276, Langkawi Island, Malaysia, July 2017.
[12] A. M. P. Braşoveanu and R. Andonie, “Integrating machine learning techniques in semantic fake news detection,” Neural Processing Letters, vol. 52, no. 2, 2020.
[13] T. C. Truong, Q. B. Diep, I. Zelinka, and R. Senkerik,
“Supervised classification methods for fake news identification,” in Proceedings of the ICAISC, pp. 445–454, Zakopane, Poland, July 2020.
[14] S. Hakak, M. Alazab, S. Khan, T. R. Gadekallu, P. K. R.
Maddikunta, and W. Z. Khan, “An ensemble machine learning approach through effective feature extraction to classify fake news,” Future Generation Computer Systems, vol. 117, pp. 47–58, 2021.
[15] N. C. Sahar Ghannay, B. Favre, and Y. Estève, “Word embeddings evaluation and combination,” in Proceedings of the Tenth International Conference on Language Resources and Evaluation, pp. 300–305, Portorož, Slovenia, May 2016.
[16] J. L. B. Diederik and P. Kingma, “ADAM: a method for stochastic optimization,” in 3rd International Conference on Learning Representations, San Diego, CA, USA, May 2015.
[17] 2021, https://fasttext.cc/docs/en/english-vectors.html “English word vectors