iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62143

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

AI Assisted Deceptive Content Analysis System Using Bi-LSTM

Negha Ajayaghosh[1], Richa Teresa Mohan[2], Sivanie Nejukumar[3], Basil Joy[4], Rotney Roy Meckamalil[5]
Department of Computer Science and Engineering, Mar Athanasius College of Engineering, Kothamangalam, Kerala

Abstract: ***In an age of widespread misinformation, creating accurate machine learning algorithms is vital for ensuring the integrity of information sharing. This project aims to develop a machine learning model capable of distinguishing between fake and true news articles. [1]The model utilizes a Bidirectional Long Short-Term Memory (Bi-LSTM) neural network architecture. Exploratory data analysis techniques are employed to gain in-sights into the characteristics and distributions within the dataset. Data visualization techniques aid in understanding patterns and relationships within the dataset. Additionally, unigram analysis is conducted to extract meaningful features from the text data. The datasets are then prepared for model training, involving preprocessing steps such as tokenization and vectorization. Fi-nally, the Bi-LSTM model is constructed, leveraging its ability to capture long-range dependencies in sequential data. The model is trained on the prepared datasets, optimized using appropri-ate techniques, and evaluated using metrics such as accuracy, precision, recall, and F1-score. The Bi-LSTM architecture offers the advantage of capturing long-range dependencies in sequential data, thereby enhancing the model's ability to discern nuanced patterns in news articles. The primary objective of this project is to develop a robust and accurate system for automatically detecting fake news, thus playing a pivotal role in enhancing the dissemination of reliable information in the digital age.***
Index Terms: ***Bidirectional Long Short Term Memory, Natu-ral Language Processing, cognitive abilities, Global Vectors for Word Representation , unigram and bigram analysis, machine learning.***

I. INTRODUCTION

Within the widespread propagation of misinformation, the pressing need for the development of accurate machine learning algorithms is a ray of hope in the face of information credibility[12]. This necessity is showcased by a project, called "AI Assisted Deceptive Content Analysis Using Bi-LSTM," that aims to create a complicated model with the ability to differentiate between true and fake news articles. Based on Bidirectional Long Short-Term Memory (Bi- LSTM) neural network architecture known for capturing intricate patterns in sequential data, this process takes on an extensive methodology throughseveral steps. It starts with getting a diverse dataset that includes genuine and deceptive from credible sources. The project goes ahead to employ various exploratory data analysis methods that can help in revealing the underlying characteristics and distributions embedded in the dataset. Moreover, using advanced data visualization approaches which can bring into light significant patterns and relationships, this sets ground for careful unigram analysis that will assist in extracting meaningful features from textual data itself.The preparation process is very important for the construction and optimization of datasets to be used in building the Bi-LSTM model that has been said to potentially strengthen the truthfulness of information flow in the digital era. This project aims at constructing and optimizing a Bi-LSTM model which can capture intricate long-range dependencies embedded within sequential data. This study tests whether this model can accurately identify patterns in news articles indicative of false contents using strict training and evaluation procedures. Although it possesses certain challenges like being time-consuming when training and requiring much more computer resources than other alternatives, no other approach can beat Biological –LSTM architecture in deciphering complex sequential data. Nevertheless, the project is committed to creating a solid reliable system able to automatically detect fake news so as to strengthen information dissemination integrity in modern times.By advancing state-of-the-art machine learning techniques, this initiative will promote a discernable and trustworthy information eco-system that safeguards against misinformation propagation across our ever-growing interconnected world.

II. RELATED WORKS

A. Overview on detecting Fake News Spreaders

Online social networks (OSNs) are a key element of societal digitalization. OSNs have the power to change people's opinions, actions, choices, and communication by exposing them to diverse popular trends in a variety of spheres of life. Malicious people and social bots are important sources of disinformation on social media and can be major cyber dangers to society.

©IJRASET: All Rights are Reserved

It is exceedingly challenging to distinguish between the user profiles of a criminal user disseminating bogus news and a cyber bot based solely on their differences.[2] Researchers have been trying to figure out how to lessen this issue for years. Identifying the propagators of false information on OSNs is still difficult, though.Based on several factors suggested in the unique taxonomy, researchers have presented a thorough evaluation of the state of the art techniques for identifying rogue users and bots. In order to assist researchers who are new to this topic, it has also been attempted to avoid the critical issue of false news identification by outlining numerous important hurdles and prospective future research fields.

B. Benign Identity Prediction Among Deceptive Accounts

Social networking services are seeing an exponential increase in the number of profiles; nevertheless, some users are hiding their identities for nefarious reasons. On the other hand, it has frequently been possible to accurately identify accounts that were generated by bots using machine learning techniques. In the current scenario, it is imperative to detect fraudulent accounts because the account holder's reviews in SMPs are used for decision-making in numerous industries. Effective benign accounts need to be forecasted in a way that produces high-quality results from this kind of study. The reviews that users post on Twitter are collected into a dataset, and the content that is tweeted—including personal details—is used to classify the reviews.[3] The techniques Support Vector Machine (SVM), Random Forest Classifier, Logistic Regression, XGBoost Classifier, and Decision Trees are used in a related method.

C. NLP Model Analysis

While current academics are focusing on social characteristics of the news, traditional detection approaches were centered around content analysis. A unique deep natural language processing (NLP) approach for artificial intelligence (AI)-assisted false news detection has been proposed[1]. The publisher layer, social media networking layer, enabled edge layer, and cloud layer are the four levels that define the suggested work. Four steps were completed in this work.

The steps involved in classifying news articles as false or real are as follows:

1) Data collecting;
2) Information retrieval (IR);
3) NLP-based data processing and feature extraction; and
4) A deep learning-based classification model that employs credibility scores of publishers, users, messages, headlines, and other factors. This suggested model outperforms other current techniques, with an average accuracy of 99.72% and an F1 score of 98.33%.

D. Toxic Fake News Detection and Classification

A study aimed to address this gap and identify toxic fake news to save time spent on examining nontoxic fake news. Multiple datasets from various online social networking sites, including Facebook and Twitter, were gathered in order to do this. The most recent samples were acquired through data collection based on the most popular keywords that were taken from the pre-existing datasets. Following the classification of the instances as toxic or nontoxic using toxicity analysis, a toxic-fake news detection and classification system was designed using conventional machine-learning (ML) techniques like linear support vector machine (SVM), conventional random forest (RF), and transformer-based ML techniques like bidirectional encoder representations from transformers (BERT). The linear SVM approach performed better than BERT SVM, RF, and BERT RF, according to the experiments. But NLP methods were found to have an even higher accuracy rate

E. Web Scraping for data collection

[6] Web scraping, sometimes referred to as data scraping, is the process of obtaining data from a certain website or web page. For data analysis, business data collection, real-time data collection, etc., data extraction is helpful. A lot of the data and information found on webpages is in an unstructured format, making it unusable for data analysis and other purposes. Thus, the data is transformed into a structured format after it has been extracted. While obtaining data from websites is beneficial, it is also prohibited. With restricted access, we are able to scrape publically available data. It is imperative to adhere to the policies of websites as some may expressly forbid or restrict scraping activity. Check the terms of service, privacy statement, and any other relevant documents before scraping data from websites.

F. A Comprehensive Review on deep learning methods

Deep learning-based algorithms are more accurate in identifying fake news than many machine learning techniques. Prior review articles primarily focused on data mining and machine learning methodologies, with little attention paid to deep learning approaches for identifying false news[8]. Nevertheless, prior studies have not included recently developed deep learning-based techniques like Attention, Generative Adversarial Networks, or Bidirectional Encoder Representations for Transformers. The explanation of the dataset and NLP approaches employed in earlier studies is carried over into this survey [5]. An extensive summary of deep learning-based approaches has been provided in order to classify representative techniques into different groups. Also covered are the well-known assessment measures in the field of fake news detection.

G. Hybrid BiLSTM-TCN Model with Attention Mechanism

Several AI and machine learning-based text detection techniques have been created in response to the spread of misinformation.[7] A study has been presented that offers a novel solution by combining attention mechanism, TCN, and Bi-LSTM. The model makes use of TCN for deep feature extraction, Glove Embedding for word representation, Bi-LSTM for capturing sequential text dependencies, and an attention method for selecting significant text pieces. Using multiple models on the publicly accessible ISOT dataset, empirical trials were conducted to evaluate the effectiveness of the model. These trials' very remarkable outcomes showed that the study project was successful.

III. PROPOSED MODEL

This work proposes a robust AI-assisted deceptive content analysis system designed to combat the spread of misinformation. The system leverages a trainable deep learning model for efficient and accurate classification of deceptive statements. Users can contribute to the system's knowledge base by providing labeled training data, consisting of statements with corresponding truth labels (true/false). Additionally, users can submit test samples for real-time analysis, allowing the system to assess its effectiveness on unseen content. To capture the nuances of language and meaning, the system utilizes pre-trained word embeddings, which represent words as numerical vectors. This allows the model to process the semantic relationships between words and identify patterns indicative of deception[9]. Before analysis, the system performs thorough data pre-processing, which may include tasks like text cleaning, normalization, and removal of irrelevant information. The core of the system lies in a Bi-LSTM classification model, a type of recurrent neural network (RNN) adept at handling sequential data like text.This model analyzes the pre-processed text, identifying linguistic features and stylistic choices often associated with deceptive content. By continuously learning from user-provided feedback on test sample labels[10], the system refines its classification ability, becoming progressively adept at distinguishing between truthful and deceptive statements. This user-centric approach fosters a collaborative environment, empowering users to contribute to a more reliable and trustworthy digital information landscape.

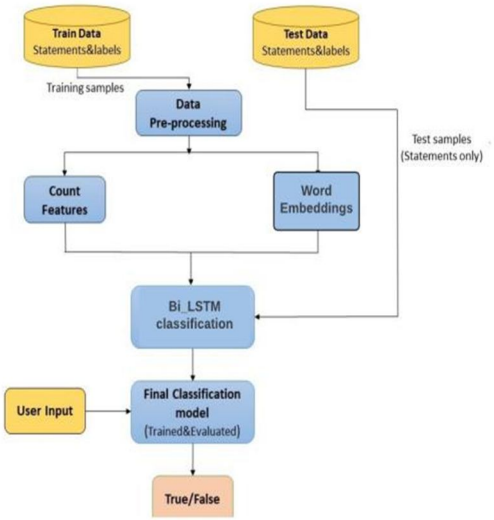


Fig. 2. Data Flow Diagram of Our Proposed Model

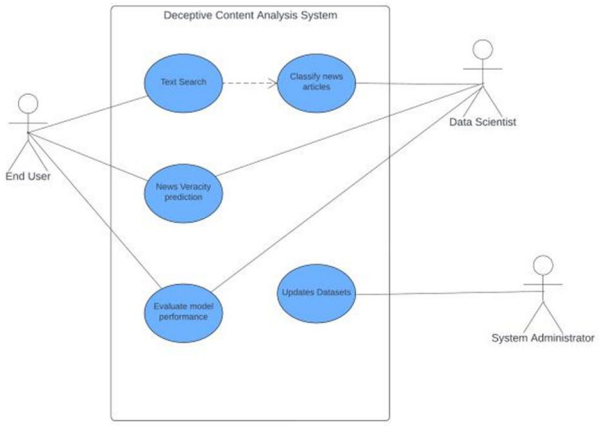


Fig. 1. Use Case Diagram of Our Proposed Model

A. Data Collection

The dataset for training and testing the Deceptive content analyzer is composed of a number of text samples classified as genuine or unauthentic. The sources that provided the data include social media sites, online fora, as well as news articles. Each sample gets a corresponding label either true or fake with respect to supervised learning. Moreover, [11]there was a pre-trained word embedding model referred to as GloVe Twitter which has been applied. It had been taught on numerous tweets. The models use the idea from how often two words appear together in Twitter account.By integrating dynamic datasets, the system can adapt to the evolving landscape of misinformation, providing timely and accurate assessments of news articles as they emerge. This enhancement will require the implementation of mechanisms for continuous data ingestion, processing, and model retraining to ensure the system remains current and effective.

Dataset	Source	Content
true.csv	Kaggle	21418 rows of true news
fake.csv	Kaggle	23503 rows of fake news
GloVe Twitter	Kaggle	pre-trained word embeddings

TABLE I
DATASET

B. Data Preprocessing

Data preprocessing, before the model is trained, underwent a number of pre-processing stages such as:
Tokenization: Breaking down text samples into individual words or subwords. Lowercasing: It entails converting all text to lower case so that analysis can be case-insensitive. Removal of special characters as well as punctuation marks: These are unnecessary symbols for analysis. Stopword removal: These are common words with no meaning. Text Embedding: Tokens are converted into numerical vectors using techniques like Word2Vec and GloVe. This captures semantic relationships between words.

C. Bi-LSTM Architecture

The Deceptive Content Analyzer utilizes a deep learning model, specifically a Bidirectional Long Short-Term Memory (Bi-LSTM) cells. The Bi-LSTM architecture is chosen for its ability to capture sequential dependencies in text data effectively both in the forward and reverse direction. Using the labeled dataset and the proper loss function and optimization technique, the model is trained. The model gains the ability to identify patterns and characteristics linked to legitimacy in text samples through training. Throughout the training process, the model's parameters are modified continuously through the dataset in order to reduce loss and increase accuracy

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Deceptive Content Analysis

In the domain of deceptive content analysis, our system represents a pioneering leap forward, harnessing the formidable capabilities of AI, particularly through the deployment of the Bi-LSTM architecture. Departing from conventional methodologies that often struggle to keep pace with the dynamic landscape of digital misinformation, our approach is distinguished by its adeptness in discerning subtle linguistic cues indicative of deceptive intent. By iteratively training on meticulously labeled datasets and continuously refining model parameters, our system acquires a nuanced understanding of the contextual nuances inherent in deceptive language. Furthermore,[12] the integration of pre-trained word embedding models like GloVe Twitter augments our system's capacity to grasp semantic intricacies, thereby fortifying its proficiency in detecting deceptive narratives across a spectrum of sources, spanning social media platforms to news articles. This fusion of cutting-edge AI techniques with robust data preprocessing mechanisms not only endows our Deceptive Content Analyzer with the adaptability to navigate the ever-evolving landscape of misinformation but also furnishes it with the agility to furnish timely and precise assessments, thereby serving as a potent bulwark against the proliferation of deceptive content in the digital sphere.

IV. RESULTS

The outcomes of our study illuminate the efficacy and practical implications of our AI-assisted deceptive content analysis system. We commence by delineating key performance metrics, elucidating the system's adoption rates and effectiveness in discerning deceptive content across various digital platforms. Subsequently, we conduct a thorough analysis of the system's performance metrics, encompassing precision, recall, and F1-score, across different categories of deceptive content. Through a combination of quantitative evaluation and qualitative feedback analysis,

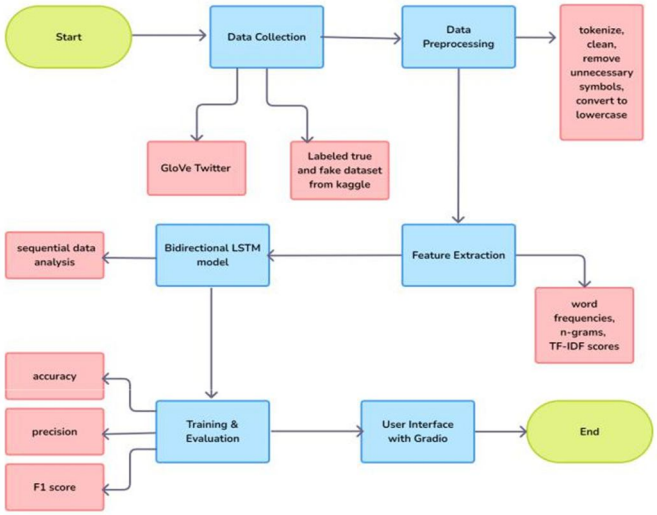


Fig. 3. Architecture of Deceptive content analysis System

we unearth invaluable insights into the system's ability to accurately identify deceptive narratives and mitigate the spread of misinformation. Moreover, by presenting the results in a clear and comprehensive manner, we aim to provide stakeholders with a holistic understanding of the system's efficacy in combating deceptive content and safeguarding the integrity of digital discourse.

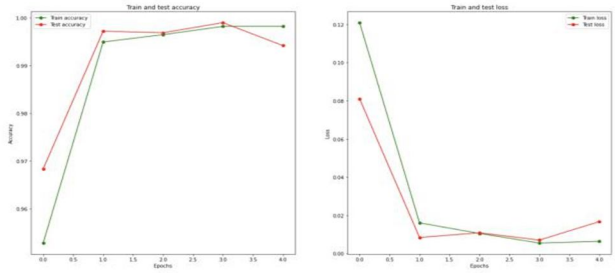


Fig. 4. Efficiency of Bi-LSTM model

V. FUTURE SCOPE

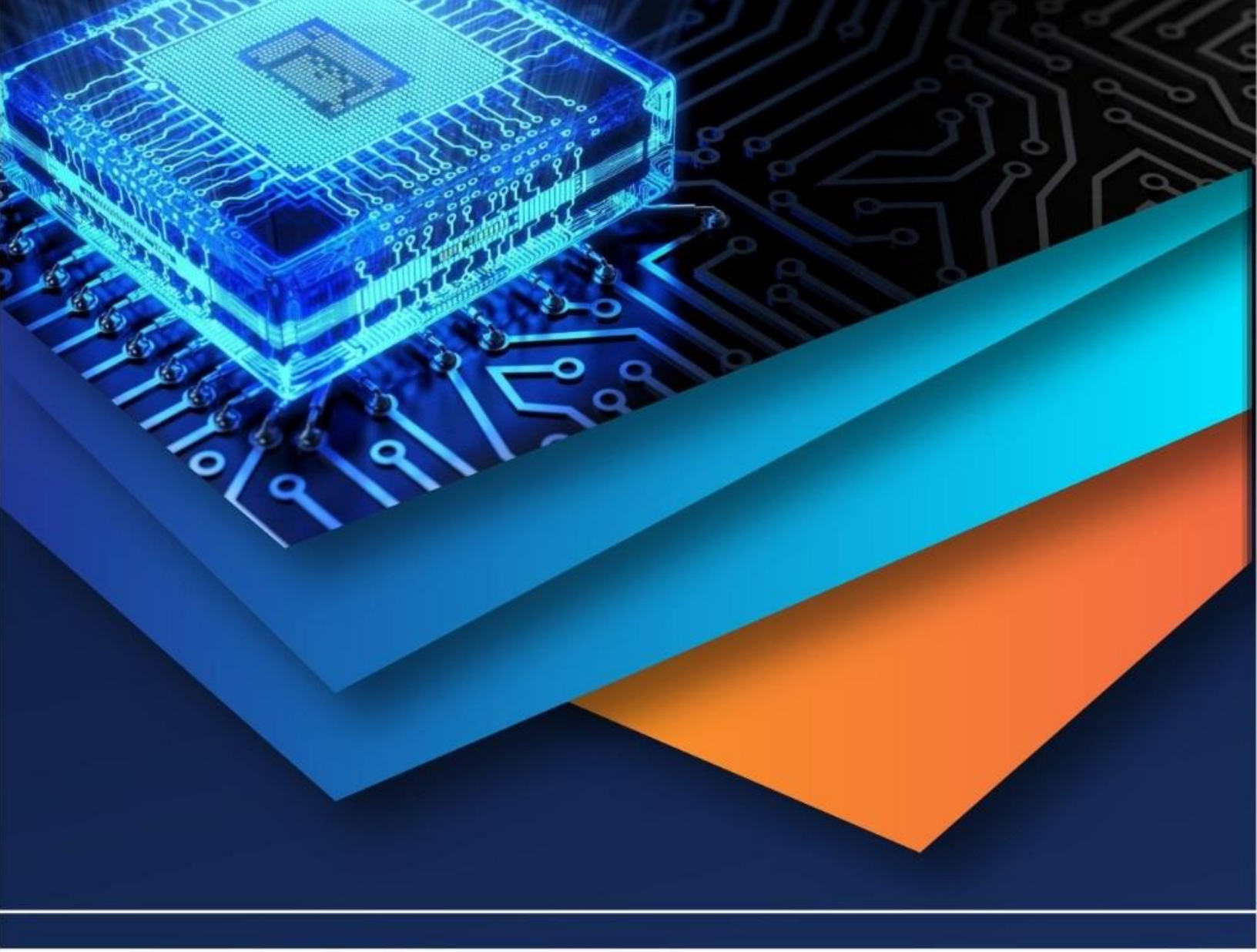
The future scope of the project is a shift from static datasets to dynamic ones in which live news feeds are included for analyzing and categorizing fake or real time content. Incorporating dynamic datasets however equips the system with the ability to adjust to the ever-changing landscapes of misinformation this will lead to timely and precise news article assessments as they are generated. In addition, this upgrade would require developing mechanisms for constant data ingestion, processing, and model retraining in order to maintain relevance and efficiency of such a system. Moreover, moving forward, the model's enhancements entail tackling some inherent challenges that come with working on dynamic data including but not limited to; changing language styles, topics or sources among others. Besides other ways like training from pre-trained models and multiple algorithms that could help enhance the ability of a model in a dynamic environment. Thus, embracing dynamic datasets while advancing model's intricacy is the aim of this project which tries to stay ahead in fighting falsehoods and also promoting evidence-based instant information availability.

VI. CONCLUSION

The project titled "AI Assisted Deceptive Content Analysis Using Bi-LSTM" takes on the challenge of misinformation by creating a machine learning model which can differentiate fake news articles from real ones. This project adopts a Bidirectional Long Short-Term Memory (Bi-LSTM) neural network architecture and goes through various stages like data loading, exploratory data analysis, and preprocessing. To tap into the Bi-LSTM model's ability to capture long-term dependencies in sequential data, the solution seeks to fully understand sensitive patterns within news articles.While it admits that the Bi-LSTM architecture draws much computational power, this study gives priority to the development of an efficient and reliable system for automatic recognition of fake news thereby contributing greatly to information credibility online.

REFERENCES

[1] G. G. Devarajan, S. M. Nagarajan, S. I. Amanullah, S. A. S. A. Mary and A. K. Bashir, "AI-Assisted Deep NLP-Based Approach for Prediction of Fake News From Social Media Users," in IEEE Transactions on Computational Social Systems, doi: 10.1109/TCSS.2023.3259480.
[2] W. Shahid, Y. Li, D. Staples, G. Amin, S. Hakak and A. Ghorbani, "Are You a Cyborg, Bot or Human?—A Survey on Detecting Fake News Spreaders," in IEEE Access, vol. 10, pp. 27069-27083, 2022, doi: 10.1109/ACCESS.2022.3157724.
[3] T. K, S. Ks, S. G, S. V. S and S. V, "Effective Benign Identity Prediction Among Deceptive Accounts using Machine Learning," 2023 Interna-tional Conference on Recent Advances in Science and Engineering Technology (ICRASET), B G NAGARA, India, 2023, pp. 1-6, doi: 10.1109/ICRASET59632.2023.10420028.
[4] M. A. Wani et al., "Toxic Fake News Detection and Classification for Combating COVID-19 Misinformation," in IEEE Transactions on Computational Social Systems, doi: 10.1109/TCSS.2023.3276764.
[5] M. F. Mridha, A. J. Keya, M. A. Hamid, M. M. Monowar and M. S. Rahman, "A Comprehensive Review on Fake News Detection With Deep Learning," in IEEE Access, vol. 9, pp. 156151-156170, 2021, doi: 10.1109/ACCESS.2021.3129329.
[6] C. Bhatt, Gaitri, D. Kumar, R. Chauhan, A. Vishvakarma and T. Singh, "Web Scraping: Huge Data Collection from Web," 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET), Ghaziabad, India, 2023, pp. 375-378, doi: 10.1109/ICSEIET58677.2023.10303037.
[7] M. B. Shikalgar and C. S. Arage, "Fake News Detection us-ing Hybrid BiLSTM-TCN Model with Attention Mechanism," 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2023, pp. 1130-1136, doi: 10.1109/ICAAIC56838.2023.10140491.
[8] Olusegun, R., Oladunni, T., Halima Audu, YAO Houkpati, Staphord Bengesi. (2023). Text Mining and Emotion Classification on Monkeypox Twitter Dataset: A Deep Learning-Natural Language Processing (NLP) Approach. 11, 49882–49894. https://doi.org/10.1109/access.2023.3277868.
[9] Giner, M., Luque, F. M., Zayat, D., Mart´ın Kondratzky, Moro, A., Serrati, P., . . . Cotik, V. (2023). Assessing the Impact of Contextual Information in Hate Speech Detection. IEEE Access, 11, 30575–30590. https://doi.org/10.1109/access.2023.3258973.
[10] Correia, A., Diogo Guimaraes,˜ Paredes, H., Fonseca, B., Paulino, D., Trigo, L., . . . Jameel, S. (2023b). NLP-Crowdsourcing Hybrid Framework for Inter-Researcher Similarity Detection. IEEE Transactions on Human-Machine Systems, 53(6), 1017–1026. https://doi.org/10.1109/thms.2023.3319290.
[11] Gayo-Avello, D. (2013). A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data. Social Science Computer Review, 31(6), 649–679. https://doi.org/10.1177/0894439313493979.
[12] Babar, M., Ahmad, A., Muhammad Usman Tariq, Kaleem, S. (2024). Real-Time Fake News Detection Using Big Data Analytics and Deep Neural Network. IEEE Transactions on Computational Social Systems, 1–10. https://doi.org/10.1109/tcss.2023.3309704



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)