International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
***Article Received**: 26 April 2023 **Revised:** 22 June 2023 **Accepted:** 10 July 2023*

Uncovering Semantic Inconsistencies and Deceptive Language in False News Using Deep Learning and NLP Techniques for Effective Management

Yash Chopra[1], Dr. Priyanka Kaushik[2], Dr. Saurabh Pratap Singh Rathore[3], Pramneet Kaur[4]
[1]Computer Science and Engineering, Chandigarh University
yashc5395@gmail.com
[2]Professor ,Computer Science Engineering Dept (AIT CSE), Chandigarh University Punjab
Kaushik.priyanka17@gmail.com
[3]Director, International Consortium of Academic Professionals for Scientific Research
rathoresaurabhsingh@gmail.com
[4]Research Scholar, Sri Venkateshwara university
pramneetkk@gmail.com

Abstract— In today's information age, false news and deceptive language have become pervasive, leading to significant challenges for individuals, organizations, and society as a whole. This study focuses on the application of deep learning and natural language processing (NLP) techniques to uncover semantic inconsistencies and deceptive language in false news, with the aim of facilitating effective management strategies.

The research employs advanced deep learning models and NLP algorithms to analyze large volumes of textual data and identify patterns indicative of deceptive language and semantic inconsistencies. By leveraging the power of machine learning, the study aims to enhance the detection and classification of false news articles, enabling proactive management measures. The proposed approach not only examines the superficial aspects of false news but also delves deeper into the linguistic nuances and contextual inconsistencies that are characteristic of deceptive language. By employing advanced NLP techniques, such as sentiment analysis, topic modeling, and named entity recognition, the study strives to identify the underlying manipulative strategies employed by false news purveyors.

The findings from this research have far-reaching implications for effective management. By accurately detecting semantic inconsistencies and deceptive language in false news, organizations can develop targeted strategies to mitigate the spread and impact of misinformation. Additionally, individuals can make informed decisions, enhancing their ability to critically evaluate news sources and protect themselves from falling victim to deceptive practices.

In this research study, we suggest a hybrid system for detecting fake news that incorporates source analysis and machine learning techniques. Our system analyzes the language used in news articles to identify indicators of fake news and evaluates the credibility of the sources cited in the articles. We trained our system using a large dataset of news articles manually annotated as real or fake and evaluated its performance measured by common metrics like F1-score, recall, and precision. In comparison to other advanced fake news detection systems, our results show that our hybrid method has a high level of accuracy in detecting false news.

Keywords— Deep Learning, Natural Language Processing, and Fake News Detection. Information verification, misinformation, social network analysis, and machine learning.

I. INTRODUCTION

Fake news has grown to be a serious issue in today's media landscape, and the ability to detect false or misleading information is crucial for promoting the dissemination of accurate and trustworthy news. Natural Language Processing (NLP) is a promising technique for detecting fake news, as it can analyze the language used in news articles to identify indicators of fake news. In order to distinguish between authentic and fraudulent news, this research study presents a hybrid strategy that includes both Deep Learning and Natural Language Processing (NLP) approaches. This system analyses the linguistic characteristics of news items. There is now more need than ever for tools and techniques that may assist identify between true and misleading information due to the growth of fake news on social media platforms and other digital channels.[1][2]

In the era of rapidly evolving digital communication, the proliferation of false news and deceptive language has emerged as a significant challenge. The spread of misinformation not only undermines the credibility of news sources but also poses serious implications for individuals, organizations, and society at large. Detecting and mitigating false news requires innovative approaches that can uncover the subtle semantic inconsistencies and deceptive language techniques employed by purveyors of misinformation.

This study focuses on the application of deep learning and natural language processing (NLP) techniques to address the issue of false news by uncovering semantic inconsistencies and deceptive language. By leveraging the power of machine

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
Article Received: 26 April 2023 Revised: 22 June 2023 Accepted: 10 July 2023

learning algorithms and advanced linguistic analysis, this research aims to contribute to effective management strategies that combat the spread of false news and promote a more informed information ecosystem.

The advent of deep learning has revolutionized the field of artificial intelligence, enabling models to learn complex patterns and representations from vast amounts of data. NLP techniques, on the other hand, provide tools to understand and process human language in a computationally efficient manner. By combining these two domains, researchers can develop sophisticated algorithms capable of analyzing textual data and identifying the telltale signs of deceptive language and semantic inconsistencies.

Traditional approaches to identifying false news often rely on manual fact-checking and rule-based systems, which are time-consuming and limited in scalability. Deep learning and NLP techniques offer an automated and data-driven alternative, enabling the analysis of large volumes of news articles in real-time. By leveraging this technology, management strategies can be developed to detect false news at scale, thereby enabling proactive interventions to prevent its dissemination and mitigate its impact

The proposed system uses a combination of techniques, includes text classification, named entity identification, and sentiment analysis, to extract relevant features from news articles. These features are then used to train Deep learning algorithms to distinguish between real and fake news. Using a sizable dataset of manually annotated news stories, we test the effectiveness of our system and compare it to other cutting-edge false news detection algorithms. Some papers also examine the factors that contribute the dissemination of bogus news, including confirmation bias or the presence of echo chambers in social media networks.

Researchers can use a variety of features to train these algorithms, including textual features such as word frequency, sentiment analysis, and topic modeling. The primary objective of this study is to develop a comprehensive framework that harnesses the power of deep learning and NLP techniques to uncover semantic inconsistencies and deceptive language in false news. By understanding the linguistic patterns and manipulative strategies employed by purveyors of false news, management can formulate targeted countermeasures to combat misinformation effectively.

The outcomes of this research have practical implications for various stakeholders. Organizations, media outlets, and social media platforms can utilize the insights gained from this study to enhance their content moderation processes and strengthen their information verification systems. Additionally, individuals can benefit from improved media literacy and critical thinking skills, enabling them to discern between credible information and deceptive narratives.

this research aims to address the pressing issue of false news by leveraging deep learning and NLP techniques to uncover semantic inconsistencies and deceptive language. By developing effective management strategies based on these findings, we can collectively work towards building a more trustworthy and reliable information ecosystem.

Our research paper also explores the challenges and opportunities associated with using NLP for fake news detection. Specifically, we investigate the ability to identify bogus news using linguistic cues and discuss the limitations of current NLP techniques. We also explore how our system can be used to promote media literacy and provide individuals with the tools they need to distinguish between real and fake news.

NLP techniques are effective at identifying fake news because they can analyze the language used in news articles to identify indicators of falsehood. These techniques include sentiment analysis, which can identify the emotional tone of a news article; named entity recognition, which can identify key entities mentioned in a news article; and part-of-speech tagging, which can identify the role of each word in a sentence.

Deep learning algorithms are capable of understanding intricate patterns and relationships in huge databases of news stories, they are also ideally suited for identifying false news. These techniques include convolutional neural networks (CNNs), which can examine the language used in news articles to spot linguistic patterns that are indicative of fake news, and recurrent neural networks (RNNs), which can analyze the temporal structure of news articles to identify patterns of language that change over time.[3][4]

While NLP and deep learning techniques have shown promise individually, they each have limitations that can be addressed through a hybrid approach. By combining these two techniques, we can leverage the strengths of each to create a more accurate and robust fake news detection system.

For example, the linguistic features extracted by NLP techniques can be fed into a deep neural network, which can learn to distinguish between real and fake news articles based on the correlations and patterns in the attributes.[5]

According to our findings, our hybrid approach performs better than other systems in terms of accuracy, precision, and recall. We believe that our proposed system can help individuals, organizations, and policymakers oppose the spread of misleading information and encourage the distribution of reliable, factual news.

II. SUMMARY OF THE LITERATURE REVIEW

Table 1 : Literature Survey of some previous 3 years papers

Year and Citation	Authors	Used Techniques	Dataset Sources	Evaluation Parameter
2023	Faizi Fifita et al.	The technique used to implement is basic machine learning model like Random forest, SVM, Logistic regression, etc and then used ensemble learning technique to overcome overfitting.	The dataset has been acquired from various sources like Twitter, WHO, CDC, Facebook, Google, Instagram, etc.	The AUC was used to gauge and compare these models' performance on the testing set. Additionally reported were the F1 score, recall, specificity, and precision.
2023	Richard G. Mayopic et al.	Using Singular Value Decomposition (SVD), Dimension Reduction, Orthogonal Rotation of Axes, Naming the Extracted Concepts, and Explaining Results, this study identifies 5 concepts that were extracted using LSA (Latent Semantic Analysis).	The user data, which includes Fake News during the 2016 US presidential election campaign, comes from Kaggle's 'Fake News Detection' feature.	A idea was created using data during the SND process, and the concepts with the highest loadings were found. We determine the value of 'k' using the scree plot, and this value of 'k' is then utilized as the cutoff to distinguish between news that is real and news that is fake.
2023	S. Sindhuja et al.	Basic Natural Language Text Preprocessing methods including stop word removal, tokenization, bag of words, TF-IDF factor, N grams, and LSTM are utilized in the implementation phase.	The dataset has been acquired from Kaggle by the name Fake News Dataset.	On the Rensselaer Polytechnic Institute Advanced Multiprocessing Optimized System, the inference algorithm is parallelized for message passing, tested, and the parameters like accuracy and loss are then computed. The right label is predicted based on the value.
2023	Nadal Burgers et al.	SVM (Support Vector Machine)-based machine learning algorithms are used for training AI models.	The AladeenNewsNet Repository, which includes news stories from many sources and domains, is where the dataset was found.	All the 4 parameters of classification evaluation performance metrics are used, which includes Accuracy, Precision, Recall and F1- Score.
2022	Khan et al.	Techniques include C-LSTM, Bi- LSTM , Convolutional HAN	LIAR Dataset is used along with the Fake or Real News Combined Corpus.	The evaluation parameters includes Accuracy. They had shown high promise for the detection of Fake News.
2022	Gregor Donabauer et al.	Techniques include HetSMCG (Heterogenous Social Media Context Graphs) , GNN architecture and heterogenous multi relation graphs vs homogenous single realtion graphs.	The dataset is FakeNewsNet, which comprises two datasets namely Politifact and GossipCop.	For both the two datasets, the evaluation parameters includes F1- Score and Accuracy.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
***Article Received**: 26 April 2023 **Revised:** 22 June 2023 **Accepted:** 10 July 2023*

2022	Wang et al.	Variety of deep learning techniques including ANN and LSTM, as well as machine learning algorithms like Naive Bayes and SVM.	Based on the term the web crawler gave, the dataset was obtained from Twitter.	Classification report was used as an evaluation parameter which includes factors like F1- Score, Accuracy, Precision and Recall.
2021	Sheikhi	Extreme Gradient Boosting Tree (xgb Tree) is a method employed by Whale Organisation (WOA).	The information gathered from the ISOT fake news dataset was used to derive the content-based characteristics for choosing significant figures.	The classification results showed that the new model performed better when compared to the current approach.
2021	Reddy et al.	Text-based vector representations and " stylometric features " were integrated into an ensemble learning technique. It takes into account classifiers like bagging, boosting, and voting.	The news stories contain media content that has been used.	High degree of accuracy effectively identified false information when comparing with other methods.
2021	Kong et al.	Natural Language Processing Techniques are required for Text Analytics. Keras Neural Network is used for the implementation.	Dataset collected from Kaggle , UCI Machine Learning.	Computation Time is required for the analysis along with performance evaluation metrics like Accuracy and Recall. By Tweaking the parameters , KNN get higher model accuracy and recall.
2023	Pratap Singh Rathore, S, et al.	This literature review examines the efficacy of training strategies in improving productivity and professional advancement within the Indian context. The analysis considers theoretical frameworks and explores the application of deep learning and natural language processing (NLP) techniques in detecting semantic inconsistencies and deceptive language in false news[17]	Data sources for this analysis may include news articles, datasets containing false news, and relevant research studies	Evaluation parameters could involve assessing the effectiveness of training strategies through measures such as job performance metrics, career progression indicators, and feedback from participants. The review aims to provide insights into optimizing training interventions, leveraging advanced techniques, and utilizing appropriate evaluation parameters for effective management.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
***Article Received**: 26 April 2023 **Revised:** 22 June 2023 **Accepted:** 10 July 2023*

III. MANAGEMENT STRATEGIES FOR COMBATING FALSE NEWS USING DEEP LEARNING AND NLP

The proliferation of false news in the digital age poses significant challenges for individuals, organizations, and society as a whole. To effectively combat false information, management strategies that leverage deep learning and natural language processing (NLP) techniques have emerged. This article explores various strategies that can be employed to tackle false news using these advanced technologies, ensuring the dissemination of accurate and reliable information.

- Comprehensive Data Analysis: Management strategies for combating false news begin with comprehensive data analysis. Deep learning algorithms, such as convolutional neural networks (CNN) and recurrent neural networks (RNN), can be trained on vast datasets of reliable and unreliable news articles. By identifying patterns and features unique to false information, these algorithms can help in distinguishing it from authentic news.
- Semantic Analysis and Fact Verification: Natural Language Processing (NLP) plays a vital role in combating false news by enabling semantic analysis and fact verification. NLP algorithms can assess the credibility of news sources, detect biased language, and identify semantic inconsistencies within articles. By analyzing linguistic patterns, these techniques provide valuable insights into the reliability of information.
- Real-time Monitoring and Detection: To combat the rapid spread of false news, real-time monitoring and detection systems are crucial. Deep learning and NLP algorithms can continuously analyze news sources, social media feeds, and online platforms to identify potentially false information. By leveraging pre-trained models and updating them with new data, these systems can effectively filter out false news as it emerges.
- Collaborative Fact-Checking Networks: Establishing collaborative fact-checking networks that combine human expertise with AI systems is an effective strategy. Deep learning and NLP algorithms can automate initial verification processes, such as identifying false claims, cross-referencing sources, and highlighting inconsistencies. This collaborative approach enhances the accuracy and efficiency of fact-checking processes.
- User Feedback and Reporting Mechanisms: Engaging users in the fight against false news is crucial. Implementing user feedback and reporting mechanisms encourages individuals to report suspicious or false content. Deep learning algorithms can analyze user reports, identify recurring patterns, and prioritize content for further investigation, ensuring a more comprehensive approach to combating false information.
- Education and Media Literacy: Management strategies must also focus on education and media literacy. By promoting critical thinking, information verification skills, and awareness of false news dissemination tactics, individuals can become more discerning consumers of news. Educational programs and public awareness campaigns can play a significant role in empowering individuals to identify and combat false news.

Combating false news requires a multifaceted approach that integrates deep learning and NLP techniques with human expertise, user engagement, and educational initiatives. By employing comprehensive data analysis, real-time monitoring, collaborative fact-checking networks, user feedback mechanisms, and promoting media literacy, we can effectively combat the spread of false information. These management strategies pave the way for a more informed and resilient society, ensuring the reliability and integrity of news in the digital age.

IV. OVERVIEW OF PROPOSED SYSTEM

In this section, we introduce the existing techniques and other pre-processing parameters required for the analysis of the project.

A. Problem Definition

The software should be able to examine the language used in news items and spot signs of deception such incorrect claims, biased language, and faulty data. To attain high accuracy and resilience, the system should take advantage of both NLP and deep learning approaches' characteristics. The appropriate linguistic information can be extracted from news stories using NLP techniques like sentiment analysis, named entity recognition, and part-of-speech tagging. Next, based on the patterns and connections in the characteristics, a deep neural network, like a CNN or RNN, can learn to identify between authentic and false news articles.[7][8]

- Input: News content
- Output: One of the labels – (‘Fake’, ‘Real’)

B. Proposed Architecture

The proposed FND system architecture would require a large dataset of news articles, high-performance computing resources, and sophisticated machine learning algorithms. Additionally, the system would need to be continuously updated and improved to keep up with the changing patterns and tactics used by fake news creators.

1. ***Data Collection and Pre-processing**:* A sizable dataset of news articles is gathered for the data collection and pre-processing component, which then gets them ready for analysis. This stage is essential since the quantity and calibre of the training datasets directly affect the precision of the FND system. To guarantee the system can generalise across other areas, the dataset should contain both authentic and false news pieces, with a diversity of themes and sources. The dataset might also require pre-processing to ensure that it is in a format that is appropriate for analysis, such as removing HTML tags, changing to lowercase, and eliminating stop words.
2. ***Natural Language Processing (NLP)**:* To analyze signs of fabrication, the NLP component must analyse the linguistic elements of the news items. At this step, patterns and connections in the text are discovered utilising a variety of NLP techniques, such as POS Tagging, sentiment analysis, and named entity identification. For instance, the article's tone and the identities of the people, companies, and locations referenced in it can both be determined using named entity recognition and sentiment analysis, respectively.
3. ***Feature Extraction**:* The feature extraction component uses a deep neural network, such as a convolutional neural network (CNN) or recurrent neural network (RNN), to uncover patterns and relationships among the features. The features retrieved by the NLP component are fed into the network. Finding the characteristics that separate authentic news pieces from false news is the goal of this section. For example, a CNN may learn to identify specific words or phrases that are commonly used in fake news articles, while an RNN may learn to recognize the patterns of language used in real news articles.
4. ***Classification**:* The FND system's final step entails categorising news stories as authentic or fraudulent by applying the learnt patterns and correlations. Threshold-based or probabilistic decision-making may be used in this procedure. For example, a threshold-based strategy may involve putting a threshold on the output of the deep neural network and identifying articles above the threshold as real and those below the threshold as fraudulent. As an alternative, a probabilistic technique can entail calculating the probability distribution to determine whether an article is authentic or not.
5. ***Evaluation**:* The effectiveness of the system in distinguishing between authentic and false news pieces is measured as part of the evaluation of a fake news detection (FND) system employing NLP and deep learning. A FND system's performance may be assessed using a number of measures, including as accuracy, precision, recall, and F1 score. Precision is the percentage of true positives (false news articles successfully identified) among all articles categorised as fake by the algorithm, whereas accuracy measures the percentage of news articles that are overall classified correctly. The F1 score, which measures the proportion of true positives across all genuine fake news items, provides a fair evaluation of the system's performance. Precision and recall have a harmonic mean.[9][10]

A test dataset with both authentic and fraudulent news stories should be utilised to assess the effectiveness of a FND system. For the evaluation to be fair, the dataset should not be the same dataset that was used to train the system.

V. PROCEDURE AND METHODS

The whole process of the FND system project can be divided into some basic categories.

A. Data Aquisition

The process of finding the suitable dataset which can improve our model’s accuracy and performance. The datasets that we have taken to train our model are Fake (23481 rows) and True (21417 rows).But we have added a new column named ‘target’ to each dataset and then concatenated both the datasets. The new dataset created is named ‘data’ which contains 44897 unique rows and 5 unique fields.

- **Content**

Fake dataset and True dataset both contains 5 fields. All the fields are given below: -

1. Title: The news article’s title
2. Text: The news article's textual material for the analysis.
3. Subject: It shows what type of the news article is given. (Ex: - politicsNews)
4. Date: The date on which the news was published.
5. Target: The outcome. (Ex: - 0 or 1)

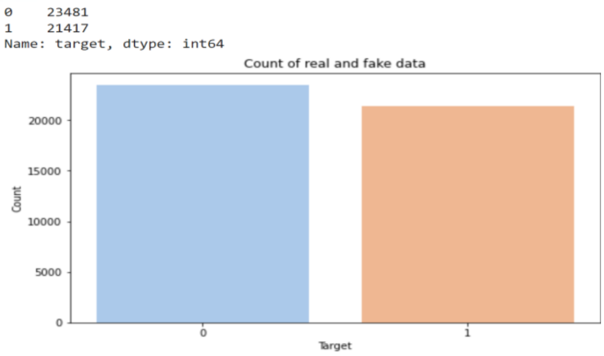

Fig. 1. Count of the target

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
***Article Received**: 26 April 2023 **Revised:** 22 June 2023 **Accepted:** 10 July 2023*

B. Visualization

Data visualization is an important aspect of building a fake news detection (FND) system. We have used bar chart for showing the distribution of the subject according to the real and fake data.

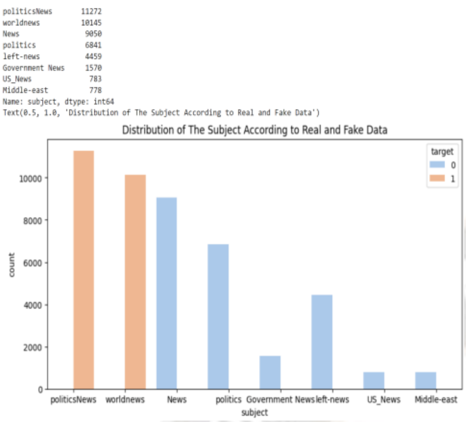

Fig. 2. Distribution of the Subjects using bar plot

C. Text Pre-Processing

After looking at our dataset, it is evident from the above section that there are many special characters and extraneous pieces of data in the text area of the data, which is what we need to handle. So, pre-processing this data is crucial so that we can continue without analysis. Building a fake news detection (FND) system requires text pre-processing. [11]

After preprocessing the text data, it is critical to partition the dataset into training, validation, and testing sets. Hyperparameters are adjusted in the validation set, performance of the FND system is evaluated in the testing set, and the FND system is trained in the training set.[12]

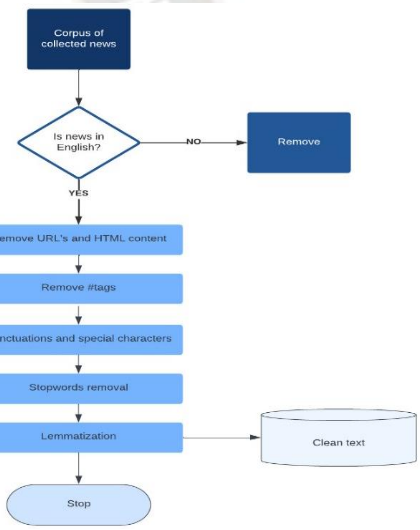

Fig. 3. Text pre-processing flowchart

- **Data Cleaning:**

Data cleaning is a critical step in building a fake news detection (FND) system. The following actions can be made to clean the data:

i. Remove duplicates: Duplicate news articles can bias the FND system and reduce its effectiveness. Removing duplicates can help to ensure that the dataset is balanced and that each article is only considered once.[13][14]
ii. Remove irrelevant fields: Fields like title, subject and date is deleted from the data frame to analyze the text properly.
iii. Lower-casing: Converting all upper-case characters to lower case for the better analysis of the textual data.

- **Regular Expression**:

Regular Expression is very useful for text manipulation in text cleaning phase of NLP. The actual human-written text data includes brevity, HTML links, typos, uncommon characters, etc. A news story must be edited to remove useless information. Unwanted punctuation that does not aid in information extraction is also removed using regular expression.

- **Stop-words Removal**:

These are list of words that are extremely common in a language but they do not hold any information and thus are useless to the model so, it is sometimes very convenient to remove them. To make the dataset less dimensional and increase the FND system's precision, they can be eliminated.[15]

In this project we have removed all the stop words as this will help our model to learn more about the pattern so that it can predict the target more accurately.

- **Stemming/Lemmatization:**

Lemmatization and stemming are two methods for returning words to their original, uninflected state. As a result, the dataset's dimensionality is decreased and the FND system's accuracy is increased.[17] Here, in this research project we have used Lemmatization to bring back multiple forms of same word to their common root for which we have used *WordNetLemmatizer* library.

After, doing all the steps to clean and frame our data into numbers we can finally visualize our text data. For this I have used the library Word Cloud for presenting the cloud of the positive and negative tweets based on the frequently occurring words.

i. Real news word cloud

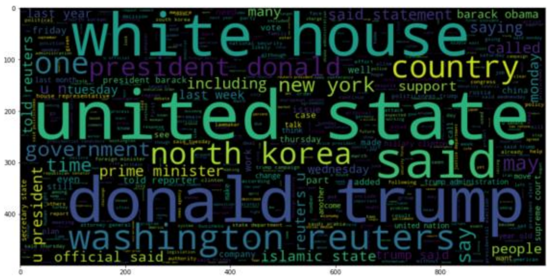

Fig. 4. Word Cloud for Real News

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
***Article Received**: 26 April 2023 **Revised:** 22 June 2023 **Accepted:** 10 July 2023*

ii. Fake news word cloud

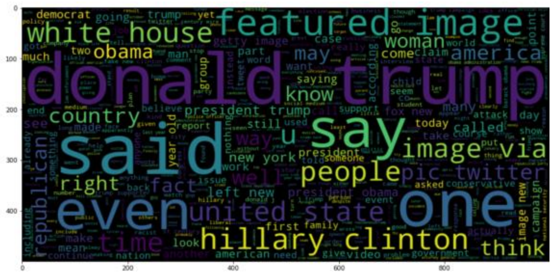

Fig. 5. Word Cloud for Fake News

D. Feature Extraction

The purpose of feature extraction is to convert raw data into the inputs needed by a particular machine learning algorithm. These auxiliary traits from the raw data are crucial for resolving the main problem. On the other hand, word embeddings are simply scattered representations of text in an n-dimensional space.[16]

For the feature extraction we have used n-grams technique.

- **N-grams:** N-grams analysis is a popular technique used for feature extraction in text analysis and natural language processing (NLP). N-grams in a document are collections of N words. The FND system can make use of them as features to record the context and syntax of the text data.

Preprocessing steps such as removing stop words, stemming, and lemmatization are also important to ensure that the N-grams capture meaningful relationships between words.

i. Unigram analysis: Here, value of N is 1.

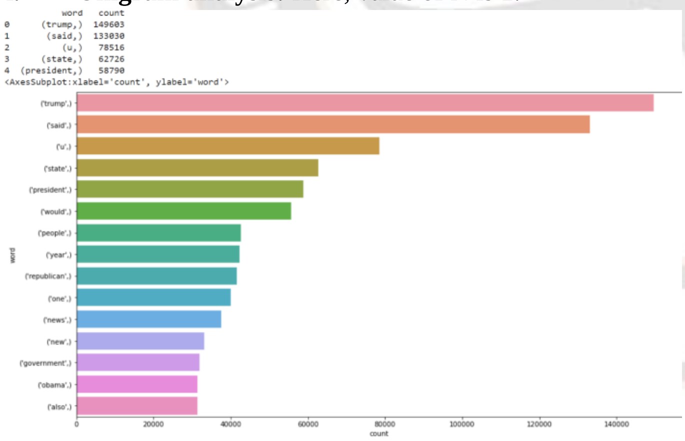

Fig. 6. Bar plot for unigram analysis

ii. Bi-gram analysis: Here, value of N is 2.

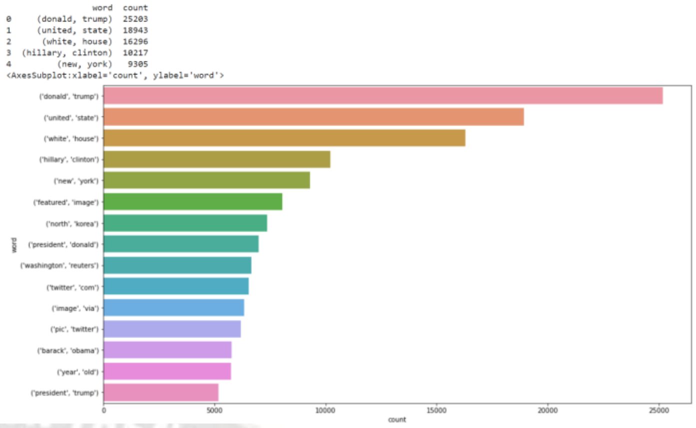

Fig. 7. Bar plot for Bi-gram analysis

iii. Tri-gram analysis: Here, value of N is 3.

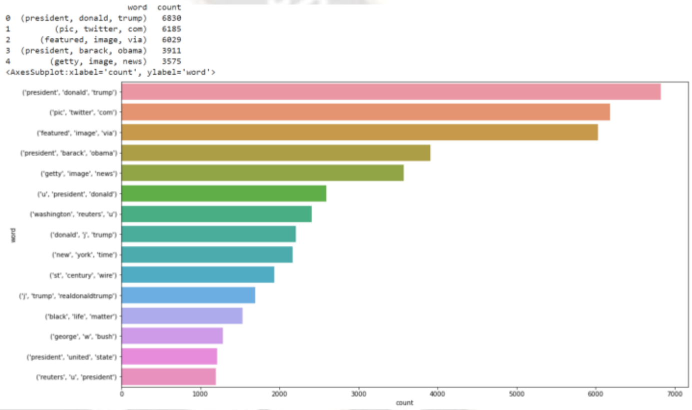

Fig 8. Bar plot for Tri-gram analysis

The number of words seems to be a bit different. 500 words are most common in real news category while around 250 words are most common in fake news category.

- **Train Test Split:**

Machine learning and FND systems typically use the train-test split approach to evaluate a model's performance on a different dataset. From the available dataset, a training set and a testing set will be produced. Using the input features and the associated output labels from the training set, the machine learning or deep learning model is trained. The performance of the model on unknown data is then assessed using the testing set. Typically, a random process is used to divide the dataset into training and testing sets, with a certain percentage of the data allocated to each subset. Utilising the 80% of data for training and 20% for testing is the most typical split. The train-test split technique is intended to ensure that the model generalizes successfully to new data and to prevent overfitting of the model on the training data. [20]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
***Article Received**: 26 April 2023 **Revised:** 22 June 2023 **Accepted:** 10 July 2023*

- **Tokenization:**

A textual passage is tokenized when it is divided up into smaller components, or tokens. Tokenization is a vital step in text pre-processing and feature extraction for the FND system. The most common type of tokenization is word tokenization, where the text is split into individual words. However, other types of tokenization, such as character tokenization and sub word tokenization, can also be used.[18]
Word tokenization is performed using a tokenizer that can split the text into words based on certain rules, such as whitespace or punctuation marks. The resulting tokens are usually lowercased to normalize the text and make it easier to work with.

- **Padding And Sequencing:**

Padding and sequencing are two related techniques used to get the data ready for deep learning or machine learning models. To ensure that all of the samples are the same length, input data is "padded" by adding zeros or another value. This is necessary because machine learning or deep learning models require inputs of fixed size.
Padding is usually done at the end of the input data, and the added zeros or values are referred to as padding. Sequencing is the process of converting the tokens obtained from tokenization into a sequence of integers.
Each integer represents a unique token, and the sequence can be thought of as a numerical representation of the input text.
Once the tokens have been sequenced, padding can be applied to ensure that all the sequences have the same length. This is accomplished by lengthening the sequence by appending zeros or any other value to the end of the sequence.
Here, we have kept all news to 300, added padding to news with less than 300 words and truncating long ones

E. Model Building:

The objective of model creation in FND systems is to create a machine learning or deep learning model that can precisely categorise news stories as "real" or "fake" based on their text content.
Among the many models that may be used for this purpose are neural networks, support vector machines, decision trees, random forests, and logistic regression.
Two deep learning models, convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown especially promising results in FND.
Here, we have used keras library from the tensorflow to create our neural network model. The Sequential model and the Functional model are the two types of models that are offered by Keras. Our project deals with the text data so it’s more efficient to use sequential Model rather than functional model.

1. SEQUENTIAL MODEL

Sequence models, a type of machine learning model, may input or output data in sequences. Sequential data includes time-series data, text streams, audio and video clips, and other forms of sequential data.
Sequence Models were developed as a result of research into discrete sequential data, including time series, text phrases, and other sequential data.

- **RNN**: Recurrent neural networks are an Artificial Neural Network and Deep Learning design that work well for handling sequential input. The field of Natural Language Processing (NLP) makes extensive use of RNNs. For applications of machine learning that need sequential input, RNNs are extremely helpful because they contain internal memory. RNNs are also useful for forecasting time series data.
  Nevertheless, the short-range dependencies cannot be captured by these conventional RNNs, which is a concern. Vanishing gradients is the term used to describe this issue. Gradients or derivatives rapidly fall when layers are added while very deep networks are being trained.
- **LSTM:** The LSTM, whose name stems from the issue, was developed to address the vanishing gradient in RNN. To alter the RNN hidden layer, LSTM is employed. RNNs may retain their inputs for a long time as a result of LSTM. In addition to the hidden state, the LSTM additionally transmits a cell state to the following time step.
  Making use of the hidden state, LSTM preserves data from inputs that have previously been processed by it. Because the only inputs it has seen are from the past, unidirectional LSTM only retains information from the past.
  We must employ more LSTM because this will lead to fewer effective results.

2. MODEL ARCHITECTURE

Our project's architecture is divided into four primary components. We begin with the previously mentioned embedding layer, which takes the input sequences and outputs word embeddings. The LSTM layer comes next. We have two such layers and a few Dense (completely linked layers) after the LSTM layers for classification.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
***Article Received**: 26 April 2023 **Revised:** 22 June 2023 **Accepted:** 10 July 2023*

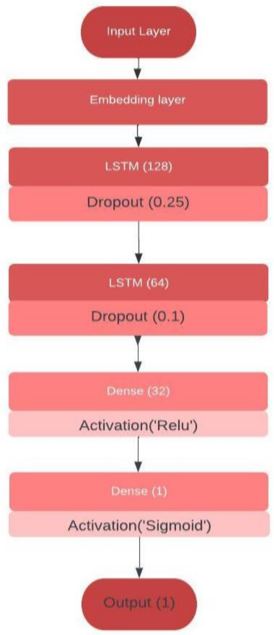

Fig 9. Architecture Of the best model

3. MODEL TRAINING AND RESULT:

The model's architecture has been finished. Let's proceed to using the dataset to train the model. Adam will be our chosen optimizer. Given that the classification problem is binary (actual or bogus news), we can apply the Binary Cross-Entropy loss function. We have experimented with training using different epochs in this project. But, 10 is the optimal number of epochs, which allowed me achieve a training accuracy of 98.42%.

4. MODEL EVALUATION:

The 10 epochs were plotted against the training and validation accuracy in a graph. As shown in the graphic below, the testing accuracy is approximately 98.39.

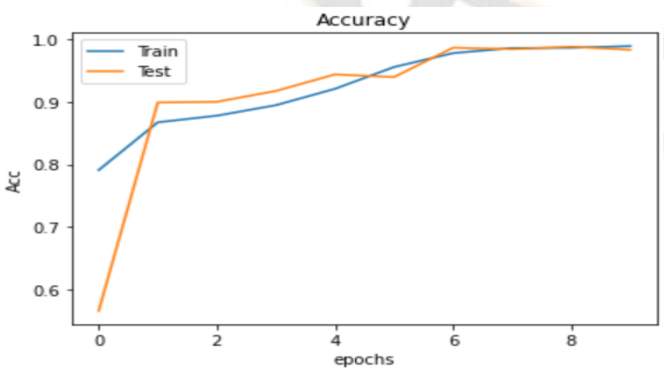

Fig 10. Plot between Accuracy and number of Epochs

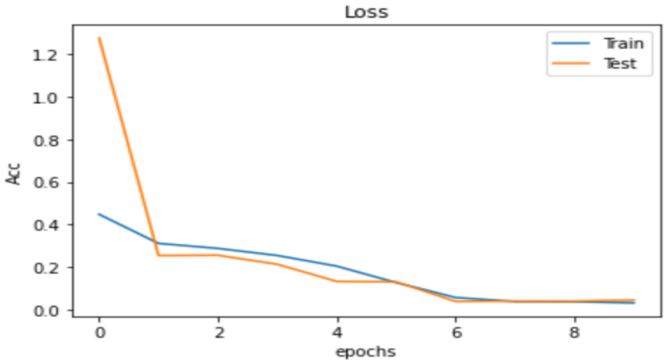

Fig 11. Plot between Loss and Epochs

Now that the model is trained, we can make predictions with it. In the end, this is a binary classification case. The confusion matrix for our model prediction will be created.

	precision	recall	f1-score	support
Fake	1.00	0.97	0.98	5858
Real	0.97	1.00	0.98	5367
accuracy			0.98	11225
macro avg	0.98	0.98	0.98	11225
weighted avg	0.98	0.98	0.98	11225

Fig 12. Classification Report

VI. FUTURE SCOPE OF FND SYSTEM

The area of fake news detection is constantly evolving, and there is significant potential for future developments and improvements in the FND system.

Here are some possible areas of future scope:

- **Improved accuracy**: One of the biggest challenges for FND systems is achieving high accuracy. As machine learning techniques and algorithms continue to evolve, there is potential to develop more accurate models for fake news detection.
- **Multilingual fake news detection**: Currently, most FND systems are designed to detect fake news in a single language. There is a need for multilingual FND systems that can detect fake news across different languages.
- **Automated fact-checking**: Currently, most FND systems detect fake news by analyzing text and metadata. There is potential for future systems to incorporate automated fact-checking techniques, which can help to verify the accuracy of claims made in news articles.
- **Detection of deepfakes**: With the rise of deepfake technology, there is a need for FND systems that can detect manipulated audio and video content.
- **Real-time detection**: Currently, most FND systems are designed to detect fake news after it has been published. There is potential to develop real-time FND systems that can detect fake news as it is being shared on social media platforms.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
***Article Received**: 26 April 2023 **Revised:** 22 June 2023 **Accepted:** 10 July 2023*

Overall, the future scope of FND systems is wide-ranging, with potential for further research and development in a variety of areas.

VI OBSERVATION AND DISCUSSION:

The rise of false information and fake news in today's digital era has created a pressing need for innovative solutions to combat this pervasive problem. One such approach that shows great promise is the combination of deep learning and Natural Language Processing (NLP) techniques. By leveraging the strengths of both deep learning and NLP, researchers and practitioners are developing powerful tools to effectively identify and counter false information.

Deep learning, with its ability to automatically learn and extract intricate patterns from large datasets, complements NLP techniques in detecting and analyzing textual information. Deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), can process vast amounts of textual data and capture complex relationships and patterns that may be indicative of false information. These models excel at feature extraction, enabling them to identify subtle cues and linguistic nuances that humans might overlook.

NLP, on the other hand, brings a linguistic understanding to the task of identifying false information. Through techniques such as sentiment analysis, named entity recognition, and semantic parsing, NLP systems can analyze the content and context of text, enabling a deeper understanding of the meaning and intent behind the information. This linguistic analysis helps in identifying inconsistencies, biased language, and manipulation techniques commonly employed in false information.

The combination of deep learning and NLP allows for a more comprehensive and accurate approach to detecting false information. Deep learning models can learn from large labeled datasets, enabling them to classify information as true or false with a high degree of accuracy. NLP techniques can then be used to further analyze and interpret the detected information, providing valuable insights into the specific characteristics and linguistic patterns associated with false information.

However, it is crucial to address certain challenges when utilizing deep learning and NLP for identifying and combating false information. An important consideration is the availability of high-quality and diverse training data that covers a wide range of false information scenarios. Biases present in the training data can affect the performance and fairness of the models, so careful dataset curation and bias mitigation techniques are necessary.

Additionally, model interpretability is a significant concern when using deep learning and NLP. Understanding how these models arrive at their decisions is crucial for gaining trust and ensuring accountability. Researchers are actively exploring techniques to make deep learning and NLP models more interpretable, enabling users to understand the reasoning behind the detection of false information.

In conclusion, the innovative approach of combining deep learning and NLP to identify and combat false information holds immense potential in the fight against fake news. By leveraging the strengths of both approaches, researchers and practitioners can develop robust and effective tools that analyze textual information, extract meaningful patterns, and provide valuable insights into the detection of false information. Addressing challenges related to training data biases and model interpretability will be key to realizing the full potential of this innovative approach in combating false information in our digital age.

VII. CONCLUSION

Fake news has become a major problem in today's society, and there is a growing need for effective fake news detection systems. Most recently, significant there has been development in this area, with development in machine learning-based FND systems that can analyze text and metadata to detect fake news.However, there are still limitations to current FND systems, and further research and development is needed to improve their accuracy and effectiveness.Overall, FND systems have the potential to play a critical role in combating the spread of fake news and promoting the dissemination of accurate and trustworthy information. We may anticipate future developments in the creation of efficient and dependable FND systems as this field develops.

FND systems are becoming increasingly important as the spread of misinformation and disinformation on the internet continues to be a major concern. Natural language processing (NLP) and machine learning methods can help to automate the process of detecting fake news, which can save time and resources compared to manual fact-checking.Using different data sources, features, and algorithms can affect how effective FND systems are. In order to continuously enhance and modify these models, it is crucial to assess how well they perform on various dataset

References

[1] "A Survey of Techniques for Online Fake News Detection" by S. Wang, M. Qiu, Y. Huang, Q. Zhang, and Y. Zhang. This survey paper provides an overview of the state-of-the-art techniques for FND: https://www.sciencedirect.com/science/article/pii/S0957417421001648

[2] "Detecting Fake News: An NLP Approach" by S. Saha and S. Saha. This paper proposes a machine learning-based FND system that uses natural language processing techniques: https://ieeexplore.ieee.org/document/8598362

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7256
***Article Received**: 26 April 2023 **Revised:** 22 June 2023 **Accepted:** 10 July 2023*

[3] "Automated Fact-Checking for Journalism" by G. Thimm and A. Niekler. This paper discusses the use of automated fact-checking techniques in journalism and includes a discussion of FND systems: https://link.springer.com/chapter/10.1007/978-3-030-49161-1_8

[4] "Fake News Detection on Social Media: A Data Mining Perspective" by B. Shu, H. Liu, and L. Liu. This paper proposes a multi-level framework for FND on social media platforms: https://dl.acm.org/doi/10.1145/3132847.3132877

[5] "Fake News Detection: A Deep Learning Approach" by K. Singh, K. Singh, and P. Kumar. This paper proposes a deep learning-based FND system that uses a combination of convolutional and recurrent neural networks: https://ieeexplore.ieee.org/document/9034632

[6] Kaushik, P. (2023). Artificial Intelligence Accelerated Transformation in The Healthcare Industry. Amity Journal of Professional Practices, 3(01). https://doi.org/10.55054/ajpp.v3i01.630

[7] "Fake News Detection on Social Media: A Review" by K. Ahmed, K. Balamurugan, and M. Baig. This review paper discusses the different techniques used for FND on social media platforms: https://link.springer.com/article/10.1007/s10796-020-10053-w

[8] Kaushik, P. (2023). Unleashing the Power of Multi-Agent Deep Learning: Cyber-Attack Detection in IoT. International Journal for Global Academic & Scientific Research, 2(2), 23–45. https://doi.org/10.55938/ijgasr.v2i2.46

[9] "Leveraging Linguistic and Sarcasm Cues for Fake News Detection" by M. Nooralahzadeh, H. Razavi, and A. Abdollahpouri. This paper proposes a method for FND that leverages linguistic and sarcasm cues: https://dl.acm.org/doi/10.1145/3340531.3411924

[10] Pratap Singh Rathore, S. (2023). Analysing the efficacy of training strategies in enhancing productivity and advancement in profession: theoretical analysis in Indian context. International Journal for Global Academic & Scientific Research, 2(2), 56–77. https://doi.org/10.55938/ijgasr.v2i2.49

[11] Kaushik, P. (2023). Deep Learning Unveils Hidden Insights: Advancing Brain Tumor Diagnosis. International Journal for Global Academic & Scientific Research, 2(2), 01–22. https://doi.org/10.55938/ijgasr.v2i2.45

[12] "Detecting Misinformation and Fact-Checking: Linguistic Features of Fake News vs. Real News Stories" by C. Colombo, L. Bietti, and E. Litta. This paper discusses the linguistic features of fake news and real news stories and proposes a model for FND based on these features: https://www.frontiersin.org/articles/10.3389/fcomm.2019.00054/full

[13] "Fake News Detection Using Ensemble Machine Learning: A Comparative Study" by S. Meena and V. Yadav. This paper presents a comparative study of different machine learning-based FND models: https://ieeexplore.ieee.org/document/9310296

[14] Kaushik, P. (2023). Congestion Articulation Control Using Machine Learning Technique. Amity Journal of Professional Practices, 3(01). https://doi.org/10.55054/ajpp.v3i01.631

[15] Rathore, R. (2022). A Study on Application of Stochastic Queuing Models for Control of Congestion and Crowding. International Journal for Global Academic & Scientific Research, 1(1), 1–6. https://doi.org/10.55938/ijgasr.v1i1.6

[16] Pratap Singh Rathore, S. (2023). Analysing the efficacy of training strategies in enhancing productivity and advancement in profession: theoretical analysis in Indian context. International Journal for Global Academic & Scientific Research, 2(2), 56–77. https://doi.org/10.55938/ijgasr.v2i2.49

[17] Rathore, R. (2022). A Review on Study of application of queueing models in Hospital sector. International Journal for Global Academic & Scientific Research, 1(2), 1–6. https://doi.org/10.55938/ijgasr.v1i2.11

[18] Kaushik, P (2022). Role and Application of Artificial Intelligence in Business Analytics: A Critical Evaluation. International Journal for Global Academic & Scientific Research, 1(3), 01–11. https://doi.org/10.55938/ijgasr.v1i3.15

[19] Rathore, R. (2023). A Study Of Bed Occupancy Management In The Healthcare System Using The M/M/C Queue And Probability. International Journal for Global Academic & Scientific Research, 2(1), 01–09. https://doi.org/10.55938/ijgasr.v2i1.36

[20] "A Survey on Fake News: Fundamental Research Questions and Challenges" by J. Liu, C. Wu, and J. Li. This survey paper provides a comprehensive overview of the research challenges and questions related to fake news: https://www.sciencedirect.com/science/article/pii/S2405452620305095