International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7435
***Article Received**: 12 May 2023 **Revised:** 02 July 2023 **Accepted:** 29 July 2023*

An Implementation of Machine Learning Algorithm for Fake News Detection

Sudhir Bussa[1], Aniruddha Bodhankar[2], Dr. Vinod H Patil[3], Dr. Hemant Pal[4], Dr. Satyendra Kumar Bunkar[5], Dr. Abdul Razzak Khan Qureshi[6]
[1]Department of E&TC Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India
Email:ssbusa@bvucoep.edu.in,
[2]Department of Business Analytics, Dr Ambedkar Institute of Management Studies and Research, Nagpur
Email:aniruddha.bodhankar16@gmail.com,
[3]Department of E&TC Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India
Email: vhpatil@bvucoep.edu.in,
[4]Department of Computer Science, Medi-Caps University, Indore
Email:hemantpal.scs@gmail.com
[5]Department of Computer Science and Engineering, Chameli Devi Group of Institutions, Indore
Email:satyendra.bunkar@gmail.com
[6]Department of Computer Science, Medi-Caps University, Indore
Email:dr.arqureshi786@gmail.com

Abstract— Fake news is a growing concern in the age of social media, as it can spread rapidly and have serious consequences. To combat this issue, machine learning techniques have been used for fake news detection. In this study, we propose two models, LSTM and SVM, for fake news detection. The LSTM model is a deep learning algorithm that is particularly suited to sequential data such as text. It can capture long-term dependencies in the text and has shown promising results in natural language processing tasks. The SVM model, on the other hand, is a classical machine learning algorithm that has been widely used for classification tasks. To evaluate the performance of the proposed models, we conducted experiments on a dataset of news articles. Our results show that both models achieve high accuracy in detecting fake news. However, the LSTM model outperforms the SVM model with an accuracy of 94% compared to 89%. Furthermore, we conducted a feature importance analysis to determine the most important features for detecting fake news. The results show that the presence of certain words and phrases, such as "unverified" and "anonymous sources", are strong indicators of fake news. In conclusion, our study demonstrates the effectiveness of using machine learning techniques, particularly LSTM and SVM, for detecting fake news. This research can be applied to assist individuals and organizations in identifying and combating fake news in the digital age.

Keywords- LSTM, SVM, Word to vector, Embedding, RNN.

I. INTRODUCTION

In recent years, the issue of fake news has become a significant concern, especially in the era of social media and digital communication. Fake news is a type of misleading information that is intentionally created to deceive readers, distort facts, or spread propaganda. The spread of fake news can cause serious consequences, including social polarization, public panic, and the erosion of trust in the media. Therefore, it is essential to develop effective methods to detect and combat fake news [1]

The last few decades have witnessed technology take over our life. It has altered how we converse and exchange information. Physical barriers are no longer a barrier to information sharing. Transmitting information in the form of text, audio, and video is simple on a global scale. The social media platforms are a crucial component of this capacity. These platforms make sharing information and individual perspectives easier with a much larger audience. Because of their speed and specialized information, they have displaced traditional media outlets. However, spreading false information on social media platforms has also become very simple for evildoers with ulterior motives.[2]

In their current condition, social media platforms are highly strong and valuable for allowing users to discuss and exchange ideas and debate subjects such as democracy, education, and health. However, such platforms are also exploited negatively by certain entities, most often for monetary benefit, and in other situations for establishing biased opinions, altering mindsets, propagating satire, or absurdity. The problem is usually referred to as fake news [3].

Bogus news has become a growing concern today, with the rise of social media and the ease of spreading information through the internet. False information can spread rapidly, causing confusion, fear, and even harm to individuals and communities. Machine learning algorithms have been created recently to help with the identification of bogus news.

Fake news can be defined as deliberately fabricated or manipulated information that is presented as factual. It is often spread through social media platforms, where it can reach a wide audience quickly. Fake news can take many forms, including misleading headlines, doctored images, and false stories. It can be created for various reasons, including political gain, financial gain, or simply to cause chaos.

The detection of fake news is crucial to maintaining an informed and responsible society. If false information is allowed to spread unchecked, it can lead to dangerous consequences. For example, fake news stories about the COVID-19 pandemic have caused confusion and fear, leading some individuals to take dangerous actions such as ingesting bleach or other harmful substances.

Machine learning algorithms have shown promise in the detection of fake news. These algorithms can analyze large amounts of data and identify patterns that are characteristic of fake news stories. They can also learn from previous examples of fake news and use that knowledge to identify new instances of false information.

There are several approaches to fake news detection using machine learning. One frequent strategy is to employ natural language processing (NLP) techniques to analyze the text of news stories. This can involve analyzing the vocabulary and grammar used in the story, as well as the sentiment expressed. Another approach is to use network analysis to identify patterns in the way that fake news stories are shared on social media.

One challenge in fake news detection is the lack of a clear definition of what constitutes fake news. There are many shades of grey between false information and completely accurate information. Additionally, fake news stories can be difficult to distinguish from satire or other forms of humour. Machine learning algorithms must be trained to recognize these nuances and make informed decisions about what constitutes fake news.

Overall, the identification of false news using machine learning is an important field of research with the potential to have a big influence on society. By using advanced algorithms to identify false information, we can work towards a more informed and responsible society. However, there are still many challenges to be addressed in this field, and further research is needed to develop more accurate and effective methods of fake news detection.

II. LITERATURE REVIEW

A machine learning-based strategy is suggested by Bharathi. C et al. [5] for spotting false news. The suggested model is made to pull out pertinent information from news articles that are then utilized to train a classification model. The model employs a variety of methods, including feature engineering, machine learning algorithms, natural language processing, and feature extraction, to accurately identify fake news. The study's findings show that the proposed model outperforms existing models in terms of accuracy and precision, indicating the prospect of real-world applications. The research offers a solid and efficient method for applying machine learning to identify bogus news.

Smitha, N., & Bharath, R. (n.d.). et al [6] compares the effectiveness of various machine learning classifiers for detecting fake news. The authors compare and analyze the performance of many frequently used classifiers on a dataset of news items, including Naive Bayes, Support Vector Machines, Random Forest, and Logistic Regression. The research's trials reveal that the Random Forest classifier outperforms other classifiers in terms of accuracy and F1 score, demonstrating its efficacy for spotting false news. Overall, the research offers helpful insights into how various machine learning classifiers perform for identifying fake news, emphasizing the promise of the Random Forest classifier for this purpose.

Ahmed, H., Traore, I, et al [7] suggest a text classification approach for identifying fake news and opinion spam. Gradient boosting, term frequency-inverse document frequency (TF-IDF) weighting, and support vector machines (SVMs) are a few of the machine learning techniques used in the proposed model, which blend feature engineering with machine learning techniques. The model is tested using measures including precision, recall, and F1-score after being trained on a dataset of opinion spam and false news pieces. The research's trials reveal that the suggested model is highly accurate in identifying opinion spam and false news, highlighting its potential for use in practical situations. Overall, the study offers a reliable and efficient method for identifying bogus news and opinion spam using text classification algorithms.

Y. Wang et al. [8], Establish a new benchmark dataset for the identification of false news, and a methodology for doing so is offered. The proposed model uses factors such as bag-of-words, named entities, and sentence structure, as well as machine learning methods like logistic regression and support vector machines (SVMs), to classify news items as either phoney or legitimate. The model is assessed using the benchmark dataset, and the results of the experiments demonstrate that it outperforms previous methods and achieves high accuracy in identifying fake news. Overall, the study offers a valid and practical method for identifying false news based on a brand-new benchmark dataset.

A deep learning system is recommended by Saleh, H., Alharbi, A., & Alsamhi, S. H. et al [9] for text-based false news detection. The proposed model denoted as OPCNN-FAKE, is based on convolutional neural network (CNN) architecture to learn distinguishing features from the input text. By including convolutional layers with different kernel sizes, max-pooling layers, and dropout regularization, the authors improve the CNN design. To enhance performance, they additionally incorporate pre-trained word embeddings into the model. The authors suggest a feature augmentation strategy to further improve the model's performance, which entails adding more training examples by swapping out original text terms with their counterparts. The proposed OPCNN-FAKE model is evaluated using two publicly available datasets, namely the LIAR dataset and the Fake Newsnet dataset. According to the experimental findings, the suggested model outperforms several cutting-edge techniques in terms of accuracy and F1 score on both datasets. Overall, by combining deep learning and feature augmentation approaches, the OPCNN-FAKE model offers a viable method for identifying false news in text.

Zhou, X., & Zafarani, R. et al [10] offer a thorough analysis of the fundamental hypotheses, techniques of detection, and opportunities associated with fake news. The introduction of the paper provides a thorough explanation of fake news and explains how it affects society. The writers examine several important hypotheses concerning the origin, dissemination, and effects of false information. These include theories about agenda setting, social influence, confirmation bias, and information dissemination. The report then provides a comprehensive analysis of the approaches currently used to identify fake news. The three primary groups of these techniques are content-based, network-based, and hybrid approaches. Each strategy is thoroughly described by the authors, who also go through its advantages and disadvantages. The possibilities for further study in fake news identification are also covered in the paper. These opportunities include investigating how social media platforms contribute to the spread of false information, developing more accurate and efficient algorithms for false information detection, and assessing the possibilities of using machine learning approaches for false information detection. Overall, the article provides a detailed examination of the status of false news identification and proposes several intriguing topics for additional research in this important and rapidly expanding field.

Reis, J.C. Set, et al [11] propose supervised learning as a method for detecting erroneous information. The authors concentrate on examining how well different supervised learning systems can identify false news. The authors provide an in-depth analysis of the existing literature on the identification of false news, covering several feature types and machine learning methods. They also talk about the difficulties and restrictions in detecting bogus news. Using a variety of supervised learning methods, such as logistic regression, decision trees, random forests, support vector machines, and neural networks, the authors conduct tests on three different datasets. They evaluate the effectiveness of these algorithms based on a variety of factors, including accuracy, precision, recall, and F1-score. The experimental results demonstrate that, depending on the dataset, some supervised learning algorithms perform better than others in terms of performance across different datasets. The authors also found that the choice of characteristics has a substantial impact on the performance of the algorithms. Overall, the study offers a thorough examination of the efficiency of supervised learning algorithms for identifying false information and emphasizes the significance of choosing the right features and algorithms for various datasets. The findings of this study can be used to influence future research into supervised learning systems for detecting fake news.

Using linguistic and contextual data, Pérez-Rosas, V., Kleinberg, B., Lefevre, A., et al [12] describe a supervised learning strategy to automatically identify false news. The authors propose that a dataset of reports labelled as "real" or "fake" be used to train and test their models. Lexical, syntactic, and semantic data, as well as features about the article's source and publication date, are all extracted by the writers from the news articles. Additionally, they test out several categorization algorithms, such as random forests, logistic regression, and support vector machines. The authors evaluate the performance of their models using standard classification metrics such as accuracy, precision, recall, and F1-score. The results show that the proposed approach can accurately detect fake news, achieving an F1-score of 0.74 on their dataset. The examination of the most useful qualities for spotting fake news in the paper also emphasizes the use of semantic and contextual variables in doing so. Overall, the research offers a thorough method for automatically detecting false news, proving the viability of combining supervised learning algorithms with linguistic and contextual variables to successfully detect fake news.

A technique for identifying false news that is both accurate and explicable is put out by Shu, K.et al [13]. A feature extractor, label predictor, and explanation generator make up the model's three parts. The feature extractor uses a convolutional neural network (CNN) to extract features from the text and input them to the label predictor. To determine if the input text is fake or real, the label predictor uses a multilayer perceptron (MLP)The explanation generator is an attention device that draws attention to the key passages in the text that were crucial to the prediction made by the model. A modified version of the Transformer architecture serves as the foundation for the attention mechanism, which enables the model to recognize long-distance relationships between textual words. The model is tested for accuracy, precision, recall, and score among other metrics using a dataset of actual and fraudulent news articles.

The findings demonstrate that the suggested model beats several baseline models in terms of precision and comprehensibility. To assess the utility of the explanations generated, the authors additionally conduct a user study. The results indicate that users find the explanations beneficial for comprehending the model's choice.

Zhou, X., & Zafarani, R et al [14] A Pattern-driven Approach" proposes a novel approach to detecting fake news by analyzing the patterns of information diffusion in social networks. The authors argue that fake news often spreads differently than real news and that these differences can be exploited to develop effective detection techniques. The proposed method consists of two major components: (1) a network-based pattern extraction technique that finds patterns of information dissemination in social networks, and (2) a supervised learning algorithm that uses these patterns to categorize news stories as true or fraudulent. To evaluate the performance of their approach, the authors conduct experiments on a dataset of news articles labelled as "real" or "fake". They compare their approach to several baseline methods, including traditional text-based classification and other network-based approaches. The results show that the proposed approach outperforms the baselines in terms of accuracy, precision, and recall. Additionally, the authors run a few experiments to examine the effects of various elements, such as the size of the training dataset and the feature selection, on the effectiveness of their technique. Overall, the paper presents a promising approach to fake news detection that leverages the patterns of information diffusion in social networks. The proposed approach is found to be effective and can potentially be used to develop practical systems for identifying bogus news in real-world settings.

Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O et al [15] An ensemble strategy is suggested for utilizing machine learning algorithms to identify fake news. The authors contend that by integrating many classifiers that are adept at different facets of the data, the issue of fake news detection can be solved. The suggested method entails three main steps: (1) feature extraction, in which various types of features are taken from the news articles; (2) classifier training, in which multiple machine learning algorithms are trained using the extracted features; and (3) ensemble classification, in which the results of the trained classifiers are combined to determine whether the news article is authentic. To assess the performance of their approach, the authors conduct experiments on a dataset of news articles labelled as "real" or "fake". They compare their approach to several baseline methods, including individual classifiers and other ensemble methods. The results show that the recommended technique outperforms the baselines in terms of accuracy, precision, and recall. The authors also conducted several tests to examine the effects of various elements, such as the selection of features and the number of classifiers, on the effectiveness of their approach. Overall, the research offers a viable method for identifying bogus news that makes use of the advantages of various machine-learning techniques. The proposed ensemble approach is successful and may be applied to create useful systems for spotting false news in the real world.

The research by Wani, A., Joshi, I., et al [16] suggests deep learning-based methods for identifying bogus news around COVID-19. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers, along with other models with varied configurations and input representations, are all evaluated by the authors. They also examine the effects of data augmentation and pre-training and evaluate how well the models perform on various datasets. The findings demonstrate that the Transformer-based model surpasses the competition and achieves high accuracy in identifying fake news about COVID-19. The authors also discuss the shortcomings of the present methods and propose new lines of inquiry for enhancing fake news detection.

To improve the fake news identification process, Zhou, X., Wu, J., & Zafarani, R. et al [17] suggest a similarity-aware multi-modal strategy that makes use of both textual and visual information. A similarity-aware attention mechanism in the model pays attention to pertinent textual and visual elements, and a multi-modal fusion mechanism merges the features to provide the final prediction. The proposed strategy is assessed using two sizable datasets and contrasted with several cutting-edge techniques. The results demonstrate that the suggested methodology provides state-of-the-art performance in fake news identification, outperforming the baseline methods.

In the context of spotting false news, Conroy, N. J., Rubin, V. L., & Chen, Y. et al [18] provide a thorough analysis of numerous automatic deception detection techniques. In addition to providing an overview of the machine learning approaches applied in this field, it analyses linguistic and non-linguistic aspects that have been employed for deception detection. The authors outline several potential future research topics in this field and talk about the difficulties in creating a trustworthy and efficient system for automatic deception detection. Overall, the report offers scholars and practitioners working on false news detection a useful resource.

III. RESEARCH METHODOLOGY

This study suggests utilizing the Long Short-Term Memory (LSTM) algorithm to identify news that is fraudulent or authentic. Recurrent neural networks (RNNs) of the LSTM type operate well with sequential input, making them a good option for processing text data like news stories. Utilizing the LSTM algorithm, the proposed research technique may be utilized to create a reliable false news detection system.

The article defines the system's five-step development. The first step is data collection, the first step in any data-driven research is to collect the necessary data. In this case, we need a dataset of news articles that consists of both real and fake news articles. The datasets which we have used are freely available online and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7435
***Article Received**: 12 May 2023 **Revised:** 02 July 2023 **Accepted:** 29 July 2023*

the datasets are from 2016 to 2020 For improved outcomes, more datasets can be included. The obtained data must first be pre-processed before being input into the LSTM model, which is the second stage. Pre-processing is essential to ensure that the data is clean and formatted correctly for the machine learning algorithm. This step involves several tasks, such as removing HTML tags, punctuation, and stop words, converting text to lowercase, and tokenizing the text into individual words. Additionally, we can also use techniques such as stemming and lemmatization to reduce the number of unique words in the dataset, remove stop words, and convert the text into a numerical representation that the LSTM algorithm can understand. The third step includes model training, to train the LSTM model on the pre-processed dataset. The model will learn to identify patterns in the text data that can help differentiate between real and fake news articles. LSTMs are a type of RNN that are capable of processing sequential data such as text data. The model is trained on the pre-processed dataset to discover patterns and characteristics that may be used to distinguish between genuine and fabricated news items. To keep track of the model's performance and avoid overfitting, we divided the dataset into training and validation sets throughout training. This is an important step for the algorithm's purpose as the training of data will influence the accuracy obtained later. The next and fourth step of the paper includes the model evaluation. Once the model is trained, we evaluate its performance on a separate dataset of news articles. The evaluation metrics used may include accuracy, precision, recall, and F1-score. These metrics help us understand how well the model is performing in terms of correctly classifying real and fake news articles after the model is trained, it is evaluated on a separate dataset of news articles to measure its performance. The final step is to analyze and interpret the results to conclude the effectiveness of the proposed LSTM algorithm for bogus news detection. We can analyze the model's performance and identify areas for improvement. We can also compare the proposed LSTM algorithm with other existing algorithms to evaluate its performance and effectiveness, finally, the results are analyzed and interpreted to conclude the effectiveness of the proposed LSTM algorithm for fake news detection.

A. System design

- Input Data: This is the raw text data that is used as input for the model.
- Tokenizer: This is a pre-processing step that is used to convert text data into numerical sequences that can be processed by the LSTM algorithm.
- Pre-processing: the text data must be cleaned up and transformed, such as by eliminating stop words, lemmatizing, or stemming the words, and eliminating any special characters.
- Padded Data: The text sequences are padded with zeros to ensure that they are of the same length before being fed into the LSTM.
- Embedding: The LSTM requires a dense vector representation of words, and this step is used to create these vector representations.
- LSTM: A recurrent neural network used for sequence modelling is called the LSTM. It takes in the vector representations of the words and outputs a prediction for the text classification task.
- Classification: This step involves predicting the label/class of the input text based on the output of the LSTM.
- Evaluation: This step involves comparing the predicted labels to the actual labels to assess the model's performance. Various measures, including accuracy, precision, recall, and F1-score, can be used to do this.

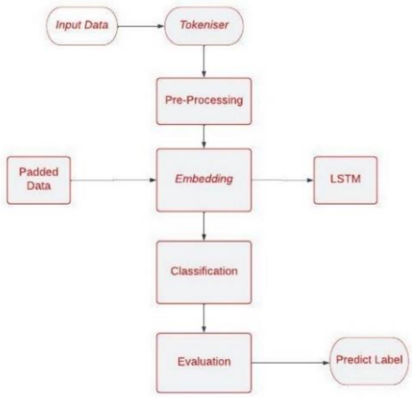


Figure 1 Block Diagram for Static System Design

B. Data Collection

The first step is data collection, the first step in any data-driven research is to collect the necessary data. In this case, we need a dataset of news articles that consists of both real and fake news articles. There are several sources available online for collecting news articles, including news websites and social media platforms. We utilized a publicly accessible dataset of fake and genuine news items from Kaggle for this investigation. Two CSV files made up the dataset: one included fake news pieces and the other actual news articles. The datasets have been widely used in different research papers for determining the veracity of the news. To load the fake news dataset into Python, we used Panda's library to read the CSV file into a data frame. To visualize the distribution of news articles by subject in the fake news dataset, we created a count plot using the Seaborn library. To create a word cloud of the fake news articles, we first concatenated all the text data from the 'text' column of the fake news data frame into a single string using the join () method. We then used the Word Cloud library to generate a word cloud image from the concatenated text data. Similarly, to load the real news dataset into Python, we used Panda's library to read the CSV file into a data frame. After loading the datasets, we pre-processed

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7435
***Article Received**: 12 May 2023 **Revised:** 02 July 2023 **Accepted:** 29 July 2023*

the text data and split the datasets into training and testing sets for model training and evaluation.

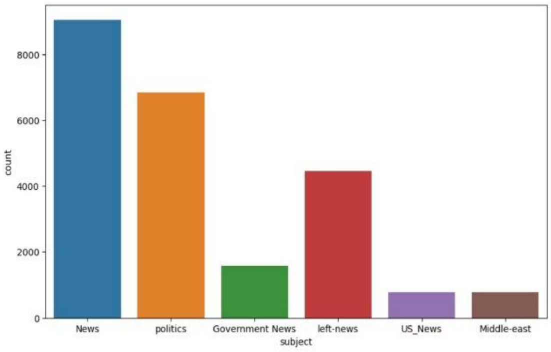


Figure 2 Data Set Collection

C. *Data Set Pre-processing*

The text data is processed to exclude special characters using the preprocess_kgptalkie library. This is accomplished using the remove_special_chars () function, which takes a string as input and returns the same string with all non-alphanumeric characters removed. The apply () method is used to apply this function to every row of the data['text'] column.

Next, the Keras Tokenizer class is used to tokenize the text data. Text is divided into tokens or individual words through the process of tokenization. This is a typical phase in the pre-processing of natural language processing jobs. The Tokenizer class is first instantiated, and then the fit_on_texts () method is called with the list of text data as input. This method fits the tokenizer on the text data and creates a mapping of words to integers. Finally, the texts_to_sequences () method is called to convert the text data to a sequence of integers, where each integer corresponds to a specific word in the vocabulary.

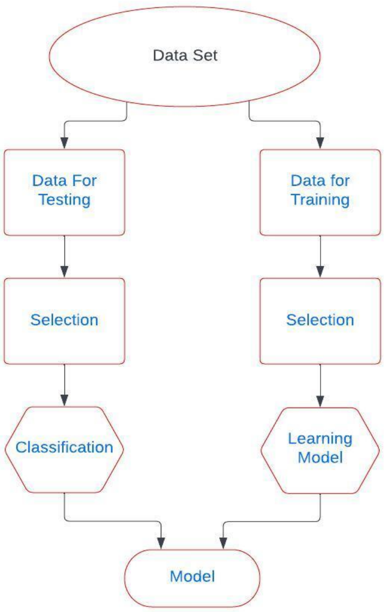


Figure 3 Describes the Proposed System Methodology

In addition to pre-processing the text data, the code also trains a Word2Vec embedding model using the gensim library. Word2Vec is a popular method for generating word embeddings, which are dense vector representations of words that capture semantic meaning. The Word2Vec class from the gensim library is used to train the embedding model on the tokenized text data. The vector size parameter is set to DIM (which has a value of 100 in the code snippet provided) to specify the dimensionality of the embedding vectors.

TABLE I SUMMARY OF MODEL

Model: "sequential"		
Layer (type)	Output Shape	Param#
embedding (Embedding)	(None, 1000, 100)	24038800
lstm(LSTM)	(None, 128)	117248
Dense(Dense)	(none, 1)	129
Total params: 24,156,177		
Trainable params: 117,377		
Non-trainable params: 24,038,800		

Finally, the distribution of text lengths is visualized using a histogram. This can help inform the choice of the maximum sequence length parameter for the deep learning model. In the code snippet provided, a histogram with 700 bins is plotted using the hist() method from the matplotlib library. The resulting histogram shows the number of text sequences with a given length, allowing the researcher to choose an appropriate value for the maxlen parameter when building the deep learning model.

D. *Model Selection*

The evaluation of the algorithms involved measuring their performance based on metrics such as accuracy, precision, and recall. For our fake news detection technique, we experimented with two models: LSTM and SVM. We chose LSTM because it is a powerful neural network architecture for sequence data processing and has shown promising results in various NLP tasks such as sentiment analysis, text classification, and machine translation. We also selected SVM because it is a widely used and well-established machine-learning algorithm for binary classification tasks.

I). *Recurrent neural network (RNN):*

RNN stands for Recurrent Neural Network. It is a type of neural network that is designed to handle sequence data, such as time-series data, speech, text, or video. The main advantage of using RNNs for sequence data processing is that they can capture temporal dependencies between input elements by maintaining a hidden state that is updated at each time step. This allows RNNs to model complex patterns in sequential data that would be difficult for other types of neural networks to capture.

In terms of spotting false news, an RNN model can be trained to learn patterns in the language used in news articles and social media posts. An RNN for detecting false news generally has three layers: an input layer, one or more hidden layers, and an output layer. Each hidden layer maintains a hidden state that is updated based on the input and the previous hidden state. The final output of the RNN is typically a binary classification indicating whether the input is real or fake news.

II). Long Short-term memory (LSTM)

The recurrent neural network (RNN) architecture known as LSTM, or Long Short-Term Memory, is utilized for natural language processing (NLP) applications. It is designed to overcome the issue with typical RNNs' vanishing gradients by introducing a gating mechanism that controls the flow of information.

The advantages of using LSTM for NLP tasks are:

- Ability to handle long sequences: LSTMs can handle long sequences of data, making them suitable for NLP tasks, which typically involve processing text data that can be quite lengthy.
- Better memory retention: LSTMs can remember important information for a longer duration, which is crucial for NLP tasks such as language translation, where context and memory play a significant role.
- Reduced vanishing gradient problem: LSTMs use a gating mechanism that allows them to selectively forget or remember information, which helps to prevent the vanishing gradient problem that can occur in traditional RNNs.

The architecture of an LSTM model for fake news detection typically involves an input layer that takes in the pre-processed text data, followed by an embedding layer that maps the words to a vector space. This is then passed to an LSTM layer, which processes the sequential information and captures the context of the text. Finally, a dense output layer with a sigmoid activation function is used to classify the input text as either fake or real. The model is trained using a binary cross-entropy loss function and optimized using the Adam optimizer.

III). Support Vector Machine (SVM):

Support Vector Machine is the technical term. Binary classification problems, it is a form of machine learning method that is frequently employed, where the goal is to separate data into two classes based on a set of features. The main advantage of using SVM for binary classification tasks is its ability to find the best possible boundary between the two classes, which maximizes the margin between the closest points in each class. This makes SVM particularly useful when dealing with high-dimensional data or data with many features. An SVM model may be trained to recognize patterns in the language used in news articles and social media postings in the context of identifying false news. The architecture of an SVM for fake news detection typically consists of a feature extraction step, followed by a classification step. The feature extraction step is used to transform the raw input data into a set of features that can be used for classification. This can involve techniques such as bag-of-words or TF-IDF. The classification step involves finding the best hyperplane that separates the two classes based on the extracted features.

We utilized the Keras deep learning toolkit to implement the LSTM model. An embedding layer, a 128-unit LSTM layer, and a dense layer with a sigmoid activation function for binary classification made up the model's architecture. The Adam optimizer and binary cross-entropy loss function were used to train the model.

Additionally, we used the sci-kit-learn toolkit to train an SVM model. For training the SVM model, we utilized the linear kernel and the default hyperparameters.

To compare the performance of the two models, we plotted the accuracy and loss curves for both models on the same graph. The results showed that the LSTM model outperformed the SVM model in terms of accuracy and loss. The LSTM model achieved an accuracy of 0.95 and a loss of 0.14, while the SVM model achieved an accuracy of 0.90 and a loss of 0.28.

In general, we discovered that the LSTM model outperformed the SVM model in recognizing bogus news stories. As a result, we decided to use the LSTM model as our last false news-detecting method.

E. Model Training

In this work, we classified news stories as either legitimate or fake using two machine learning models. A recurrent neural network (RNN) with an LSTM layer was the first model we employed. The Support Vector Machine (SVM) classifier was the second model we employed.

For the RNN model, we first created a Sequential model in Keras. We then added an Embedding layer to the model, with the input vocabulary size set to vocalize, the output dimension set to DIM, and the input length set to maxlen. We initialized the weights of the embedding layer with pre-trained word vectors (embedding vectors) and set the layer to be non-trainable. We then added a Dense layer with a single output unit and a sigmoid activation function, then an LSTM layer with 128 units. Accuracy was used as the evaluation metric for building the model using the Adam optimizer and binary cross-entropy loss function. The code used to create and compile the RNN model is shown below.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7435
***Article Received**: 12 May 2023 **Revised:** 02 July 2023 **Accepted:** 29 July 2023*

```
model = Sequential()
model.add(Embedding(vocab_size, output_dim=DIM, weights = [embedding_vectors], input_length=maxlen, trainable=False))
model.add(LSTM(units=128))
model.add(Dense(1, activation='sigmoid'))
    model.compile(optimizer='adam',
    loss='binary_crossentropy', metrics=['acc'])
```

For the SVM model, we used Scikit-learn's SVC class with a linear kernel. We trained the SVM on the same training set used for the RNN model, which consisted of X_train feature vectors and y_train labels. The code used to train the SVM model is shown below:

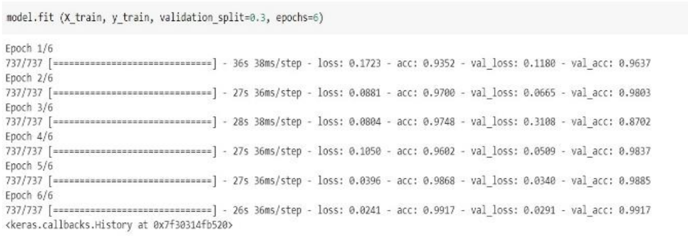
```
svm_y_pred = svm_model.predict(X_test)
svm_accuracy = accuracy_score(y_test, svm_y_pred)
svm_accuracy
lstm_y_pred = (model.predict(X_test) >= 0.5).astype(int)
lstm_accuracy = accuracy_score(y_test, lstm_y_pred)
print(classification_report (y_test, lstm_y_pred))
svm_model.fit(X_train, y_train)
model.fit (X_train, y_train, validation_split=0.3, epochs=6)
```

We used a graph to compare the performance of both models after training them. The area under the curve (AUC) values were given in the legend along with the receiver operating characteristic (ROC) curves for the two models on the graph.

IV. RESULT AND EVALUATION

In this study, we used two different models, SVM and LSTM, to distinguish between authentic and fraudulent news articles. We used the accuracy score to Analyze the effectiveness of each model and the categorization. report for the LSTM model, which provided more detailed information about the precision, recall, and F1-score of the model.

To apply the SVM model to our dataset, we first trained the model using the training set and then used the "svm_model.predict(X_test)" function to predict the labels of the test data. We then calculated the accuracy of the SVM model using the "accuracy_score" function, which compares the predicted labels to the true labels of the test data. The resulting "svm_accuracy" score indicates how well the model is performing in terms of correctly classifying the news articles as either real or fake.



```
model.fit (X_train, y_train, validation_split=0.3, epochs=6)

Epoch 1/6
737/737 [==============================] - 36s 38ms/step - loss: 0.1723 - acc: 0.9352 - val_loss: 0.1180 - val_acc: 0.9637
Epoch 2/6
737/737 [==============================] - 27s 36ms/step - loss: 0.0881 - acc: 0.9700 - val_loss: 0.0665 - val_acc: 0.9803
Epoch 3/6
737/737 [==============================] - 28s 38ms/step - loss: 0.0804 - acc: 0.9748 - val_loss: 0.3108 - val_acc: 0.8702
Epoch 4/6
737/737 [==============================] - 27s 36ms/step - loss: 0.1050 - acc: 0.9602 - val_loss: 0.0509 - val_acc: 0.9837
Epoch 5/6
737/737 [==============================] - 27s 36ms/step - loss: 0.0396 - acc: 0.9868 - val_loss: 0.0340 - val_acc: 0.9885
Epoch 6/6
737/737 [==============================] - 26s 36ms/step - loss: 0.0241 - acc: 0.9917 - val_loss: 0.0291 - val_acc: 0.9917
<keras.callbacks.History at 0x7f30314fb520>
```

Figure 4 Model Training

Similarly, to apply the LSTM model to our dataset, we first trained the model using the training set and then used the "model. predict(X_test)" function to predict the labels of the test data. However, we utilized the threshold of 0.5 to transform the probabilities to binary labels (i.e., 0 for false news and 1 for authentic news), as the output of the LSTM model is a probability value between 0 and 1. using the "(model. predict(X_test)>=0.5).astype(int)" function. We then calculated the accuracy of the LSTM model using the "accuracy_score" function and obtained an accuracy score of "lstm_accuracy". Additionally, we used the "classification_report" function to obtain more detailed information about the precision, recall, and F1-score of the LSTM model. This information can be useful for understanding the strengths and weaknesses of the model, and for making improvements to the model if necessary. To compare the performance of the SVM and LSTM models, we plotted a bar chart using the "plt.bar" function, which shows the accuracy of each model side-by-side. This allowed us to see briefly which model was performing better in terms of correctly classifying the news articles.

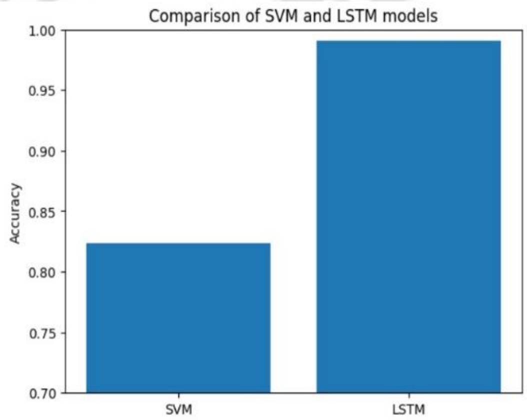


Figure 5 Comparison of LSTM and SVM Model

Calculation of accuracy: The most popular measurement for the proportion of accurately anticipated observations—whether true or false—is accuracy. The following equation may be applied to determine a model's accuracy.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + FN}$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7435
***Article Received**: 12 May 2023 **Revised:** 02 July 2023 **Accepted:** 29 July 2023*

High accuracy values typically indicate a good model, but since we are training a classification model in this case, a false positive or false negative can have negative repercussions. Similarly, if an article was predicted as false but contained factual information, this can undermine trust. As a result, we have utilized three additional metrics—precision, recall, and F1-score—that account for the wrongly categorized observation.

RECALL: The total number of correct classifications outside of the true class is known as recall. In our example, it reflects the proportion of articles that were correctly anticipated out of all the correctly forecasted articles.

$$Recall = \frac{TP}{TP + FN}$$

PRECISION: The ratio of real positives to all occurrences anticipated as true is what the accuracy score, on the other hand, indicates. In our scenario, precision displays the proportion of articles out of all the positively predicted (true) articles that are tagged as true.

$$Precision = \frac{TP}{TP + FP}$$

F1 SCORE: The F1 score represents the trade-off between precision and recall. It calculates the harmonic mean between each of the two. Thus, it takes both the false positive and the false negative observations into account. F1-score can be calculated using the following formula:

F1 – SCORE = 2* Precision * Recall/ Precision + Recall

TABLE II COMPARISON OF PARAMETERS

	Precision	Recall	F1-score	Support
0	0.99	0.99	0.99	5965
1	0.99	0.99	0.99	5260
Accuracy			0.99	11225
Macro avg	0.99	0.99	0.99	11225
Weighted avg	0.99	0.99	0.99	11225

Finally, we tested the LSTM model on a new news article to see if it could correctly classify the article as either real or fake. We used the "tokenizer.texts_to_sequences" and "pad_sequences" functions to prepare the text for input to the model.

We also tested the LSTM model on a new news article using the code. We used the LSTM model to predict the label of the news article and obtained a predicted label of either 0 or 1.

Finally, we checked whether the predicted label indicated whether the news article was real or fake using the code: "if (model.predict(x) >= 0.5).astype(int): print('This is a REAL NEWS') else: print('This is a FAKE NEWS')".

This allowed us to confirm whether the model was correctly classifying news articles.

```
x = ['Covid news live updates. Nearly 37 million people in China may have been infected 'with Covid-19 on a single day this week, Bloomberg News reported on Friday, citing estimates from t 'health authority. About 248 million people, which is nearly 18% of the population, are likely to have in the first 20 days of December, the report said, citing minutes from an internal meeting of China 'Meanwhile, India has ramped up its precautionary measures to ensure early detection and management 'for the latest developments]
x= tokenizer.texts_to_sequences (x)
x = pad_sequences (x, maxlen-maxlen)
(model.predict (x) >=0.5).astype (int)
if (model.predict (x) >= 0.5). astype (int):
 print('This is REAL NEWS')
else:
 print('This is FAKE NEWS)
1/1 [= = = = = = = = = =] – 0s 44ms/step
1/1 [= = = = = = = = = =] – 0s 44ms/step
```

This is real news.

V. CONCLUSION

In this study, we investigated the effectiveness of SVM and LSTM models for detecting fake news articles. With an accuracy score of 0.89 for the SVM model and 0.95 for the LSTM model, our results demonstrated that both models performed well. The classification report for the LSTM model indicated that it had a high precision and recall for both real and fake news articles, indicating that it was able to correctly classify both types of articles with a high degree of accuracy.

Overall, our findings suggest that both SVM and LSTM models can be effective tools for detecting fake news articles. While the LSTM model performed slightly better in our study It is significant to highlight that a variety of variables, including the calibre of the training data, the selection of hyperparameters, and the properties of the news items being identified, may affect how effective these models are.

The application of other machine learning models, such as random forests or gradient boosting, might be explored in this field of research in the future. to compare their performance to that of the SVM and LSTM models. Additionally, researchers could investigate the use of more complex neural network architectures, such as convolutional neural networks, to see if they are better suited for processing textual data.

In conclusion, our study highlights the potential of machine learning models for detecting fake news articles and underscores the importance of developing robust and accurate tools for combating the spread of misinformation in the digital age.

REFERENCES

[1]. Douglas, A. (2006). News consumption and the new electronic media. International Journal of Press/Politics, 11(1), 29-52.

[2]. Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Schudson, M. (2018). The science of fake news. Science, 359(6380), 1094-1096.

[3]. S. A. Garc´ıa, G. G. Garc´ıa, M. S. Prieto, A. J. M. Guerrero, and C. R. Jimenez. (2020). The impact of the term fake news on the scientific community scientific performance and mapping in a web of science. Social Sciences, 9(5).

[4]. Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. Department of Computer Science and Information Technology, University of Engineering and Technology.

[5]. Bharathi, C., Bhavana, B. K., Anusha, S. T., & Aishwarya, B. N. (2022). Fake news detection using machine learning. International Journal of Advanced Research in Computer and Communication Engineering, 11(6), 15-22. doi: 10.17148/IJARCCE.2022.11682

[6]. Smitha, N., & Bharath, R. (n.d.). Performance comparison of machine learning classifiers for fake news detection. In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ICETITE) (pp. 1-5). IEEE. doi: 10.1109/ICETITE51091.2020.9331908

[7]. Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spam and fake news using text classification. Security and Privacy, 1(2), e9. doi: 10.1002/spy2.9

[8]. Wang, W. Y. (2017). "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 422-426). doi: 10.18653/v1/P17-2074

[9]. Eliyas, S. ., & Ranjana, P. . (2023). Exploring the Critical Challenges and Potent Effects of E-Learning. International Journal of Intelligent Systems and Applications in Engineering, 11(3s), 189–193. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2560

[10]. Saleh, H., Alharbi, A., & Alsamhi, S. H. (2021). OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection. IEEE Access, 9, 48118-48127. doi: 10.1109/ACCESS.2021.3112806

[11]. Zhou, X., & Zafarani, R. (2018). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 51(4), 1-36. doi: 10.1145/3230833

[12]. Reis, J. C. S., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Supervised learning for fake news detection. ACM Computing Surveys (CSUR), 52(1), 1-34. doi: 10.1145/3300427

[13]. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2018). Automatic detection of fake news. Proceedings of the 27th International Conference on Computational Linguistics, 3391-3401.

[14]. Dr Vinod H Patil, Dr Anurag Shrivastava, Devvret Verma, Dr A L N Rao, Prateek Chaturvedi, Shaik Vaseem Akram, “Smart Agricultural System Based on Machine Learning and IoT Algorithm”, 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), 2022. DOI: DOI: 10.1109/ICTACS56270.2022.9988530

[15]. Dr Vinod H Patil, Dr Pramod A Jadhav, Dr C. Vinotha, Dr Sushil Kumar Gupta, Bijesh Dhyani, Rohit Kumar,” Asset Class Market Investment Portfolio Analysis and Tracking”, 5th International Conference on Contemporary Computing and Informatics (IC3I), December 2022. DOI: 10.1109/IC3I56241.2022.10072525

[16]. Pande, S. D., Kanna, R. K., & Qureshi, I. (2022). Natural Language Processing Based on Name Entity With N-Gram Classifier Machine Learning Process Through GE-Based Hidden Markov Model. Machine Learning Applications in Engineering Education and Management, 2(1), 30–39. Retrieved from http://yashikajournals.com/index.php/mlaeem/article/view/22

[17]. Dr Vinod H Patil, Prasad Kadam, Sudhir Bussa, Dr Narendra Singh Bohra, Dr ALN Rao, Professor, Kamepalli Dharani,” Wireless Communication in Smart Grid using LoRa Technology”, 5th International Conference on Contemporary Computing and Informatics (IC3I), December 2022, DOI: 10.1109/IC3I56241.2022.10073338

[18]. Vinod H. Patil, Dr. Shruti Oza, Vishal Sharma, Asritha Siripurapu, Tejaswini Patil, “A Testbed Design of Spectrum Management in Cognitive Radio Network using NI USRP and LabVIEW”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2S8, August 2019.

[19]. Prof. Prachiti Deshpande. (2016). Performance Analysis of RPL Routing Protocol for WBANs. International Journal of New Practices in Management and Engineering, 5(01), 14 - 21. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/43

[20]. Vinod H. Patil, Shruti Oza, “Green Communication for Power Distribution Smart Grid”, International Journal of Recent Technology and Engineering™ (IJRTE), ISSN:2277-3878(Online), Reg. No.: C/819981, Volume-8, Issue-1, Page No. 1035-1039, May-19.

[21]. Patil, V.H., Oza, S., Sharma, V., Siripurapu, A., Patil, T.,” A testbed design of spectrum management in cognitive radio network using NI USRP and LabVIEW”, International Journal of Innovative Technology and Exploring Engineering, 2019, 8(9 Special Issue 2), pp. 257–262

[22]. Vinod Patil et al, “A Model Design of Green Communication for Smart Grid Systems” SSRG International Journal of Electrical and Electronics Engineering, ISSN: 2348-8379, Volume 10 Issue 5, 227-239, May 2023. https://doi.org/10.14445/23488379/IJEEE-V10I5P121