

¹ T V Divya² Figlu
Mohanty

An Adaptive Word Vector Integrated With Enhanced Lstm For Classification Of Fake News



Abstract: - Social-media has great impact on society by spreading news which can be fake or real without analysis. Before confessing the news predictions has to be considered. To analyse a text through machines only NLP was best method. This focuses on five stages of NLP for detection of fake or real news. For creating vector to words a neural network technique word2vec was utilized for accurate prediction of word. For identifying meaning full words tokenization was utilized. By creating the vectors, the words have to be embedded for inputting the data to train and feature learning. In dense and dropout, the neural network method LSTM is initialized for the extraction of features or words. Dropout layer involves LSTM method for co-adopt the internal data. These layers mitigate overfitting by stochastically disabling neurons during training, hence fostering resilient feature acquisition. Dense layers, by use of non-linear mappings, allow the network to comprehend intricate patterns in data. This extensive investigation offers valuable insights into the development and fine-tuning of NLP models, which in turn enhances the accuracy of false news detection. To evaluate the performances epochs are utilized were accuracy and loss was validated. Five epochs are calculated were accuracy was gradually increased and loss had been decreased. The highest accuracy achieved was 99.8% and low loss value 0.01.

Keywords: Model Fine-tuning, Overfitting Mitigation, Text Embedding, Word2Vec, Tokenization, Feature Learning

I. INTRODUCTION

LSTM (Long Short-Term Memory) networks are a class of recurrent neural network (RNN) architectures tailored to apprehend and model enduring relationships within sequential data [13]. Numerous LSTM network variants have been put forth throughout time, each bringing unique modifications to the fundamental LSTM architecture. LSTM typically consists of three gates.

(a) Forget Gate: The "forget gate," a crucial part of an LSTM (Long Short-Term Memory) network, chooses which information from the prior cell state should be kept or removed [19]. This gate acts as a sigmoid layer by using the inputs of the previous cell state (C_{t-1}) and the current input (X_t). For each cell state element, it produces a value between 0 and 1. The forget gate at time step t has the following mathematical equation (1)

$$\text{Forget}_t = \sigma(\text{Weight}_{\text{forget}} * [\text{hidden}_{t-1}, \text{Input}_t] + \text{bias}_{\text{forget}}) - (1)$$

Where,

σ is an activation function

(b) Input Gate: The input gate in an LSTM network determines whether information from the current input should be incorporated into the cell state while accounting for the values from the previous cell state [14]. Through the use of a sigmoid activation function in the input gate, the network is able to control the amount of new data that is injected, ensuring that only relevant data is added to the cell state and irrelevant data is removed. Finding fresh information that the cell state ought to comprise is its primary duty [15]. The input gate is realized by the cooperation of the hyperbolic tangent (tanh) layer and the sigmoid layer. The computations for the updates are shown in equation (2)

$$\text{Input}_t = \text{forget}_t * \tanh(\text{Weight}_{\text{matrix}} * [\text{hidden}_{t-1}, \text{Input}_t] + \text{bias}_{\text{forget}}) - (2)$$

(c) Output Gate: In an LSTM (Long Short-Term Memory) network, the "output gate" makes the decision of which cell state data is used at a particular time step to create the output and hidden state. For this process to be successful, both the current input and the previous hidden state are needed [18]. A hyperbolic tangent (tanh) function and a sigmoid activation function interact to implement the output gate [16]. The network gains control over the data transmission from the cell state by employing a sigmoid activation function in the output gate. This ensures that relevant information is absorbed into the output and hidden state while discarding irrelevant data. The computation is shown in equation (2)

$$\text{Output}_t = \text{Output}_{t-1} * \tanh(\text{forget}_t * \text{Cell}_{t-1} + \text{Input}_t) - (2)$$

¹ Research Scholar, Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation, Aziz Nagar, Hyderabad, Telangana, Md2829@Gmail.Com

² Associate Professor, Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation, Aziz Nagar, Hyderabad, Telangana, Figlu@Klh.Edu.In

Copyright © JES 2024 on-line : journal.esrgroups.org

Figure 1 presents the architecture of traditional LSTM

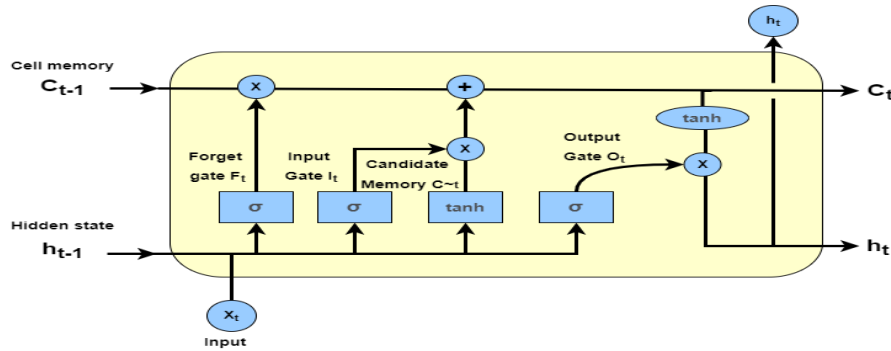


Figure 1: Traditional Architecture of LSTM

1.1. Need of LSTM for Textual Information: Recurrent neural networks (RNNs) include Long Short-Term Memory (LSTM) networks, which are especially good at processing sequential input, like textual documents [20]. When it comes to text, LSTMs have several benefits that make them an excellent choice for a wide range of natural language processing (NLP) applications. Because they can maintain a hidden state that contains information from previous time steps, LSTMs are unique in that they can handle sequences of different lengths [17]. One notable feature of LSTMs is its ability to capture subtle textual context. They have the capacity to retain and utilize historical data in order to forecast or categorize. Moreover, LSTMs are purposefully designed to mitigate the vanishing gradient issue that besets traditional RNNs. This pivotal attribute enables them to capture enduring dependencies within the text, a critical facet for comprehending the semantics and structure of textual data.

II. LITERATURE SURVEY

Ehtesham Hashmi et al [1] has developed a hybrid model for the detection of fake news text. Three individual data was introduced WELFake, Fake newsnet, and fake news prediction. Almost 11 data was present in the three dataset. The data was processed to pre-processing and filters three stage of removal of techniques are utilized tags, special characters, tokenization. Next goes to extraction of features in two categories of fast text. In unsupervised crawl embedding, embedding task specific, and handling of oov words. In supervised couple of techniques are utilized task specific and training the embeddings. The data was split and processed in three stages. Those are transmitted to three learning methodologies DL, ML, and Transformers. Among those LSTM and CNN was the appropriate method to retrieve efficient and accurate prediction of news. The layer involved are input, embedding, conv1d, pooling, LSTM segment, and Dense. Finally the evaluation of the model is categorized in two parts classification and model interpretability.

R. Uma Maheswari et al [2] has implemented a customized RNN model for the identification of fake news. The dataset was collected from LIAR which was public accessible in websites. This datasets contains different contents related to news which are taken from different websites relevant to political statements. 44k articles are present among those 21 are realm and remaining are fake. The data was transmitted to pre-processing stage where three kind of clearers are involved. Tokenization is utilized for removing all unwanted punctuations and initializing tokens to the words, removing stop words like expression, stop words means removal of articles, and finally, stemming finds the accurate and root of the topic. These data are transmitted to Lexicon method for extraction of features. This holds two stages Explicit and Latent features. These models create a variables with accurate prediction of words those are considered as extraction of features. Lexicon model identifies the words positive and negative and those are removed based on the mathematical identification. To those features COA is applied for selection of features. In just two steps the initialization and exploration where the extracted features are given and through COA the process was initiated to extract accurate words. In exploration, kind of strategy to hunt and attacks are considered and those are derived. As of now EDL is utilized for this RNN, vgg16 and resnet-50 are utilized for the actual prediction of words either fake or real. Hence the evaluation is performed and accuracies are derived.

In [3], Sheng Sciences et al remarked that a big issue in the internet era is bogus news. This study suggests utilising NLP approaches to identify and remove false information from sources that are unreliable on social media. Social media platforms may utilise models of neural networks trained with news information to accurately identify false news. N-gram vector-trained models outperform sequence vector-trained models somewhat. A technique for converting words into numerical vectors is called one-hot encoding. This project uses unigram, bi-gram, and tri-gram vectorization with TF-IDF encoding to produce numerical vectors which reflect the statistics of each n-gram. The TFIDF Vectorizer algorithms of the NLTK package are utilised to do N-gram vectorization and estimate TF-IDF. A sparse matrix of size $n \times m$ is created as a result, where n is the total number of rows and m is the

vocabulary size. The researchers claim that recurrent neural networks with LSTM algorithms can enhance NLP's capacity to identify fake news.

In [4], Iftikhar Ahmad et al reported that, information sharing has been transformed by the internet's and social media's explosive expansion. There were two voting classifiers utilised, one made up of random forest, logistic regression, and K-nearest neighbour, and the other of logistic regression, classification, linear support vector machine, and regression trees. The authors suggest utilising a machine learning ensemble strategy, which has demonstrated greater performance when compared to individual learners, to categorise news stories. Using the random forest approach with Perez-linear support vector machine, the greatest accuracy on DS1 was 99%. On DS1, individual learners' accuracy was 95.25% on average, while ensemble learners' accuracy was 97.67%. Wang-bidirectional long short-term memory networks, which had an accuracy of 62%, was the poorest algorithm. The bagging classification algorithm has the best performance. The study's main objective was to categorise false news stories using ensemble methods and machine learning models. The analysis used 4 actual world datasets.

In [5], Álvaro Science et al stated that false news has grown to be a serious problem, especially during the 2016 US election campaign. This study suggests three brand-new architectures for spotting bogus news, one of which is tailored from BERT. As observed in the US election campaign, the growth of fake news has been attributed to the expansion of internet access. Deep learning methods are being investigated to recognise and differentiate false news using text. These models perform better than benchmarks and imply that textual characteristics can be used by neural networks to identify bogus news. In order to prevent online fraud, deep learning models may be used by a variety of players, including social network providers and individual users. The researchers read 20015 pieces of news.

In [6], Somya Ranjan Sahoo et al. cite the intention of the study as the increasing use of social media, It is becoming more and more important to fight the spread of false messages, information and reduce our dependency on social media as a source of knowledge. Social media platforms are relatively failed to offer effective solutions to this problem caused by users interactions with false and not related news may lead to its spread at the ground level. It is necessary to use contemporary techniques to combat this dissemination of false information since it has the potential to negatively affect how the public perceives crucial issues. Convolutional Neural Network methods, long short term memory methods, ensemble methods, and attention mechanisms are just a few of the cutting-edge false news detection techniques that are being taken into consideration.

In [7] Jamal Abdul Nasir et al. proposed that Fake news and message detection in online social networks is a challenging problem, but deep learning techniques have shown promise in addressing it. Hybrid models that combine CNNs, RNNs, and attention mechanisms have achieved better results on public datasets. These algorithms concentrate on key characteristics that are suggestive of fake news and are capable of learning complicated patterns from text and imagery. On 2 of fake news datasets namely iso and fa-kes respectively, the proposed model was effectively tested, and the detection results it produced were practically better than those of previous traditional baseline approaches. The findings of additional experiments on the suggested model's generalization across other datasets were encouraging.

In [8] Ning Zhang et al. share his views on fake news detection has been limited, especially in the financial sector. There is no publicly accessible dataset of wrong financial news in China. A recent work builds out of the box dataset from clarification meetings intended for financial news using world wide web to fill this gap. It suggests a deep learning method that takes into account financial, contextual, and content aspects in order to identify bogus financial news. The proposed method outperforms several alternative baseline models with an accuracy of 94.38%. The characteristics of the article's content play a big part in spotting fake financial news, according to ablation studies. The Shapley value analysis's findings demonstrate how misleading financial news has characteristics that are different from those of authentic financial news.

In [9] Sachin Kumar et al. shares that Fake information is a impactful issue in the contemporary digital world, particularly on social media platforms. Deep learning has developed as a method with potential for detecting fake news, with a number of cutting-edge systems achieving high accuracy levels. The success percentage of a study by Ko et al. that attempted to identify bogus news was 85%. To extract elements from the news stories, including text, photos, and social media participation, they utilized a deep learning model. The model was then taught to determine whether the articles were authentic or not. These experiments show the potential of deep learning for the identification of not correct news. The demand for more labeled data and the creation of generalizable methodologies for use with new and unexplored data are two issues that still need to be resolved.

In [10] Chaitra K Hiramath et al research says that news is crucial to our daily lives, but fake news can be misleading and harmful. To address this problem, we propose a fake news detection system based on machine learning classification techniques. By comparing with Support Vector Machine models, Logistic Regression models, Deep Neural Network models, Random Forest models, and Naive Bayes models for detecting fake news. Found that DNN achieved the highest accuracy, followed by other available methods. However, DNN is also the most computationally expensive model. Given its high speed and precision, RF could be a preferable option for real-time applications. It is possible to discover false information on social media, news websites, and other

internet platforms by utilizing the suggested approach. Making better informed judgments and recognizing and avoiding fake news are two benefits it may provide.

In [11] Rohit Kumar Kaliyar et al. report a rise in interest in building false news detection systems in recent years. One effective method is bidirectional training, which may identify semantic and long-distance relationships in sentences. FakeBERT is a new approach for detecting false news. FakeBERT combines a bidirectional encoder representation from transformers with a single level deeper convolutional neural network (CNN). It is utilized by the CNN to deal with ambiguity, which is a significant obstacle for natural language interpretation. On 2 of the fake news datasets that were made accessible to the public, FakeBERT was tested, and it outperformed previous models with an accuracy of 98.90%. This shows that methods for bidirectional training can be quite successful in identifying bogus news.

In [12] Awf Baykara et al. Author analyzes the transition from traditional TV stations to the internet for news consumption. It also highlights the use of ml and dl classifiers to identify fake news on social media. AdaBoost for feature extraction which had a maximum accuracy in the research of 100%. Three phases made up the study's structure. This study extracted semantic characteristics from the clean dataset in the first level, using specific data filtering and cleaning procedures. Prepositions were eliminated from the stop word filters are used to sort the data into categories. In this study, the data were filtered and contaminants that prevented categorization were removed using Hypertext Markup Language tags and non-English character removal methods. Semantical traits were transformed into feature list in the second level and in the third level, the dataset's objects were categorized using deep learning and machine learning techniques.

Table 1: Comparative Analysis on the Existing Approaches

S. No	Author	Method	Merits	Demerits	Accuracy
1. 1	Ehtesham Hashmi et al	DL and ML	Same accuracy was achieved with couple of datasets.	Huge process for detection of text when the data is large then time-complexity increases.	99%
2. 2	R. Uma Maheswari et al	RNN, VGG16 & ResNet50	By customizing the layers the data has been predicted from early stages.	Highly computational complex.	96.23%
3. 3	Sheng Sciences	N-grams and Keras neural network models	fast computation time and high recall rate	Can be improved by tuning, RNN+LSTM can achieve better results.	90.3%
4. 4	Iftikhar Ahmad	ML ensemble	Parameters were tuned, evaluated using different metrics.	Can't identify the complex wordings, assumes the linearity.	92.25%
5	Álvaro Science	LSTM +CNN+BERT	Uses transformers, hyper parameter tuning for better results.	High consumption of power, should study on more articles.	93%
6	Somya Ranjan Sahoo	long short-term memories (LSTM)	The suggested technique for spotting fake news on Facebook looks at user content as well as features of news content.	Deep learning algorithms demand more training and testing time than machine learning algorithms do.	99.4
7	Jamal Abdul Nasir	Hybrid CNN – RNN	Exhibit clearly a clear dominance against the traditional methods.	Poor generalization is shown in the cross dataset validation results.	-
8	Ning Zhang	Extended Deep neural network	Proposed model given greater accuracy than simple neural network.	Accuracy may decrease when the message are long and elaborated.	94%
9	Sachin Kumar	Bidirectional LSTM, CNN and LSTM, Bidirectional Lstm and Lstm, Cnn and Lstm	offers models that performed far better than state-of-the-art methods at the time, with a maximum accuracy of 88.78%.	Not able to spot the semantic shift from real information to fake information throughout its transmission, mostly due to the comparative style of proposed study.	87%
10	Chaitra K Hiramath	Deep neural networks with NLP	Requires less memory than all the researched models.	High cost and takes more computational time.	
11	Rohit Kumar Kaliyar	BERT-based deep convolutional	Proposed model Gave great accurate results As of comparing with the existing thresholds of other models of 98.90%.	Fails to detect the existence of fake news for multi labelled datasets	98.90%
12	Abdulrahman	ANN, CNN + LSTM, RNN + LSTM	By applying optimized feature extraction parameters the model gain highest accuracy	the techniques for spreading false information on social media that don't use words, such as the use of false pictures or videos.	90.12%

A. Research Gaps Identified:

1. Overfitting still needs to be addressed, particularly with deep learning models. In order to guarantee model robustness and avoid overfitting, techniques such as regularization methods and dropout layers require additional research.
2. Although hybrid models that fuse RNNs, CNNs, and attention mechanisms demonstrate potential, generalizing these models to various datasets remains a difficulty. To confirm that these models are generalizable, more research using diverse and new datasets is required.
3. Specific domains are used to train and test most models (e.g., political news, financial news). To make sure that algorithms can reliably identify fake news in a variety of settings and domains, research is required to investigate cross-domain adaptability.

III. PROPOSED METHODOLOGY

The primary objective of this research is to optimize word representation techniques and neural network models to enhance a variety of NLP applications. This article investigates word vectorization processes such as tokenization, embedding layer adjustments, and deployment of dense and dropout LSTM networks using Word2Vec. The comprehensive approach aims to improve the identification of fake news and other text analysis problems by utilizing strong neural network designs and state-of-the-art natural language processing (NLP) approaches. The results demonstrate notable improvements in model robustness and accuracy and provide valuable new insights into the development and enhancement of NLP models.

3.1. Word Vector: Using a neural network model, the word2vec method extracts word associations from a large text corpus. After training, this model may identify related phrases or suggest ways to complete an incomplete sentence. As implied by the name, word2vec represents each distinct word using a vector, or a specific collection of integers. The syntactic and semantic characteristics of words have been meticulously crafted into these vectors. Because of this, a straightforward mathematical process may be used to determine the level of semantic similarity between words that correspond to these vectors. Word embeddings are produced using a collection of linked models called Word2vec. These fairly shallow, two-layered models are trained to predict the linguistic contexts in which words appear. Word2vec creates a vector space that usually has several hundred dimensions by using a huge text corpus as input. Within this area, a matching vector is allocated to each distinctive word in the corpus. Word2vec provides the option to generate distributed word representations using one of the two model architectures: Continuous Skip-Gram or Continuous Bag of Words (CBOW). Word2vec traverses the corpus in both topologies, taking into account individual words as well as a dynamic context window.

3.2. Tokenization: Breaking a sentence into meaningful words, phrases, symbols etc is known as tokenization which is a basic task related to NLP. The splitting of words are particularly depends on task mainstay. Division of words is the initial stage for NLP which can leads to analyze the words in sentiment analysis, recognition of name's, and translate of machine. So, here are some of the breakdown tokenization word, sentence, character, sub-word, special. Word token is the base and common split of sentences this varies based on language. Suppose "NewYork" can be divided in two words. This is possible based on the entity dependent i.e, how to split the word. Sentence token is most easiest way to divide the sentences where it divides accordingly from the paragraphs. This is mainly depends on the level of sentences derivations for analysis of machine translation. Character token is the specially for the splitting each character in level wise. This kind of token is mainly utilized for the generation of text or modelling the characters in levels. Subword token is mainly to overcome the limitation in character. It divide the single word into small sub words. This was useful in complex morphology which can deal with vocabulary. Special token name itself define the special characters in the sentences.

3.2. Modified Embedded Layer: An embedding layer plays a pivotal role in numerous natural language processing (NLP) and deep learning models, especially when dealing with text or categorical data. Its primary purpose is to convert discrete input data, such as words or categorical attributes, into continuous vector representations. This transformation simplifies the task of training neural networks with such data. In the realm of NLP, particularly in text data applications, the embedding layer is commonly referred to as a "word embedding layer." Its role is to associate each word within a vocabulary with a compact vector containing real numbers. These vectors are acquired through the training process and effectively encapsulate the semantic connections between words. By converting categorical values into continuous vectors, the embedding layer equips the network with the capacity to learn and comprehend the relationships among distinct categories. The embedding layer acts as a bridge between the neural network and the text or category data. It allows the network to effectively learn from these types of data by representing them in a continuous, low-dimensional vector space. Figure 2 shows the embedded layer architecture

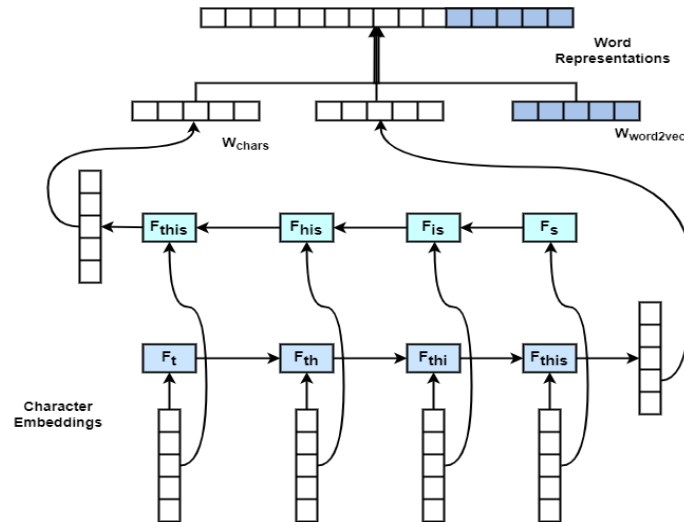


Figure 2: Embedded Layer Working

The proposed model computes the weight matrix to identify the unique words of the text. The weight matrix is updated as shown in equation (3)

$$E_{\text{update}} = E - \alpha * \frac{\partial L}{\partial E} \quad (3)$$

Weighted matrix is also known as simple word embedding or embedding matrix in NLP. Matrix represents the words related to vocabulary in the form of dense vector. The working of matrix is in six simple steps namely creating, initializing weights, assigning embedding, fine-tuning, usage in NLP, and updating embed. In creating stage, the data is collected from the corpus which has unique vocabulary were duplicates are removed. Weight initialization was done once the words are arranged in the rows and a fix number for column. To assign embedding for word they are randomly arranged. This words hold the semantic info related the word. The representation of the word is typically performed with the row form related to weights. To train the words three major techniques are utilized which has learning the text corpora which can hold general information in semantic form about words. Now the trained data was initiated to NLP which can process the text based on required task. Finally, the values are updated according to matrix weights for utilizing in other models.

3.3. Working of Dropout Layer: The dropout layer is a regularization method frequently applied in neural networks, including those featuring LSTM (Long Short-Term Memory) layers. Its primary objective is to address the issue of overfitting, a common concern in deep learning models, especially in the context of intricate and high-capacity networks like LSTMs. During the training phase, dropout introduces random inactivation of a fraction of neurons to counteract overfitting. This technique encourages the network to acquire more robust and broadly applicable features, as it discourages an excessive reliance on any specific neuron or connection. By doing this, dropout assists in reducing the issue of co-adaptation, which occurs when specific neurons or clusters of neurons work together to commit the training data to memory instead of deciphering important patterns. Given that LSTMs contain a large number of internal relationships that can co-adapt to the training set, this can be a serious issue. While lower dropout rates give milder regularization, higher dropout rates offer greater regularization however may slow down training[21].

3.4. Working of Dense Layer: Sequential data's temporal relationships and context are captured by LSTM layers, yet the output of LSTM is frequently in a high-dimensional space. By generating non-linear mappings, the dense layer allows the network to employ activation functions such as sigmoid or ReLU to identify complex patterns in the input. To create more complex and advanced models, these thick layers may be combined with various neural network architectures and are very customizable. Dense layers, for example, can be combined with convolutional layers in a convolutional neural network (CNN) for content or image processing, or stacked to form deep feedforward networks. To avoid overfitting, dense layers can also be utilized for regularization and dropout. By include dropout in the dense layer, the model becomes more resilient and has better generalization skills. The dense layer(s) provide the flexibility to adapt the output from the LSTM to the specific requirements of the task.

IV. RESULTS & DISCUSSION

```
array([ 1.7911379 , -0.23086372, -1.2384303 ,  0.41087487, -0.04847303,
       -0.17432329,  0.9885312 , -1.381359 , -0.19936675,  2.1544063 ,
        0.9801826 ,  0.81040865, -0.36999494,  2.192504 , -0.9478089 ,
        1.855237 ,  0.89092207, -1.1284126 ,  0.1605584 ,  0.7245316 ,
        3.2476208 ,  0.2860338 ,  1.225569 , -1.5010508 ,  1.0329179 ,
        1.2394263 ,  1.2832338 ,  0.11910871,  1.6482319 ,  0.21370089,
        1.6440533 , -1.9183382 , -1.1155337 , -0.83083296, -0.3960549 ,
       -1.7886214 ,  0.7419282 ,  0.02013973,  2.3721476 ,  0.870497 ,
       -0.81956345,  0.48219934,  0.5375499 , -0.37568343, -2.278334 ,
        1.4235444 , -2.7055788 ,  0.46141508, -1.2038699 , -1.3470327 ,
        0.4231823 ,  0.50537145, -2.0841286 , -1.2457424 , -0.94154805,
       -0.65247536, -0.19601114,  0.0532829 , -0.07003755,  1.5808731 ,
        0.9274005 ,  2.019521 , -1.2395371 ,  1.8678569 ,  0.9282163 ,
        3.3755949 , -0.7733084 ,  1.3425415 ,  0.9648181 , -0.87400144,
       -0.48112467,  1.7195612 , -0.3760573 , -0.46213034, -2.008174 ,
       -1.3306111 , -0.6811786 ,  2.5099654 ,  0.03317074, -0.8346526 ,
        0.26884046,  1.1883427 ,  2.677131 , -0.98547006,  1.2932502 ,
        0.4166407 ,  0.20991144,  0.15625773,  1.1547222 ,  1.2415354 ,
        0.33885446, -2.2756493 , -1.7779553 , -0.0436683 ,  1.215095 ,
       -1.437294 , -0.7115056 , -1.180643 ,  1.0807924 , -0.55654  ],
      dtype=float32)
```

Figure 4: Word to Vector Transformation

Figure 4 illustrates the words to vector transformers. These words are combination of real and fake. Each vectors are assigned with positive and negative values. The datatype of the values are float which contains 100 values are present in an array.

```
Epoch 1/5
842/842 [=====] - 185s 214ms/step - loss: 0.1272 - acc: 0.9542 - val_loss: 0.0549 - val_acc: 0.983
Epoch 2/5
842/842 [=====] - 102s 121ms/step - loss: 0.0393 - acc: 0.9873 - val_loss: 0.0422 - val_acc: 0.984
Epoch 3/5
842/842 [=====] - 71s 85ms/step - loss: 0.0153 - acc: 0.9950 - val_loss: 0.0414 - val_acc: 0.9856
Epoch 4/5
842/842 [=====] - 62s 73ms/step - loss: 0.0082 - acc: 0.9972 - val_loss: 0.0159 - val_acc: 0.9948
Epoch 5/5
842/842 [=====] - 55s 65ms/step - loss: 0.0047 - acc: 0.9987 - val_loss: 0.0181 - val_acc: 0.9952
```

Figure 4: Proposed Model Accuracy for Fake News Detection

Epoch refers to whole iteration of training data with proposed methodology. This may consist of more than one batches, with the number of epochs ranging from one to infinite. Figure 5 illustrates the five epochs which are calculates accuracy and loss for every single epoch. These epochs are calculated after the prediction. Among those the highest accuracy was achieved in 99.8% and loss was 0.01. Even though the time was just taken in 55sec.

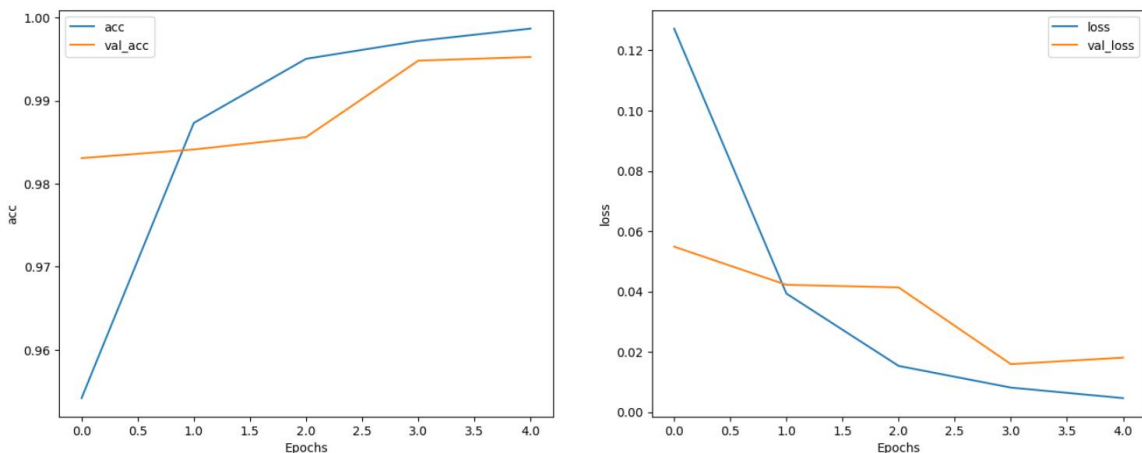


Figure 6: Accuracy and Loss of Proposed Model for Different Epochs

This is the graphical representation of five epochs one graph illustrate the accuracy and other loss. Figure 6 determines the accuracy and loss achieved by the epochs in five stages. In accuracy the graph was gradually

increased zero to 100 and value of accuracy was 98.5 to 99. In-same the loss was gradually decreased from 0.12 to 0.4 and parallely for val_loss were 0.06 to 0.01.

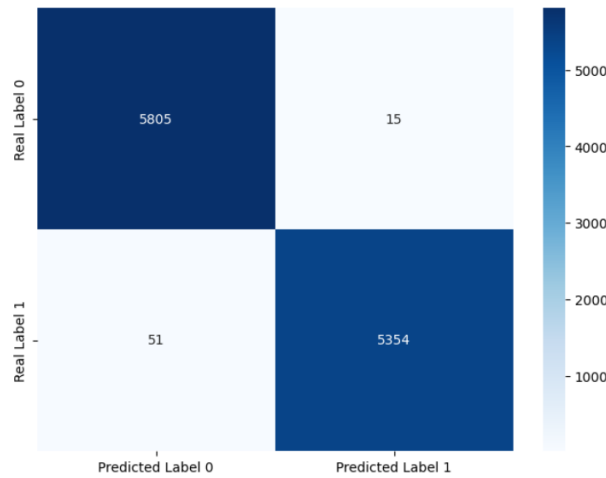
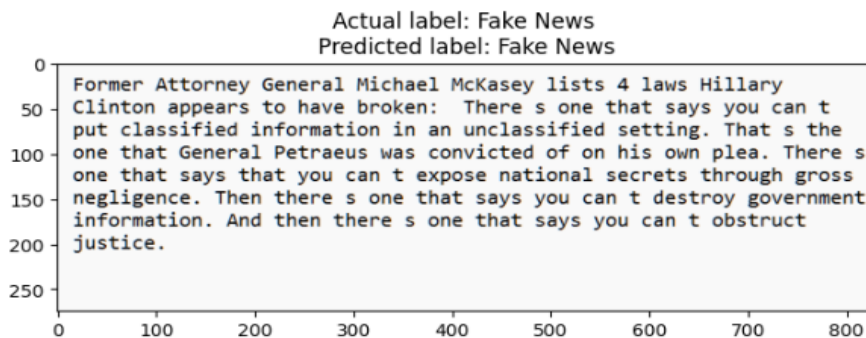


Figure 7: Confusion Matrix

Figure 7, Confusion matrix is a tabular form that provides a concise summary of a methodology performance in the field of ML. Figure x denotes the confusion matrix with prediction and real labels. In prediction label 0 and real label 1 then its valued is 51, when prediction and real is label 1 then its value is 5354, prediction is 1 and real labels 0 the obtained value is 15, and finally prediction and real labels is zero then the value is 5805.



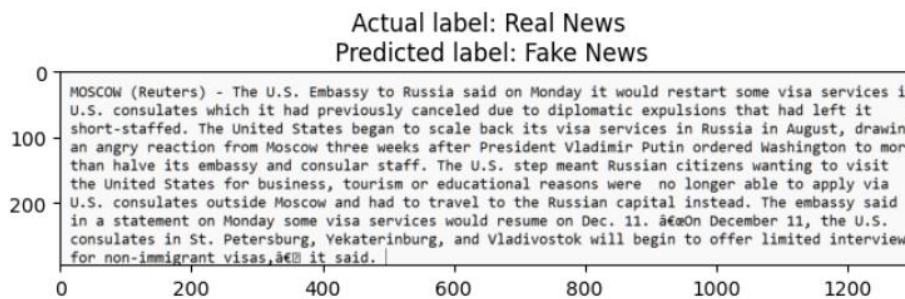
(a)



(b)



(c)



(d)

Figure 8: Outputs Obtained for Different News Articles

The analysed data was retrieved from different web-sites. The data holds with two labels actual and predicted. Actual label holds real or fake which was not clearly known but in prediction label will know the appropriate outcome. In figure 8(a), the actual data was fake and prediction is also same. In figure 8(b) the actual data is fake but the prediction is real. In figure 8(c) the actual and prediction are real news. In figure 8(d) the actual data is real and while prediction is fake news. These are the some of the predicted news from real-world statements.

V. CONCLUSION

This paper explores the fundamental components of NLP models that are essential for detecting false news. The vector representations generated by Word2vec enable the evaluation of semantic similarity, hence improving comprehension of context. Tokenization is the process of dividing text into smaller parts, which is essential for many natural language processing (NLP) tasks. Embedding layers transform categorical input into continuous vectors, streamlining the training process of neural networks. Dropout layers mitigate overfitting by promoting the model's resilience via the avoidance of dependence on individual neurons. Dense layers are capable of capturing complex data patterns, which allows for the generation of customised outputs that are suitable for certain applications. Together, these components provide a strong structure for creating effective NLP models that can differentiate between fake and real news, thereby improving the trustworthiness of information distribution. For any implementation examination of data is essential so, the data was considered from websites which contains fake and real. Evaluation was derived through epochs and confusion matrix. The accuracy was gradually increased and vice-versa for loss. The highest accuracy achieved by the epoch is 99.8%.

REFERENCES

- [1] A Hashmi, E., Yayilgan, S. Y., Yamin, M. M., Ali, S., & Abomhara, M. (2024). Advancing Fake News Detection: Hybrid Deep Learning With FastText and Explainable AI. IEEE Access, 12, 44462–44480. <https://doi.org/10.1109/ACCESS.2024.3381038>
- [2] Maheswari, R. U., & Sudha, N. (2024). An efficient fake news classification model based on ensemble deep learnintechiques. Salud, Ciencia y Tecnologia - Serie de Conferencias, 3. <https://doi.org/10.56294/sctconf2024649>
- [3] Kong, S. H., Tan, L. M., Gan, K. H., & Samsudin, N. H. (2020). Fake News Detection using Deep Learning. 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE). doi:10.1109/iscaie47305.2020.9108841
- [4] Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. Complexity, 2020, 1–11. <https://doi.org/10.1155/2020/888586>

- [5] Álvaro Ibrain Rodríguez Department of Computer Science (2019). FAKE NEWS DETECTION USING DEEP LEARNING. arXiv:1910.03496v2
- [6] Sahoo, S. R., & Gupta, B. B. (2020). Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, 106983. doi:10.1016/j.asoc.2020.106983
- [7] Nasir, J. A., Khan, O. S., & Varlamis, I. (2021). Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*, 1(1), 100007. doi:10.1016/j.ijime.2020.100007
- [8] Ning Zhang, Jiwen Liu (2021) Fake Financial News Detection with Deep Learning: Evidence from China Author links open overlay panelNing Zhang, Jiwen Liu. <https://doi.org/10.1016/j.procs.2023.07.022>
- [9] Kumar, S., Asthana, R., Upadhyay, S., Upreti, N., & Akbar, M. (2020). Fake news detection using deep learning models: A novel approach. *Transactions on Emerging Telecommunications Technologies*. doi:10.1002/ett.3767
- [10] Hiramath, C. K., & Deshpande, G. C. (2019). Fake News Detection Using Deep Learning Techniques. 2019 1st International Conference on Advances in Information Technology (ICAIT). doi:10.1109/ica.2019.89872
- [11] Kaliyar, R. K., Goswami, A., & Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimedia Tools and Applications*, 80(8), 11765–11788. doi:10.1007/s11042-020-10183-2
- [12] Agarwal, A., Mittal, M., Pathak, A., & Goyal, L. M. (2020). Fake News Detection Using a Blend of Neural Networks: An Application of Deep Learning. *SN Computer Science*, 1(3). doi:10.1007/s42979-020-00165-4
- [13] Truică, C.-O., Apostol, E.-S., & Karras, P. (2024). DANES: Deep Neural Network Ensemble Architecture for Social and Textual Context-aware Fake News Detection. In *Knowledge-Based Systems (Vol. 294, p. 111715)*. Elsevier BV. <https://doi.org/10.1016/j.knosys.2024.111715>
- [14] Luvenbe, A. M., Li, W., Li, S., Liu, F., & Wu, X. (2024). CAF-ODNN: Complementary attention fusion with optimized deep neural network for multimodal fake news detection. In *Information Processing & Management (Vol. 61, Issue 3, p. 103653)*. Elsevier BV. <https://doi.org/10.1016/j.ipm.2024.103653>
- [15] Zaheer, H., Bashir, M. Detecting fake news for COVID-19 using deep learning: a review. *Multimed Tools Appl* (2024). <https://doi.org/10.1007/s11042-024-18564-7>
- [16] Padmanabhuni, S. S. , & Gera, P. . (2023). EHFT: An Ensembled Hyperopt Fine-Tuned Neural Network for Disease Detection in Tomato Plants. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(9s), 634–646. <https://doi.org/10.17762/ijritcc.v11i9s.7478>
- [17] Abualigah, L., Al-Ajlouni, Y.Y., Daoud, M.S. et al. Fake news detection using recurrent neural network based on bidirectional LSTM and GloVe. *Soc. Netw. Anal. Min.* 14, 40 (2024). <https://doi.org/10.1007/s13278-024-01198-w>
- [18] Ganpat , R. R. ., & Ramnath , S. V. . (2024). OE-MDL: Optimized Ensemble Machine and Deep Learning for Fake News Detection. *International Journal of Intelligent Systems and Applications in Engineering*, 12(12s), 60–85. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/4492>
- [19] Datta, K. S., Naidu, G. N., Abhishek, S., & T, A. (2024). Enhancing Veracity: Empirical Evaluation of Fake News Detection Techniques. In *Procedia Computer Science (Vol. 233, pp. 97–107)*. Elsevier BV. <https://doi.org/10.1016/j.procs.2024.03.199>
- [20] Abd El-Mageed, A.A., Abohany, A.A., Ali, A.H. et al. An adaptive hybrid african vultures-aquila optimizer with Xgb-Tree algorithm for fake news detection. *J Big Data* 11, 41 (2024). <https://doi.org/10.1186/s40537-024-00895-9>
- [21] Divya, T. V., & Banik, B. G. (2021). An Empirical Study on Fake News Detection System using Deep and Machine Learning Ensemble Techniques. In *International Journal of Advanced Computer Science and Applications (Vol. 12, Issue 12)*. The Science and Information Organization. <https://doi.org/10.14569/ijacsa.2021.0121219>