J. Electrical Systems 20-7s (2024): 3988-3999

[1] Chinni Bala Vijaya Durga*

[2] Dr. Gatram Rama Mohan Babu

NBESM: A Novel BERT Integrated with Sequence Model for Classifying the News in Telugu

Journal of Electrical Systems

Abstract: - News classification helps individuals understand the political and social landscape of their region, enabling them to make informed decisions and participate in civic activities. Accurate classification ensures that readers can distinguish between news, opinion, and other types of content, maintaining the credibility and ethics of journalism. Many people prefer consuming news content in their native language. By classifying news in regional language, media outlets can cater to the language preferences of their audience. Traditional approaches have classified using Bag of Words and embedding approaches but Word embeddings are limited to the vocabulary seen during training. Words that were not present in the training data will have no pre-defined embeddings, and handling such out-of-vocabulary words can be challenging. Later, researchers started experimenting with neural networks such as RNN, GRU and LSTM. While LSTMs are designed to capture long-range dependencies, they can still struggle with very long sequences. As the sequence length increases, LSTMs may face challenges in retaining relevant information over extended periods. The proposed research integrates LSTM with transformers. Transformers excel at capturing contextual relationships across text, while LSTMs are designed to capture sequential patterns. This integration can be particularly useful for tasks that require modeling both local context and long-range dependencies.

Keywords: Recurrent Units, Text Classification, Regional Language, Bag of Words, Effective Communication

1. INTRODUCTION: Media organizations and advertisers can better target their content and advertisements to specific segments of the Telugu-speaking population based on their preferences and interest. News classification allows readers to filter and prioritize news articles based on their interests, saving them time by presenting relevant content upfront. Below section different types of text classification mechanisms and classification is presented in figure 1.

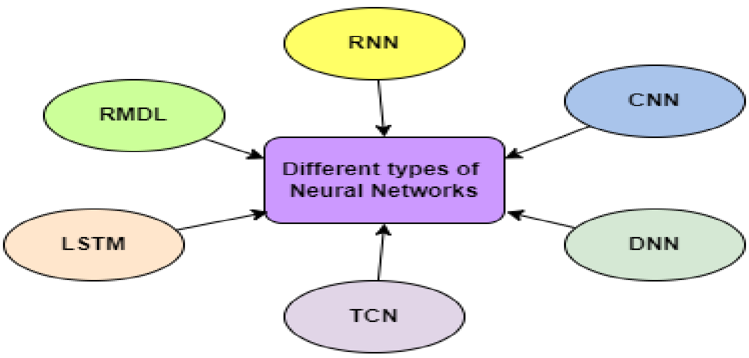

Figure 1: Categories Text Based Neural Networks

A. TCN: A dilated-causal variation of CNN is called a temporal convolutional network (TCN). It functions as a potent substitute for recurrent architectures since it can process lengthy input sequences without experiencing disappearing or ballooning gradient issues. B. DNN: Deep neural networks are built with several layers, each of which only accepts connections from the layers above it and only offers connections to the layers below it in the hidden portion. In essence, the Deep Neural Network (DNN) implementation is a discriminatively built model that uses ReLU or sigmoid as activation functions and the widely used back-propagation method. In the output layer of a multi-class categorization, Softmax should be utilised. C. RNN: The RNN gives earlier data points in the sequence more weight. This method therefore successfully categorises text, words, and sequential data. In RNN, the neural network uses a fairly complex way to take into account the data from earlier nodes, enabling a better semantic understanding of the dataset's structures.

[1] *Research Scholar, Dr. YSR ANU College Of Engineering & Technology, Department Of CSE, Acharya Nagarjuna University, Nagarjuna Nagar-522 510, Guntur (Dt.), A.P., India. 1balavijayadurga@Gmail.Com,

[2] Professor & Head, Dept. Of CSE (AI& ML), R.V.R. & J.C. College Of Engineering, Chowdavaram, Guntur - 522019, India E-Mail: 2 rmbgatram@gmail.com

*Corresponding Author: Chinni Bala Vijaya Durga *Research Scholar, Dr. YSR ANU College Of Engineering & Technology, Department Of CSE, Acharya Nagarjuna University, Nagarjuna Nagar-522 510, Guntur (Dt.), A.P., India. 1balavijayadurga@gmail.com,

D. LSTM: In contrast to standard RNNs, Long Short-Term Memory (LSTM) is a specific kind of RNN that retains long-term reliance more well. As LSTM employs several gates to precisely manage the amount of data that will be admitted into each node state, this is very helpful to address the vanishing gradient problem.E. CNN: CNNs have been successfully used to classify text with success. In order to reduce the computational complexity, CNNs use pooling, which prevents the quantity of output from one layer to the subsequent layers in the network. Utilising various pooling strategies, significant characteristics are preserved while outputs are reduced.
F. RMDL: Through ensembles of many deep learning architectures, RMDL (Random Multimodel Deep Learning) finds the ideal deep learning framework and architecture while simultaneously enhancing resilience and accuracy. Text, video, pictures, and symbols are just a few of the types of data that RDMLs may receive as input.

1.2. Different Existing Models for Text Classification:

a. CNN integrated with Mathematical Works:

Using the MaxWorth method, the suggested algorithm pulled from news the headline and the most important passage to verify. The problem of false news is not a new one, but in the past few years, the development of social media and the web has altered many of the notions around it, necessitating a reexamination. The research in this field has demonstrated that excellent results in tasks related to natural language processing may be obtained by carefully preprocessing the input text, vectorizing it, and feeding it to neural networks suitably constructed for sequence processing. Transformers, a novel neural network architecture that has just been released, have produced excellent results in NLP challenges. More data is immediately required to improve false news identification using deep neural networks. The context of the news can include social media comments and reviews, as well as non-textual news formats including images, audio, and video. The architecture of Mathematical works is presented in figure 2

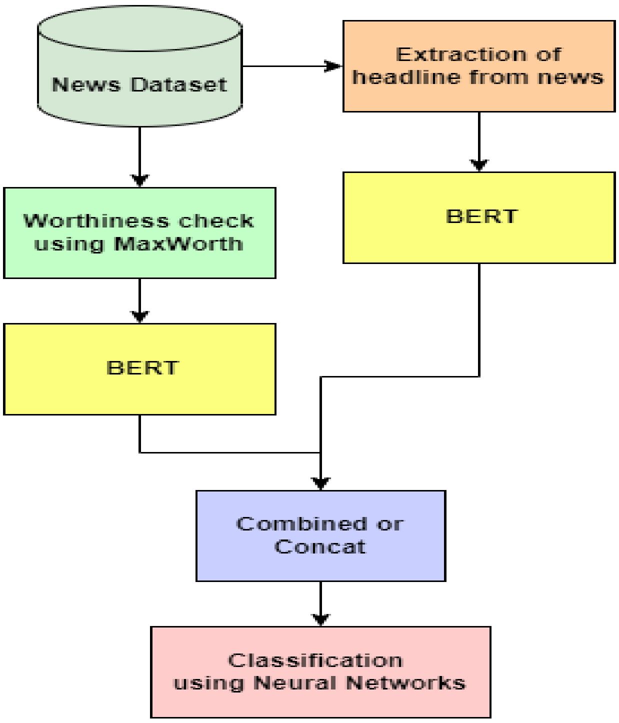

Figure 2: Mathematical Based Workflow

b. Hybrid Approach (Combination of Learning Algorithm with NLP):

The classification of fake news was accomplished in this work using deep learning (DL), machine learning (ML), and (NLP) natural language processing approaches.The phrase fake news is becoming more and more frequent. Through the use of polarising videos and viral pieces, misinformation is still rampant. The study presents a brand-new hybrid deep learning algorithm that combines recurrent and convolutional neural networks. As illustrated, two distinct datasets were integrated into one large dataset. The TF-IDF and Word2Vec methods tokenize the text as it is processed. On an open public dataset using content-level attributes, seven models have been evaluated to discover the best strategy between traditional machine learning (ML and neural network techniques. Each neural network model is more accurate than 90%. Figure presents the hybrid learning algorithms with NLP influence over it

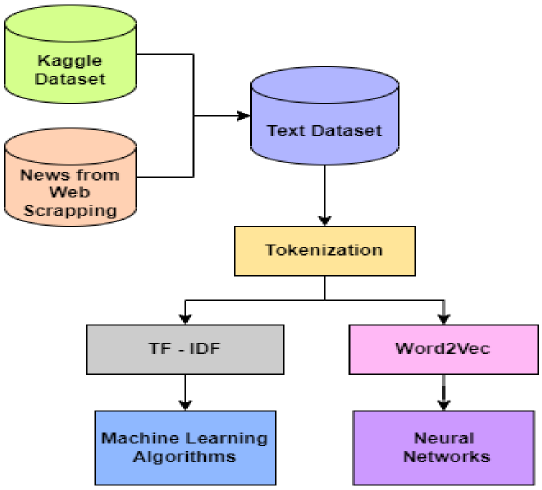

Figure 3: NLP influenced Hybrid Learning Models

c. Classification using ML-NLP

For detecting bogus news, an attention-based approach has been put out. Since the dawn of the internet era, fake news has grown significantly. The proportion of Internet consumers who make up the global population is rising annually and is at a high point right now, at 57%. The objective is to create a powerful deep-learning model that can accurately assess how false a news item is. With the use of convolutional neural networks-based deep learning ensemble architecture, Roy et al. achieved a new benchmark accuracy of 44.87%. The attention weights were computed using the speaker's key attributes. These qualities include the speech's subject, the speaker's profession, etc. An applied middle layer is the credit history. These techniques support the model's evolution. For improved outcomes, the hyperparameters are tweaked. Recall, accuracy, precision, and other performance metrics are used to validate the suggested model. Figure 5 presents the ML based NLP working

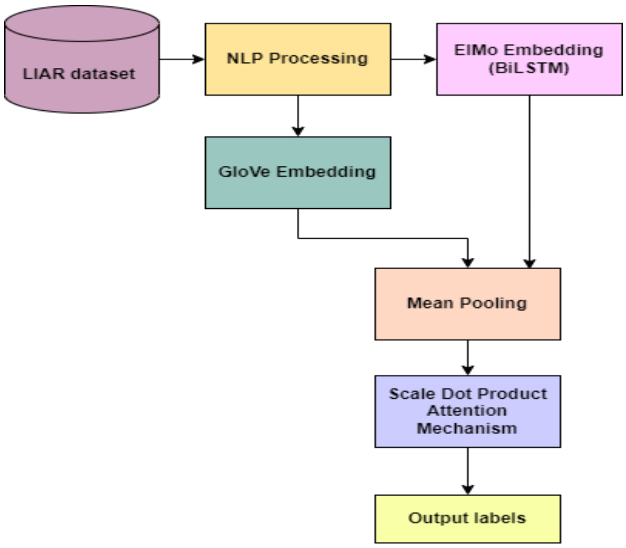

Figure 5: NLP integrated ML

1.3. Research Gaps Identified:

1. Most deep learning models require a large amount of labeled data for training. Research into few-shot or zero-shot learning approaches for text classification, where models can generalize from a limited number of examples, is an ongoing area of interest.
2. Deep learning models, including transformers, are often considered black boxes. Developing methods to make text classification models more interpretable and explainable is crucial, especially for applications requiring accountability and transparency.

2. LITERATURE SURVEY:

Nishant Rai et al. [1] used transformer-based improved LSTM and BERT to study bogus news classification. The BERT model combined with a layer of LSTM is used in this paper's proposal for the same method of categorising bogus news as either authentic or false news. The model's output has been assessed using the following criteria: precision, recall, accuracy, and F1 score. Misinformation spread by established sources like media organisations and social media is referred to as fake news. To put forth a method for identifying fake news by fusing the BERT framework with an LSTM that categorises stories as either phoney or real. For a thorough investigation on content-based categorization, hybrid models incorporating N-Gram, Word-Embeddings, and Subject Models were developed. The group suggests that grammatical and lexical characteristics of false news as found on social networks sites may be trained into the model as part of future study. Future testing of this architecture on many application domains is possible, and it could potentially enhance current benchmarks.

Yaser Wazery et al. [2] stated that a sequence-to-sequence-based abstractive Arabic text summarising method has been put out. The researchers discovered that the highest performance is obtained with a total of three layers of BiLSTM undetectable states at the encoder. a system for abstractively summarising Arabic text that use the sequence-to-sequence concept. Encoder and decoder are the two components that make up this system. The researchers employ several layers of long short-term memory (LSTM), bidirectional long short-term memory (BiLSTM), and gated recurrent units (GRUs). The group is eager to use algorithms for reinforcement learning and combine deep learning models with reinforcement learning methods. In the data pretreatment step, the AraBERT prepare has been used to produce cutting-edge outcomes by assisting the model in comprehending Arabic words.

Nankai Lin et al. [3] suggested a straightforward yet efficient technique for automated text summarization in Indonesian. Here, a method for extracting key phrases from Indonesian documents for use as summaries is suggested. It is based on the Light Gradient Boosting Machine. The goal associated with the linear regression is defined by the researchers as the formula for determining sentence score. The results of the trials show that the algorithm could solve the extractive summarization problem in the same way that Deep Learning and LightGBM did. The authors investigate how much various aspects impact the job. The researchers have put forth a multi-featured extractive summarization system for the Indonesian language based on the LightGBM regression model. Comparing the strategy to other abstract extraction models, it performs better in the F1 score of ROUGE-1 and RouGE-3. 56 publications were examined by the writers. According to the researchers, abstractive summarization performs poorly on the Indonesian automated text summarization test.

Arti Jain et al. [4] reported on real-coded genetic algorithm-based automated text summarising for Hindi. Sentence resemblance and named entity features, which are integrated with other features to compute the evaluation metrics, are among the top 14 feature differentiating characteristics among them. In order to calculate phrase ratings, one of the authors suggests summarising the Hindi text while mixing fuzzy and neural networks. Using the health corpus from the Kaggle dataset and Real Coded Genetic Algorithm (RCGA), the ATS approach is suggested for Hindi. An algorithm that uses selection and language to optimise the feature weights is suggested. By using created features, a real-valued/coded genetic algorithm is investigated and used for summarization of texts of the Hindi language. In the research, there were a total of 15 text papers with classifications. Jain and colleagues argue that other fields including banking, education, business, etc. can benefit from the automatic text summarising method. Other Indian languages including Bengali, Tamil, and Punjabi can also use the ATS system.

Mahmood Farokhianet al [5] reported on the use of parallel BERT neural networks with deep learning for bogus news identification. Veracity detection for full-text news stories is performed by MWPBert using two parallel BERT networks. The MaxWorth method is used by the researchers to put the news content into the BERT network. According to the authors' suggested model, the two BERT networks simultaneously encode the headline of the story and the news summary, and their output is then concatenated for categorization. Using the MaxWorth algorithm, the team identified the headline from the news and the most important passage to verify from the news's content. 22616 full news stories were examined by the researchers. Some of the outcomes make the assertion that past research in this area, which found that transformers have recently shown excellent results in NLP tasks, is strengthened. The researchers used the BERT network to produce one of the greatest findings in this study.

Ojas Ahuja et al. [6] indicate that the availability of substantial volumes of supervised data, such as the XSum and CNN/Daily Mail datasets, has assisted recent development in text summarization. The authors introduce ASPECTNEWS, a brand-new dataset for aspect-oriented summaries of news stories. With an emphasis on several aspects for each sub-domain, ASPECTNEWS is a collection of real-world aspect-oriented summaries. The team presents a keywordcontrolled system that was trained on fictitious data and demonstrates that, even without prior exposure to the target domains, it can successfully operate on ASPECT NEWS. The authors develop a system that, in the absence of annotations of training data regarding these aspects, can summarise a document based on a set of aspect-level keywords. The ASPECTNEWS dataset is used to test the model's performance against different baselines for aspect-oriented summarization. The study involved 42 participants.

Chun-Ming Lai et al [7] highlighted that the term "fake news" is growing in popularity among people. Using TF-Inverse Document Frequency and word embedding, modern ML and artificial neural network models based solely on content may categorise bogus news. On an accessible public database based on content-level attributes, the team compares seven models to see which one, between conventional ML and neural network techniques, performs the best. Each neural network model is more accurate than 90%. Through the use of polarising videos and viral pieces, misinformation is still rampant. The researchers used a variety of machine learning methods to illustrate and execute a fake news classification based on material level from a sizable publicly labelled corpus. 242 papers were examined by the researchers.

Weidong Zhao et al. [8] report that the weights created from feature location data are included in the pooling procedure. It is suggested to use a technique that utilises word2vec, a subject-based TF-IDF algorithm, and an enhanced convolutional neural network. To determine the topic of words, the programme employs a latent Dirichlet allocation model. The link between terms and their context was employed by the traditional word embedding method word2vec to train word vectors. It is suggested to improve the technique known as TF based on topic data, which creates the weight matrix by combining the word occurrence and the word's semantic information. 20000 articles were examined by the writers.

Santosh Mishra et al. [9] report that each research sector is publishing a sizable number of scholarly papers. Utilising a multifaceted differential evolution strategy, an innovative way of scientific document summarising has been developed. The team uses citation contextualization to first isolate significant sentences. Using the idea of multi-objective clustering, these phrases are further clustered. The organisation and separation of phrase clusters are measured by the XB index and PBM index, two objective functions. To create the best sentence clusters, a multi-objective clustering method based on the ideas of variations is used. For researchers, keeping up with the most recent advancements in their disciplines is becoming increasingly difficult. The researchers have put out a brand-new method for summarising scientific documents that makes use of citation contextualization. 40 citations were examined by the researchers.

Andy Ludwig et al [10] reported on a novel metaheuristic feature selection method for categorization that mimics human behaviour. The outcomes demonstrate that a baseline classifier's performance may be improved by using this technique with a smaller number of features. An approach for choosing features in text classification problems may be ant colony optimisation. Initialization, Learning, Consultation, and section Changing are the four primary iterative processes that make up the optimisation of individual performance. Depending on his unique feature set, the specialist offers the best function value. By utilising a brand-new metaheuristic feature selection technique, the researchers suggest a fresh categorization methodology. In order to perform 5-fold cross-validation, samples were created from the cleaned dataset. Each sample might be subjected to human behavior-based optimisation, producing an optimised feature set for every fold. The scientists looked at 1264 documents. In order to determine the ideal values for the hyperparameters, the researchers argue that further trials need be conducted. Using a more sophisticated term representation may lead to further gains.

Hozayfa Rifai et al. [11] put forward a multi-class classifier to assign Arabic news items to the correct class among the available 4. On the multi-labeled dataset, the authors put the suggested classifiers into practise and tested them. Multiple Arabic news sites have been used to create two sizable datasets. The first dataset consists of 90k articles from 4 domains that are single-labeled. There are over 290k multi-tagged items in the second dataset. To effectively address the problem of automatically labelling Arabic publications, the authors present a detailed analysis of a number of basic and deep learning techniques. The researchers have provided a single-label and multi-label tagging task-capable automated general text classification system for Arabic language. Support Vector Machines are comparable to its performance. The OVR-XGBoost classifier showed the greater accuracy in the shallow learning instance. In the investigation, 28 consonants were used. The authors argue that they will test several embedding models in the future, including BERT and ELMo. The goal of the study is to add more labels to the suggested datasets. The scientific community will be able to access all datasets, according to the authors.

Vidit Jain et al. [12] investigated that the false news has grown significantly since the dawn of the internet era, according to AENeT. Using contextual embedding and other relevant metadata, an effective deep learning network may determine how false a news item is. The organisation suggests an attention-based strategy for identifying bogus news. The accuracy obtained by the authors using the Elmo-enabled deep neural network model was 46.36%. The ultimate objective is to create a powerful deep learning model that can accurately assess how false a news item is. The group suggests a model architecture that draws connections between news assertions and metadata properties using attention processes. The research included 839 instances of pants-fire. The authors advise academics to encourage them to make use of the model for further studies in the field of false news recognition. Table 1 presents the mechanisms developed by existing approaches based on accuracy

Table 1: Comparative Study

Author	Method	Merits	Demerits	Accuracy
Nishant Rai	LSTM	The model has high learning capability	Need more instances for training to improve the	88.75%

			detection, may need manual check after the examination.	
Yaser Wazery	BiLSTM	Also investigated on the working of CBoW and skip-gram methods.	Can include reinforcement algorithms to improve quality,	Precision of 54.95
Nankai Lin	LightGBM	Simple	Features are selected manually and only work on a particular kind of data	F1 score is 0.730
Arti Jain	RCGA	Faster with the help of Simulating Binary Crossover and can be applied in multiple fields.	Cross-validation should be performed, and more features are need to be extracted.	F1 score is 87%
Mahmood Farokhian	MWPBert	Better extraction using Maxworth, GNN has been used	Can be improved to work on multiple languages.	85.4%
Ojas Ahuja	CNN	A new dataset called Aspectnews has been introduced, annoters have been used,	Complex and high time duration.	92.08%
Chun-Ming Lai	TF-IDF and word embedding	Higher accuracy	Most features are discarded, and can't get better contextual summarization.	97.76%
Weidong Zhao	word2vec + TF-IDF algorithm + CNN	LDA topic generation model improves the semantic values, improved pooling based on location.	Efficiency and time complexity can be improved.	95.76%
Santosh Mishra	Multi-Objective Clustering	Uses citation contextualization, uses two types of indices (XB, PBM)	Can be improved using various objective funtions, and decide the optimal count of clusters.	73.4%
Andy Ludwig	human behaviour-based optimization	Optimization at each step, reduction in training data because of optimization.	Can be influenced by different factors, and are unpredictable	88.4%
Hozayfa Rifai	OVR-XGBoost classifier	Faster results and fine-tuning is performed.	Usage of embedded models benefits the summarization, labels in the dataset are not enough to claim higher results.	94.85%
Vidit Jain	AENeT	Simple networks of contextual embedding with attention mechanisms can be trained even on large-sized datasets.	Didn't implement on real-time data, can use multi-modal approaches.	46.36%

3. PROPOSED METHODOLOGY: The proposed model uses the transformer model as the "encoder" to capture the contextual information of the entire text. Feed the output of the transformer as the input to an LSTM. The LSTM can then capture sequential patterns in the encoded context from the transformer. Figure 6 presents the work flow of the proposed model.

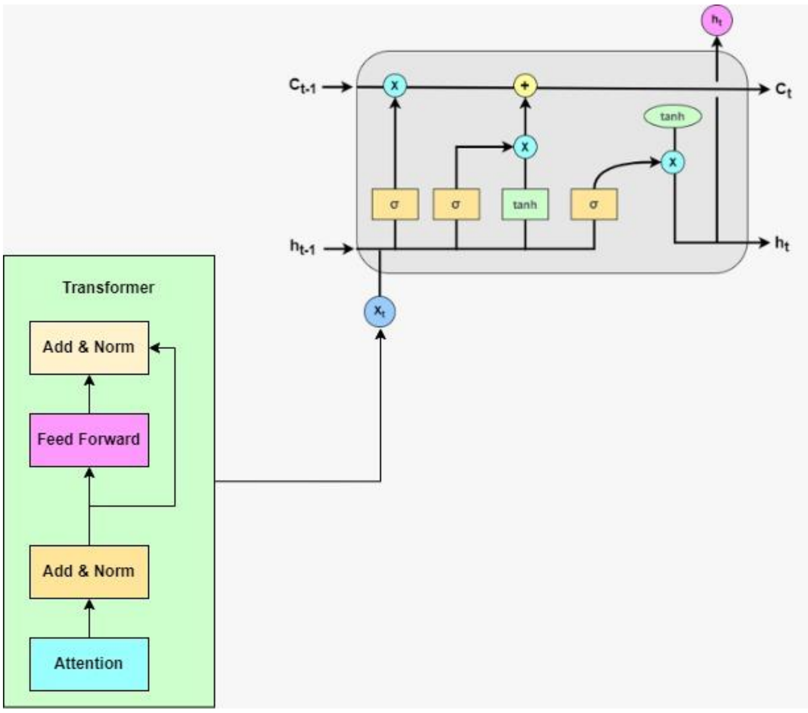

Figure 6: Integration of BERT with LSTM

3.1. Tokenization: Tokenization [13] is a method used in natural language processing to separate complex words and paragraphs into smaller, more easily understood parts. Tokenization entails breaking up the text in its raw form into smaller, more understandable chunks. Tokenization, or breaking up the original text into phrases and words, is done. These tokens help with context comprehension and NLP model building. Token occurrences in a document can be used directly as a matrix for illustrating the document. This quickly converts text or an unprocessed string into an integer representation that is appropriate for machine learning. Additionally, a computer can use them directly to start useful processes and reactions. They might also be included into a machine learning pipeline as attributes to start performing more complicated choices or actions. By examining the word order across the text, tokenization aids readers in deciphering the text's meaning. Sentence tokenization is a technical process that does the same for sentences. Word tokenization is a technological process that separates text into words. The proposed model implements BERT tokenization which employs the WordPiece tokenization algorithm. Instead of breaking text into whole words, WordPiece divides words into smaller subword tokens. This approach allows BERT to handle out-of-vocabulary words and capture morphological variations. Figure 7 presents the tokenization process

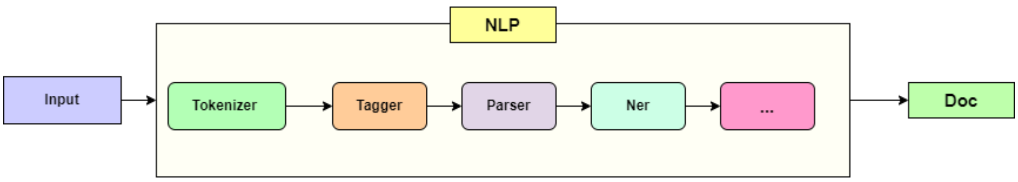

Figure 7: Process of Tokenization

3.2. Fine Tuned BERT for Text Extraction: A transformer is a deep learning architecture designed for sequential and structured data, particularly suited for natural language processing tasks. BERT is built on the transformer architecture, which includes self-attention mechanisms that allow the model to focus on different parts of the input text while considering their relationships. BERT is pre-trained on a massive amount of text data, using two main objectives:

A. masked language modelling

B. next sentence prediction.

After pre-training, the model can be fine-tuned on specific downstream tasks.

BERT generates contextualized word embeddings that take into account the entire sentence's context. This enables the model to understand word meanings based on their surrounding context. The transformer encoder is responsible for processing the input sequence, capturing contextual relationships between words, and generating meaningful representations of the input text. It's commonly used for tasks like text classification, named entity recognition, and more. The process of BERT as encoder is presented in figure 8

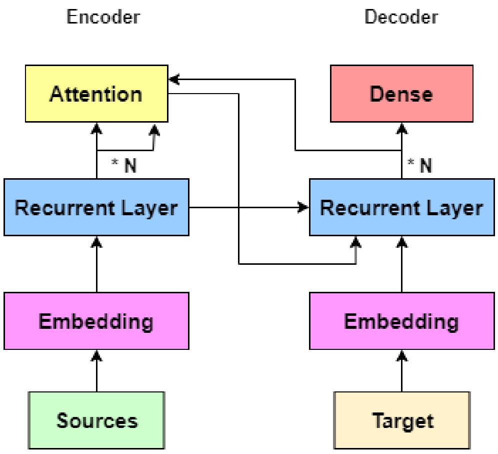

Figure 8: BERT as Encoding Transformer

The mathematical representation of the attention and other layers are presented in equation (1)

$$Attention(Q, K, V) = softmax(\frac{QK^t}{\sqrt{d_k}}) * V - 1(a)$$

$$Embedded[i] = \sum_{i=1}^{n} Sin(\frac{Position}{10000^{\frac{2*i}{disance}}}) + Cos(\frac{Position}{10000^{\frac{2*(i-1)}{distance}}}) - 1(b)$$

3.3. LSTM:

The LSTM has the ability for recalling every sequence of the data thanks to a characteristic. To help in text classification, it depends on the deletion of extraneous information. The LSTM network is a well-liked RNN for researching sequential data prediction challenges. The LSTM has a few layers, similar to other neural networks, which help in pattern detection and learning for better performance. An LSTM's basic operation may be thought of as retaining the information that is required while eliminating the data which is neither required nor useful for future prediction. The sequence through which the data is delivered can also be remembered by the LSTM. Since text data always contains a significant amount of unattended information that may be removed to speed up computation and save expenses, this final characteristic makes the LSTM a very useful tool for text categorization and other text-based tasks. The working of LSTM is presented in figure 9

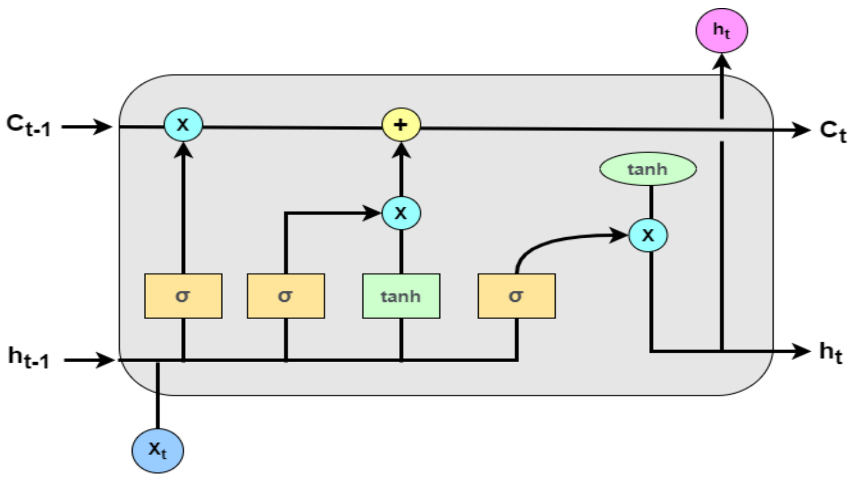

Figure 9: Working of LSTM

a. Forget gate

One of the LSTM's important features is its capacity to recall and identify information as it enters the network and to reject data which is not essential for the circuit to understand the input and generate predictions. This gate results in this LSTM feature. It helps in figuring out if data can move between network levels. It expects two separate types of input from the network: data from the levels that come before it and data at the layer of display. The working of forget gate is presented in figure 10

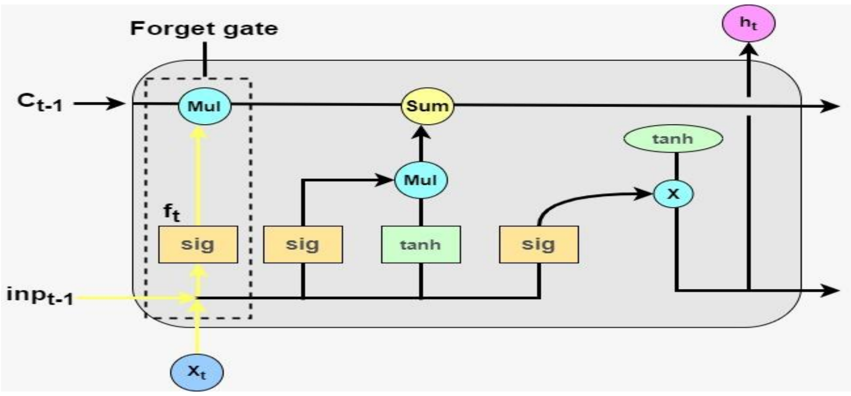

Figure 10: Forget Workflow

The mathematical representation of forget gate is presented in equation (2)

$$\boldsymbol{Forget\_Gate(Input) = \sigma(W(Input_{(h-1,t)})} \textbf{ +b) – (2)}$$

Where,
W = weight of the forget gate
h-1 = output of previous block at time t-1
t = input at current timestamp
b = bias for the forget gate
b. Input Gate
The input gate helps determine the importance of the information by altering the cell state. Other layers are helped in learning the data required for creating forecasts while the forget function assists in the removal of the data from the system. The forget function helps in the removal of the data from the network.
c. Cell State
The weight collected information passes to the cell state, that is calculated by this layer. In the cell state, the outputs of the input gate and forget gate are augmented by one another. Values that are extremely close to zero increase the risk of information disappearing. Figure 11 presents the flow of cell state execution. The mathematical notation is presented in equation (3)
$C_t = tanh(W_{(h-1,t)}) + C_{t-1} + b_c$ – (3)
C_t = cell state memory at time t
W = weight of the cell state
h-1 = output of previous block at time t-1
t = input at current timestamp
b_c= bias for the cell state

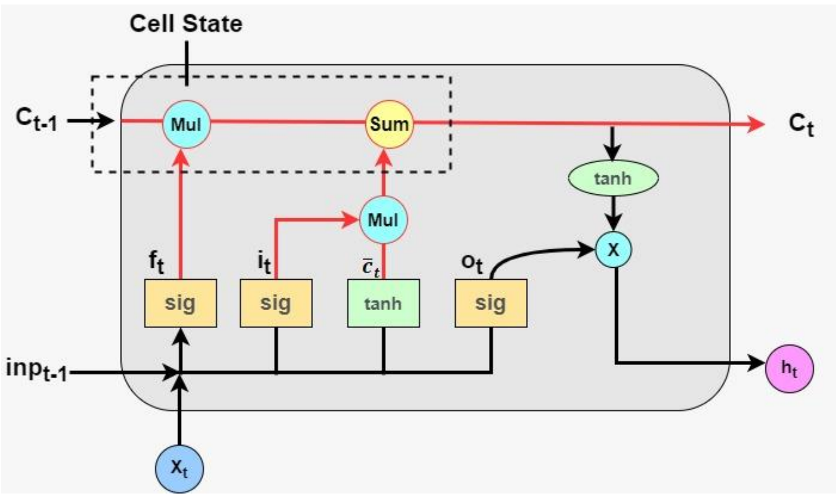

Figure 11: Flow of Cell State

d. Output Gate

The network's next hidden state, in which data is sent using the sigmoid function, is determined in part by the circuit's final gate. The output state's sigmoid function is multiplied by the tanh function, which is applied to the changed cell in the cell state. This aids the concealed state's ability to carry information.The components of LSTM are presented in figure 12

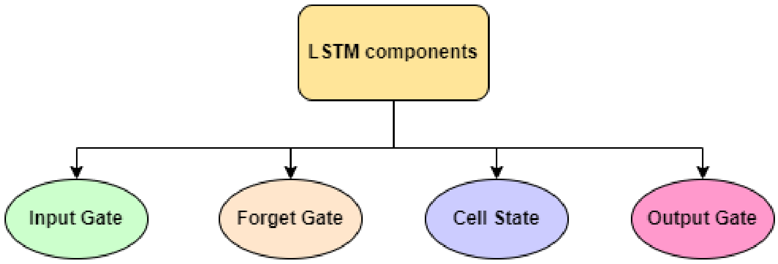

Figure 12: Different Gates in LSTM

4. EXPERIMENTAL RESULTS: Figure 13 represents the dataset utilized for the news classification in the proposed methodology. Every record contains the content with the topic name. The dataset separately contains training and testing records.

	body	topic
0	భారీ ఎత్తున మొండిబకాయిలు పెరిగిపోవడంతో ఐడిబిఐ ...	business
1	న్యూఢిల్లీ : ఆర్థిక మంత్రి అరుణ్ జైట్లీ సోమవా...	business
2	కటక్: ఇంగ్లండ్‌తో జరుగుతున్న సెకండ్ వన్డే మ్యా...	sports
3	\nఇస్లామాబాద్ : పాకిస్థాన్ అంతర్జాతీయ ఉగ్రవాది...	nation
4	స్టార్ హీరోగా వరుస సినిమాలతో బిజీగా ఉన్నప్పటి...	entertainment

Figure 13: Dataset Representation

Table 2 represents the numerical values of the class label because learning algorithms can work only on the numerical. Since the dataset contains 5 class labels i.e., multi classification, the class labels are sorted alphabetically and their representation is done using the binary bits based on the presence and absence

Table 2: Numerical Array from Vocabulary Generation

Class Label	Array Representation
Business	[1,0,0,0,0]
Sports	[0,0,0,0,1]
Nation	[0,0,0,1,0]
Editorial	[0,1,0,0,0]
Entertainment	[0,0,1,0,0]

Figure 14 represents the dimensionality of each layer with both trainable and non-trainable parameters. In this example, the shape of the input_data tensor is (20, 50), representing a batch of 32 examples, each with a sequence length of 20 words. The embedded layer transforms this input into a tensor of shape (20, 50, 100). Trainable parameters are updated through optimization algorithms like gradient descent, while non-trainable parameters are not. This makes it easier for practitioners to build and train models without manually keeping track of which parameters need to be updated.

```
Model: "sequential_20"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_20 (Embedding)    (None, 20, 50)            250000

 conv1d_17 (Conv1D)          (None, 16, 40)            10040

 global_max_pooling1d_13 (Gl  (None, 40)               0
 obalMaxPooling1D)

 dense_25 (Dense)            (None, 10)                410

 dense_26 (Dense)            (None, 5)                 55

=================================================================
Total params: 260,505
Trainable params: 260,505
Non-trainable params: 0
_________________________________________________________________
```

Figure 14: Model Representation of Proposed BETM Approach

Figure 15 presents the integrated model performance on both training and testing dataset. In the proposed model, the accuracy is efficient in both cases. The gradual improvement in accuracy represents the stability of the model

```
Epoch 1/20
433/433 [==============================] - 3s 6ms/step - loss: 1.4058 - accuracy: 0.4354 - val_loss: 1.1149 - val_accuracy: 0.6018
Epoch 2/20
433/433 [==============================] - 2s 5ms/step - loss: 0.8711 - accuracy: 0.7001 - val_loss: 0.7239 - val_accuracy: 0.7514
Epoch 3/20
433/433 [==============================] - 2s 5ms/step - loss: 0.5690 - accuracy: 0.8004 - val_loss: 0.5398 - val_accuracy: 0.8392
Epoch 4/20
433/433 [==============================] - 2s 5ms/step - loss: 0.3993 - accuracy: 0.8814 - val_loss: 0.4508 - val_accuracy: 0.8568
Epoch 5/20
433/433 [==============================] - 2s 5ms/step - loss: 0.3064 - accuracy: 0.9110 - val_loss: 0.4183 - val_accuracy: 0.8646
Epoch 6/20
433/433 [==============================] - 2s 5ms/step - loss: 0.2476 - accuracy: 0.9298 - val_loss: 0.4046 - val_accuracy: 0.8683
Epoch 7/20
433/433 [==============================] - 2s 5ms/step - loss: 0.2043 - accuracy: 0.9440 - val_loss: 0.3989 - val_accuracy: 0.8784
Epoch 8/20
433/433 [==============================] - 2s 5ms/step - loss: 0.1711 - accuracy: 0.9548 - val_loss: 0.4043 - val_accuracy: 0.8755
Epoch 9/20
433/433 [==============================] - 2s 5ms/step - loss: 0.1458 - accuracy: 0.9614 - val_loss: 0.4087 - val_accuracy: 0.8784
```

Figure 15: Model Analysis based on Epochs

5. CONCLUSION: News articles can be categorized into different topics or themes using LSTM-based classification models. This automated categorization can streamline the process of organizing and managing a large volume of news content. LSTM-based classification can be used to recommend relevant news articles to readers based on their interests, thus enhancing user engagement and personalization. LSTMs process input sequences sequentially, which makes them less suitable for parallel processing. This can result in slower training times compared to models that can take advantage of parallelism. Model can use BERT to extract features from text and then pass these features to an LSTM for sequence modeling. This way, LSTM focuses on capturing temporal patterns while benefiting from BERT's deep contextual representations. Explore advanced fine-tuning strategies for pre-trained models like BERT. Experiment with different learning rates, batch sizes, and optimization algorithms to find the best configuration for your specific task.

REFERENCES:

[1] Rai, N., Kumar, D., Kaushik, N., Raj, C., & Ali, A. (2022). Fake News Classification using transformer based enhanced LSTM and BERT. International Journal of Cognitive Computing in Engineering, 3, 98-105. https://doi.org/10.1016/j.ijcce.2022.03.003

[2] Y.M. Wazery, Marwa E. Saleh, Abdullah Alharbi, Abdelmgeid A. Ali, "Abstractive Arabic Text Summarization Based on Deep Learning", Computational Intelligence and Neuroscience, vol. 2022, Article ID 1566890, 14 pages, 2022. https://doi.org/10.1155/2022/1566890

[3] Lin, N., Li, J., & Jiang, S. (2021). A simple but effective method for Indonesian automatic text summarisation. Connection Science, 34(1), 29–43. https://doi.org/10.1080/09540091.2021.1937942

[4] Jain, A., Arora, A., Morato, J., Yadav, D., & Kumar, K. V. (2022). Automatic Text Summarization for Hindi Using Real Coded Genetic Algorithm. Applied Sciences, 12(13), 6584. https://doi.org/10.3390/app12136584

[5] Farokhian, M., Rafe, V., & Veisi, H. (2022). Fake news detection using parallel BERT deep neural networks. ArXiv. /abs/2204.04793

[6] Ojas Ahuja, Jiacheng Xu, Akshay Gupta, Kevin Horecka, and Greg Durrett. 2022. ASPECTNEWS: Aspect-Oriented Summarization of News Documents. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 6494–6506, Dublin, Ireland. Association for Computational Linguistics. 10.18653/v1/2022.acl-long.449

[7] Lai, C., Chen, M., Kristiani, E., Verma, V. K., & Yang, C. (2022). Fake News Classification Based on Content Level Features. Applied Sciences, 12(3), 1116. https://doi.org/10.3390/app12031116

[8] Weidong Zhao, Lin Zhu, Ming Wang, Xiliang Zhang &Jinming Zhang (2022) WTL-CNN: a news text classification method of convolutional neural network based on weighted word embedding, Connection Science, 34:1, 2291-2312, DOI: 10.1080/09540091.2022.2117274

[9] Mishra, S. K., Saini, N., Saha, S., & Bhattacharyya, P. (2021). Scientific document summarization in multi-objective clustering framework. Applied Intelligence. doi:10.1007/s10489-021-02376-5

[10] Ludwig, Andy, Jenny Felser, Jian Xi, Dirk Labudde, and Michael Spranger. "FoSIL at CheckThat! 2022: using human behaviour-based optimization for text classification." Working Notes of CLEF (2022).

[11] El Rifai, H., Al Qadi, L. &Elnagar, A. Arabic text classification: the need for multi-labeling systems. Neural Comput&Applic 34, 1135–1159 (2022). https://doi.org/10.1007/s00521-021-06390-z

[12] Jain, V., Kaliyar, R.K., Goswami, A. et al. AENeT: an attention-enabled neural architecture for fake news detection using contextual features. Neural Comput&Applic 34, 771–782 (2022). https://doi.org/10.1007/s00521-021-06450-4

[13] Chinni, Bala Vijaya Durga & Durga, Vijaya & Rama, Dr & Babu, Mohan. (2022). TELUGU TEXT SUMMARIZATION USING HISTO FUZZY C-MEANS AND MEDIAN SUPPORT BASED GRASSHOPPER OPTIMIZATION ALGORITHM (MSGOA). Journal of Theoretical and Applied Information Technology. 100. 5418 - 5432.