Roberta-LightGBM: A hybrid model of deep fake detection with pre-trained and binary classification

RAJKUMAR V[1]
PRIYADHARSHINI G[2]

Krishnasamy College of Engineering and Technology
Cuddalore, Tamilnadu
[1]raj_win7@yahoo.com
Dhanalakshmi College of Engineering and Technology
TamilNadu
[2]priyabeias151199@gmail.com

Abstract. In May 2023, a fabricated image depicting an explosion near the Pentagon garnered significant attention on social media, briefly impacting US markets. This incident potentially marked the first instance of an artificial intelligence (AI)-generated image influencing financial markets. Initially shared on Facebook, the misleading image portrayed a substantial column of smoke near the US military headquarters in Virginia, as claimed by a Facebook user. This research introduces the Roberta-LightGBM technique framework, combining the strengths of Roberta and LightGBM to address the challenge of detecting manipulated content in media with improved accuracy and efficiency. By leveraging a natural language processing (NLP) model in Roberta and a machine learning algorithm in LightGBM, our approach aims to swiftly identify and mitigate tampered-with content. The utilization of Roberta's NLP model enables the efficient training of large datasets in minimal time, presenting a notable advancement over traditional techniques like BERT, which necessitates substantially larger datasets for training across diverse applications. LightGBM, coupled with a decision tree, serves as a key component in our solution, enhancing training speed for handling extensive datasets with heightened accuracy and reduced memory usage. Through rigorous analysis, our proposed framework outperforms alternative techniques such as XGBoost. The Roberta-LightGBM technique achieves an impressive accuracy of 95.36%, demonstrating an overall computational time of 4.4 seconds. Experimental results in this paper showcase the implementation of Roberta, achieving an efficiency of 92.17%. This research establishes the effectiveness of our approach in efficiently identifying and combatting fake content, offering a promising solution in the realm of media integrity and accuracy.

Keywords: Accuracy, Fake News Detection, Efficiency, Natural Language Processing, And Machine Learning.

(Received December 12th, 2023 / Accepted June 28th, 2024)

1 Introduction

In the 2000s, newspapers and traditional media outlets were the primary sources of reliable information on events spanning the city, state, nation, and world. The advent of smartphones with expansive displays has transformed information accessibility, enabling people to instantly access global news. However, the rise of social networking platforms and chat apps has introduced the challenge of verifying the authenticity of news, leading to the dissemination of unverified information and potential consequences such as increased uncertainty and criminal activity in the community in world.

The advent of smartphones with expansive displays has transformed information accessibility, enabling people to instantly access global news. However, the rise of social networking platforms and chat apps has introduced the challenge of verifying the authenticity of news, leading to the dissemination of unverified information and potential consequences such as increased uncertainty and criminal activity in the community. Addressing the prevalence of deep fake digital material requires well-developed approaches. This digital content encompasses a wide range of forms, including films, artworks, images, and audio snippets. Even with a reliable mechanism to trace the history of digital material, accomplishing this goal remains challenging. As countermeasures against deep fake technology advance, there is a foreseeable evolution in the technology behind deep fake films. Establishing a system that ensures trustworthy data integrity and allows users to trace the origin and authenticity of digital content over time may provide a solution, protecting individuals from deception and manipulation, thereby fostering trust in digital content.

The BERT (Bidirectional Encoder Representations from Transformers) model, a creation of researchers at Facebook AI, is a subset of the RoBERTa (Robustly Optimized BERT Approach) model. Both models are transformer-based language models that process input sequences and generate contextualized representations of words using self-attention. Notably, RoBERTa surpasses BERT in efficiency, benefiting from a more efficient training method and a significantly larger dataset of 160GB compared to BERT's dataset. RoBERTa uses dynamic masking during training to improve the learning of accurate and flexible word representations. It does better than BERT and other models in a number of natural language processing tasks [3]. The integration method in machine learning has gained popularity, with ensemble algorithms proving more effective than individual decision tree algorithms. Booting and bagging methods, represented by AdaBoost, GBDT, and XGBoost, are primary clustering algorithms. XGBoost, known for its precision, speed, and noise-canceling abilities, has become popular in the machine- learning community. LightGBM, an upgraded version of GBDT, offers enhanced memory efficiency, faster training, and increased accuracy [20]. Text classification is a complex task, and machine learning algorithms are increasingly utilized to determine the veracity of text. Performance indicators are employed to compare and assess the effectiveness of various machine learning algorithms. This study focuses on analyzing textual characteristics to distinguish between false and legitimate news, enhancing precision through pre-processing with NLP approaches [5]. The proliferation of false information poses global social, economic, and political threats, emphasizing the crucial need to identify fake news in modern times. While researchers have employed machine learning techniques and neural networks to detect false news, challenges persist due to the low accuracy rates and intentional creation of misinformation to mislead readers [8]. Our contribution to deep fake content detection is outlined as follows:

- Theoretical Foundation and Design of Proposed Work: This research focuses on a comprehensive analysis of existing literature, identifying challenges in theoretical frameworks. The outcome is the design of the proposed Roberta-LightGBM technique, incorporating the robust Roberta Natural Language Processing (NLP) model and LightGBM to effectively detect fake content in global media.

- Experimental Techniques for Theoretical Analysis: This paper employs two experimental techniques to augment theoretical insights. The utilization of the Roberta NLP method enhances the detection of fake content with improved accuracy and efficiency.

- Memory-Efficient Implementation with LightGBM: The integration of the LightGBM machine algorithm results in reduced memory usage, particularly advantageous in applications dealing with large datasets.

- Enhanced Model Performance: The proposed model demonstrates high-speed performance, ultimately achieving the research goals outlined in this paper.

The subsequent sections of this research are structured as follows: Section 2: Deep Fake Detection Overview: This section provides an in-depth exploration of various deep fake detection methodologies, encompassing different machine learning approaches presented in existing literature. Section 3: System Framework and Method Implementation: Details regarding the system framework, the fine-tuned Roberta method, binary classification using LightGBM, and the practical implementation of the proposed technique are discussed. Section 4: Theoretical Setup and Experimental Design: This section outlines the theoretical foundation of the proposed work and delves into the experimental design. Section 5: Experimental Results and Analysis: The experimental results are presented and thoroughly analyzed in this section, providing insights into the performance and efficacy of the proposed model. Section 6: Conclusion and Research Goals Attainment: The research concludes with a comprehensive summary of the overall goals achieved, underscoring the significance of the proposed Roberta-LightGBM technique in advancing deep fake content detection.

2 Related Work

In this section, we delve into various research works related to deep fake content detection and related fields, providing insights into the methodologies employed and their respective outcomes.

Author uses a hybrid speech watermarking technique to show how digital watermarks can be inserted into a video's audio track. The detection of strong and weak watermarks can be done by

a standalone software program [2]. The terms "deep learning" and "fake" are combined to form the term "deep fake," which describes the emergence of convincing fake videos created by deep learning algorithms in easily accessible, publicly available software. These videos can spread misinformation by superimposing a target person's face on a source person's body without any discernible inconsistencies or distortions. A number of contentious deep fake films with a startlingly high level of realism have emerged online due to this form of extensively believable tempered footage, with substantial ramifications that lead to a large amount of societal discontent, serious injury, or death [3]. The purposeful production and dissemination of incorrect, modified material to mislead audiences is the major topic of this research author's description. Even though there have been several attempts to use clustering to solve the traditional classification problem of fake news detection, it remains a challenge [15]. The author primarily focuses on categorization methods. The current work uses two separate datasets to determine if the news is true or false using various classification methods, including SVM, DT, LR, and passive aggressive classifiers. In order to compare the effectiveness of the various algorithms, evaluation parameters were put into action. Later, to block the propagation of false information, they also try to determine whether the author has additional social media accounts. However, compared to our suggested approach, typical machine learning algorithms perform worse [1].

BERT-based (bidirectional encoder representations from transformers) and light gradient boosting machine (LightGBM) models are combined to produce a unique hybrid false news detection system. To verify the effectiveness of the suggested method in comparison to other methods, the author compares the proposed method39;s performance to four alternative classification approaches employing various word embedding techniques on three real-world fake news datasets. Based on the headline-only or complete text of the news material, this existing study is reviewed to identify false news. The findings demonstrate the suggested method's superiority over various state-of-the-art approaches for detecting false news. However, when compared to Bert's technique, Roberta delivers well-featured work that is studied in this paper's experimental portion [6].

This study describes the prediction of song popularity tasks. Looking back, research should lack some feature accuracy, feature importance, and types. This researcher can implement their framework based on LightGBM and logistic regression to predict the accuracy of song popularity. So the author gathered some challenges in front of the sites to multi-model extraction structure-based design their research work. LightGBM can attain better accuracy, but compared to the logistic regression algorithm, our proposed Roberta NLP model performs well [9]. A methodology for detecting fake news called DEAP-FAKED is knowledge-based. In order to encode news material and embed Knowledge Graph (KG) elements, this work uses a natural

language processing (NLP) and tensor decomposition model, respectively. A diversity of different encodings gives the detector a complementary benefit. Using two publicly accessible datasets that include articles from industries including politics, business, technology, and healthcare, this researcher evaluates the framework [12]. Intrusion detection in a generic algorithm to describe the new proposed work optimized LightGBM, which included some additional features like a recursive feature elimination algorithm and weight loss to control the insufficient balance of network traffic data [23]. This optimized LightGBM can use the CIC- IDS2017 dataset to test the experiment. The result would achieve high detection accuracy [13].

An investigation of machine learning ensemble techniques for diabetes forecasting. A recently developed light gradient boosting machine is included in the ensemble along with k-NN, Naive Bayes (Gaussian), Random Forest (RF), Adaboost, and other models. This article provides ensembles with inherited LightGBM detection capabilities to increase accuracy. This research ensemble model study outperforms other current models in fivefold cross-validation. Together, the k-NN, Adaboost, and LightGBM produce a detection accuracy of 90.76operating curve analysis shows that k-NN, RF, and LightGBM are all good at fixing the underlying dataset's class imbalance problem.

Nonetheless, it should have a lower accuracy rate than our framework [16]. During the COVID- 19 pandemic, people should wear facemasks when coming outside, which would be detected in the complex hyper-realistic face image generation. The author describes their proposed work to make it easy to detect face-masks. This deep fake detection approach predicts a fake facial image with a facemask tampered with [17].

The fault prediction in the fire control system to describe the LightGBM prediction model can be proposed to improve the fault prediction [18], and in this study, the research author proposed the light convolutional neural network (LCNN) model for the audio deep synthesis detection (ADD) task. The author adds input features in the front end, such as constant Q transform and low- frequency short-time Fourier transform. In technique synthesis, the unwanted noise audio track is predicted using training data values [21]. Challenges Identified from Existing Research:

- Comparative analyses indicate that the proposed Roberta-LightGBM approach achieves better accuracy with minimum memory usage than various machine learning algorithms, including logistic regression, K-NN, SVM, DT, Passive-Aggressive Classifier, Naive Bayes, RF, and AdaBoost.

- While some research utilizes LightGBM for binary classification to reduce memory usage in large datasets, our proposed approach outperforms in terms of both accuracy and dataset processing.

- Existing studies face challenges related to dataset size, accuracy trade-offs, and computational/training speed, which our proposed work successfully addresses.

- This comprehensive review sets the stage for our research by highlighting the strengths and limitations of existing approaches in the field of deep fake content detection and related domains.

3 Proposed Work

In the current era, the widespread dissemination of fake news presents a formidable challenge, capable of deceiving individuals, causing harm to businesses, and even influencing entire nations. Conventional methods of identifying false information often rely on rule-based or manual fact-checking processes, characterized by time-intensive efforts and limited coverage. This study proposes an innovative approach to elevate the precision and efficacy of false news detection, harnessing the power of refined Roberta and LightGBM methodologies. The methodology strategically combines LightGBM's efficiency in optimizing feature space and classification with Roberta's ability to capture intricate language patterns. The aim is to surpass existing approaches by seamlessly integrating these two methodologies for enhanced synergy. Figure 1 provides an insightful overview of the proposed system. The process initiates with pre-processing operations on the provided text, eliminating irrelevant information. Tokenization further dissects the input text into individual characters, subwords, and words, optimizing it for representation in the subsequent "fine-tuned" Roberta model. The extraction of text embeddings from the last three hidden layers of Roberta's distinctive token [CLS] enhances the model's comprehension of language nuances. In the next step, the LightGBM model is trained on the combined embedding vectors, using how they all represent the input text. This holistic approach, merging the strengths of both Roberta and LightGBM, aims to attain superior performance in false news detection.

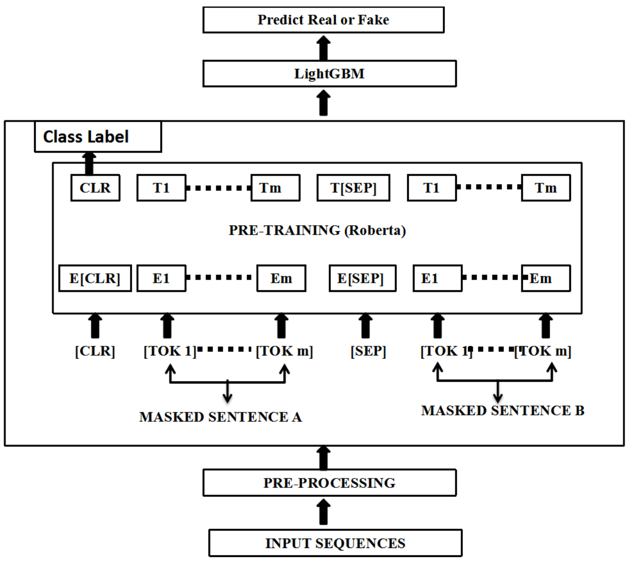

Figure 1: proposes a framework in the Roberta-LightGBM technique.

Before each figure must have a space of 6 points. Between the figure and the legend, a space of 3 points. After the legend must have a space of 6 points. The legend must be in Times or Times New Roman, size 9 points.

3.1 Roberta Fine-tuned the technique:

The Bidirectional Encoder Representation (BERT) from Transformers is expanded upon in the RoBERTa model [10]. The Transformers family, which was created for sequence-to-sequence modeling to solve the issue of long-range dependencies, includes the BERT and RoBERTa. Transformer models are made up of three parts: tokenizers, transformers, and heads. The tokenizer turns the unprocessed text into sparse index encodings. The transformers then transform the sparse material into contextual embedding for more in-depth training. To enable the usage of contextual embedding for the downstream activities, the heads are developed to surround the transformers model. BERT differs from the current language models in that it can pick up on contextual representation from both ends of sentences. BERT used 30K character- level Byte-Pair Encoding words for tokenization [11]. RoBERTa, on the other hand, utilized a bigger vocabulary set that consisted of 50K subword units and a byte-level Byte-Pair Encoding. In addition, the RoBERTa model improves the BERT model by fine-tuning it through longer training times, longer sequences, and more data. RoBERTa was trained using the following four corpora:

- The 16GB of text in the BOOK CORPUS and English Wikipedia datasets is used for training the BERT architecture.
- CC-NEWS. Between September 2016 and February 2019, 63 million English news articles were crawled and included in this data. After filtering, this dataset weighs 76 GB.
- OPENWEBTEXT: This dataset includes web material that was taken from Reddit URLs that received at least three upvotes. This dataset weighs in at 38 GB in size.
- STORIES: This dataset consists of a portion of Common Crawl data that has been filtered to correspond to the Winograd NLP task's narrative style. Text in this collection totals 31 GB.

The RoBERTa tokenizer is employed in this study. The "s" the "s" token to mark the beginning and conclusion of the sentence, the "s" the "s" and the "s" and the "s" the, and the "s" respectively, and the "s" the text to make it longer than the word vector's maximum length. The Byte-Pair Encoding tokenizer is used in the RoBERTa paradigm to break up the contents into subwords at the byte level. The tokenizer won't divide apart terms that are used frequently. However, the infrequently used terms will be divided into subwords. For example, "Transformers" will be divided into "Transform" and "ers." It is necessary to convert the text's words into meaningful numerical representations in order for the model to comprehend them. The raw text is tokenized using input ids and an attention mask by the RoBERTa tokenizer. The input ids stand in for the token indices and its numerical representation. On the other hand, the attention mask is utilised as an optional input to group the sequence together. The attention mask indicates which tokens require attention and which do not. The attention mask and input ids are supplied to the RoBERTa base model. The attention mask and input ids are supplied to the RoBERTa base model. The RoBERTa basic model has 125 million parameters, 12 RoBERTa core layers, and 768 hidden state vectors. The goal of the RoBERTa initial layers is to provide a relevant phrase to represent features so that the subsequent layers may quickly collect the insightful data from the phrase embedding. RoBERTa is a strong and successful language model that has significantly advanced the field of NLP and aided development in a variety of applications. The robustly Optimized BERT Pre- training Approach is known as RoBERTa. Researchers from Facebook and Washington University presented it. This study aimed to accelerate BERT architecture pre-training by optimizing the training process. Although RoBERTa's architecture is identical to that of BERT, the authors made a few minor adjustments to the training process and architecture to improve the outcomes for BERT. These adjustments are: Remove the Next Sentence Prediction (NSP) objective: The model is trained to predict, using an auxiliary Next Sentence Prediction (NSP) loss, whether the observed document segments originate from the same or different documents. The authors tested numerous versions with and without NSP loss, and they came to the conclusion that doing so matches or slightly enhances the performance of downstream tasks. Training with larger batch sizes longer sequences: Initially, BERT is trained for 1M steps with a batch size of 256 sequences. In this research, the model was trained using 125 steps of 2K sequences and 31K steps of batch size 8K sequences. This has two benefits: the large batches increase end-task accuracy and confusion on the masked language modeling objective. Additionally, distributed parallel training makes it simpler to parallelize large batches. The masking pattern can be changed dynamically in the BERT architecture, where the masking is only ever done once during the data pre-processing stage, resulting in a single static mask. Training data is replicated and masked ten times using a different mask strategy over 40 epochs to avoid utilizing a single static mask. This results in 4 epochs that use the same mask. This tactic is contrasted with dynamic masking, where a new mask is created each time input is passed into the model. Transfer learning, which enables employing a previously trained Roberta on big datasets in another specialized job using the fine-turning approach, is one of the keys to Roberta characteristics. The BERT settings for the false news detection job are being fine-tuned in this instance.

Roberta uses the final hidden state FHS of the initial token [CLS] to represent the entire sequence. A layer that is fully connected with a softmax classifier has been added to the Roberta to refine it, classifying the vector FHS as real or false. The completely linked layer has the dimension [768, 2]. Using the softmax function, the resultant layer of the framework forecasts the likelihood of label C(R.)

$$p(R|\, FHS\,) \;=\; softmax(M\, FHS) \tag{1}$$

Where M is the newly added fully linked layer's parameter matrix, the correct label's negative log probability is minimized to adjust all the Roberta and M parameters.

3.2 LightGBM Algorithm:

The Train Using AutoML tool employs LightGBM, a gradient-boosting ensemble technique that is based on decision trees. LightGBM is a decision tree-based technique that may be applied to both classification and regression problems. For excellent performance with dispersed systems, LightGBM has been specially designed [22]. Given a condition, LightGBM generates decision trees that develop leaf-wise, which implies that, depending on the gain, just one leaf is split for each tree, given the condition. Sometimes, especially with smaller datasets, leaf-wise trees overfit. Overfitting can be prevented by limiting the depth of the tree. Data is divided into categories using a histogram of the distribution in the LightGBM histogram-based technique. The bins are utilized instead of iterating, calculating the gain, and dividing the data. A sparse dataset can also benefit from this method's optimization.

Exclusive feature bundling, which refers to the algorithm's combination of exclusive features to minimize dimensionality and speed up processing, is another aspect of LightGBM [7]. A weak learner ensemble model called a gradient-boosting decision tree (GBDT) is based on decision trees that have been successively trained. By fitting the negative gradients (sometimes referred to as residual errors), the GBDT trains the decision trees in each iteration. One way to describe the GBDT model $\emptyset\ (x)$ is as a collection of decision trees,

$$\emptyset\ (x) = \sum_{n=1}^{N} \delta_n\, T\,(x;\ \theta_n)\,, \tag{2}$$

Where ï¿½ā$(x;\ \theta_n)$ is the n-th decision tree, N is the total number of trees in the algorithm, δ_n is the learning rate, and m is the tree's parameters. By minimizing the loss function L_f with regard to the tree parameters n, the n-th tree is trained to forecast the residual error.

$$\theta_n = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{N} L_f\ (y_{i,}\ \ \emptyset_{n-1}\,(t_i) + \delta_n\ \ T\,(x;\ \theta_n)\ \tag{3}$$

Where N is the number of training samples, $\emptyset_{n-1}\,(t_i)$ is the prediction made by the prior tree, and $y_{i,}$ is the target variable. Gradient descent is a common optimization technique where the gradient of the loss function is calculated with respect to the tree's parameter values. When working with big volumes of data, conventional GBDT takes a long time. LightGBM is an effective and scalable GBDT implementation that speeds up training while maintaining respectable accuracy. The process of creating a decision tree is what causes standard GBDT to be computationally expensive. In order to choose a feature as a split point that maximizes information gain, it is necessary to scan all the data instances. In order to get around this restriction, LightGBM is introduced. Gradient-based In LightGBM, the dataset is sampled using one-side sampling (GOSS). GOSS gives more weight to data points with bigger gradients when computing the gain. Instances that have yet to be used effectively for teaching more in this way. Certain data points are randomly deleted from the analysis to preserve accuracy and others are kept. Given the same sample rate, this strategy is often superior to random sampling. Different data instances play a variety of functions in the information gain calculation. The information gain will be greater for the examples with bigger gradients (i.e., under-trained cases). To maintain the accuracy of data gain estimation, GOSS retains examples with big gradients (greater than a predetermined threshold or in the top percentiles, for example) and only sometimes discards cases with tiny gradients. When the amount of data gain has a wide range, this method can result in a gain estimation that is more accurate than equally random sampling at the same goal sampling rate.

4 Experimental setup

4.1 Datasets

The "Fake News detection dataset" comprises articles labeled as "fake" or "real," gathered from reuters.com news updates. Two CSV files, true.csv and fake.csv, contain 12,600 articles each, collected between 2016 and 2017. In below table shows the fake and real files in Figure: 2 and Figure: 3

Donald Trump Sends Out Embarrassing New Year's Eve Message; This is Disturbing	Donald Trump just couldn t wish all Americans a Happy New Year and leave it at that. Instead, he had...	News	December 31, 2017
Drunk Bragging Trump Staffer Started Russian Collusion Investigation	House Intelligence Committee Chairman Devin Nunes is going to have a bad day. He s been under the as...	News	December 31, 2017
Sheriff David Clarke Becomes An Internet Joke For Threatening To Poke People 'In The Eye'	On Friday, it was revealed that former Milwaukee Sheriff David Clarke, who was being considered for ...	News	December 30, 2017
Trump Is So Obsessed He Even Has Obama's Name Coded Into His Website (IMAGES)	On Christmas day, Donald Trump announced that he would be back to work the following day, but he i...	News	December 29, 2017

Figure 2: Fake datasets collected from Kaggle

As U.S. budget fight looms, Republicans flip their fiscal script	WASHINGTON (Reuters) - The head of a conservative Republican faction in the U.S. Congress, who voted...	politicsNews	December 31, 2017
U.S. military to accept transgender recruits on Monday: Pentagon	WASHINGTON (Reuters) - Transgender people will be allowed for the first time to enlist in the U.S. m...	politicsNews	December 29, 2017
Senior U.S. Republican senator: 'Let Mr. Mueller do his job'	WASHINGTON (Reuters) - The special counsel investigation of links between Russia and President Trump...	politicsNews	December 31, 2017
FBI Russia probe helped by Australian diplomat tip-off: NYT	WASHINGTON (Reuters) - Trump campaign adviser George Papadopoulos told an Australian diplomat in May...	politicsNews	December 30, 2017

Figure 3: true datasets for fake news detection

4.2 Prepare the Dataset for Training with LightGBM:

1. Dataset Selection: Let $Dataset\ = \{x_i,\ y_i\}$ represent the dataset, where x_i is the feature vector for article i and y_i is the corresponding label (0 for real, 1 for fake). $Dataset$ is obtained from the "Fake News detection dataset" with true.csv and fake.csv.

2. Data Organization: The dataset is well-organized: $Dataset\ = \{x_i,\ y_i\}$ with clear labels.

3. Labeling:

$$y_i = \begin{cases} 0, & if\ article\ i\ is\ real\ news \\ 1, & if\ article\ i\ is\ fake\ news \end{cases}$$

4. Data Quality Check: Ensure $Dataset$ is clean: $\forall_i,\ x_i = null$ and y_i is well-defined.

5. Data Split: Split the dataset into training $Dataset\ _{train}$ and validation $Dataset_{val}$ sets: $Dataset\ = Dataset_{train} \bigcup\ Dataset_{val}$.
Commonly $Dataset\ _{train} \approx 80\%$ and$Dataset_{val} \approx 20\%$.

6. Feature Inclusion: Combine features: $x_i\ = features\ from\ Roberta\ +\ additional\ features.7.$ Ensuring Balance: Ensure class balance: $\frac{\sum_i y_i}{len(dat\ddot{i}_{¿}\frac{1}{2}\hat{a}set)}$is balanced.

8. Data Format Compatibility: Verify dataset format: $Dataset\ = \ \{(x_i,\ y_i)\}$ conforms to LightGBM requirements.

4.3 Convert Text Features and Other Relevant Features:

Converted text features obtained from Roberta, such as fixed-size vectors, along with other relevant features. Integrated metadata, source information, and potential structural features into the feature set.

4.4 Binary Classification Labels:

Assigned binary classification labels to each article:
0 for real news
1 for fake news

4.5 Feature Matrix Creation:

Constructed a feature matrix by combining text features and other relevant features. Ensured the feature matrix is structured appropriately for input to LightGBM.

4.6 Training and Validation Sets:

Split the dataset into training and validation sets, typically with a percentage split (e.g., 80% for training, 20% for validation). The training set is used to train the LightGBM model, while the validation set assesses its performance.

4.7 LightGBM-Compatible Format:

Prepared the feature matrix and labels in a format compatible with LightGBM's data structure. Converted the data into a LightGBM Dataset for efficient training.

4.8 Hyperparameter Considerations:

Decided on hyperparameters suitable for LightGBM, considering factors such as learning rate, tree depth, and the number of iterations. These parameters influence the model's learning and generalization abilities.

4.9 Cross-Validation:

Conducted cross-validation, if applicable, to fine-tune hyperparameters and evaluate the model's performance. This step helps ensure the robustness of the model by assessing its performance across different subsets of the data.

5 Evaluation of Result Analysis:

The experimental results are analysis based on two NLP models, BERT and Roberta, using processing task Glove, SQUAD 1.0, SQUAD 1.1. From this experiment analysis, BERT has less accuracy percentage compared to the Roberta model. That is shown in the below table: 1.

Table 1: comparative of two NLP model using processing task

Processing task	BERT Model	Roberta Model
GLovE	80.5 %	**92.3%**
SQUAD V1.1	93.2%	**96.73%**
SQUAD V1.2	83.1%	**95.43%**

The below figure: 4 can analysis the best model of a pre-trained process using the processing task such as Glove, SQUAD 1.1, and SQUAD 1.0 models to get a good processing task based on their analytic value.

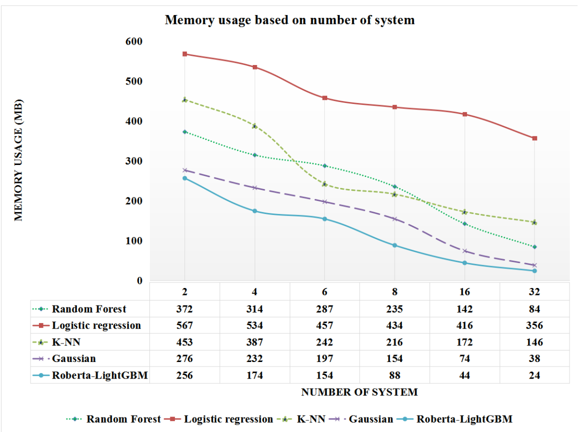

Figure 4: Compared the traditional Pre-trained BERT to the Roberta model

Evaluation metrics: We employ Auc-Roc as the metric to assess the effectiveness of our technique. The model will perform better with a higher Auc-Roc Score and accuracy [19]. To measure the classification performance using macro F1 score. It is harmonic mean calculation using precision and recall of the model measured.

$$F_1 = \frac{2}{(recall^{-1}) * (precision^{-1})} \quad (4)$$

The primary assessment metric for models is accuracy, which counts the proportion of correct predictions among all forecasts.

$$Acc\,(Fake_{news}) = \frac{RP + RN}{RP + RN + FP + FN} \quad (5)$$

Precision is a measure of the proportion of correctly predicted favourable outcomes.

$$Precision = \frac{RP}{RP + FP} \quad (6)$$

Recall measures the proportion of correctly detected true positives.

$$Recall = \frac{RP}{RP + FN} \quad (7)$$

The trade-off between the real-positive rate and the fake-positive rate at various threshold settings is illustrated by the receiver-operating characteristic (ROC) curve. By computing the area under the ROC curve, AUC condensed the efficiency of the predictor into a single metric. The AUC value is between 0.5 and 1. AUC scores range from 0.5 for a random guess predictor to 1 for a perfect predictor. Calculating the accuracy, precision and F1 score in percentage to analysis are shown in Figure 5 below.

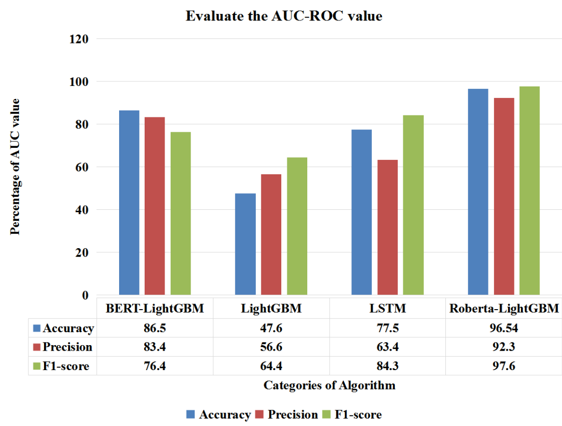

Figure 5: Evaluate the AUC-ROC value

Analysis of the memory usage in our proposed work and time taken to computation based on the number of

Table 2: Time and memory usage in per machine

Number of machines	Time Taken per computational	Memory Usage
1	627.8 sec	172 GB
2	311 sec	87 GB
4	156 sec	43 GB
8	80 sec	22 GB
16	42 sec	11 GB
32	24 sec	6 GB

machines connected to done the process. In below table: 2 shows the linear speedup of the work. The existing traditional fake news detection method is analysis with our proposed work of Roberta-LightGBM as shown in Figure: 6.

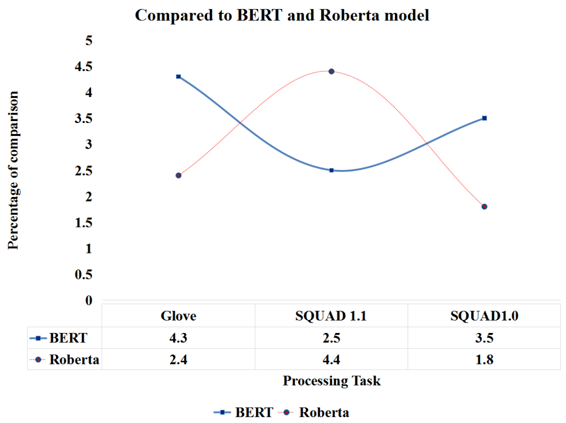

Figure 6: Memory usage based on the Number of Systems

6 Conclusion:

The research introduces the innovative Roberta-LightGBM technique framework, a powerful solution addressing tampered fake content in media. Triggered by a notable incident in May 2023, where a fabricated image briefly impacted US markets, the study merges Roberta's NLP model efficiency with LightGBM's binary classification using decision trees. Notably, Roberta's rapid training capabilities surpass traditional techniques like BERT, crucial for real-time analysis of diverse datasets. The framework achieves a remarkable 95.36% accuracy, outperforming alternatives such as XGBoost, emphasizing its ability to discern between real and fake content. With reduced memory usage and a 4.4-second computational time, the framework is efficient and practical. Individual Roberta implementation yields 92.17% efficiency. In conclusion, the Roberta-LightGBM framework emerges as a timely and robust solution against misinformation, offering a valuable tool for enhancing content detection mechanisms in the digital age.

The proposed false news detection approach, combining adjusted Roberta with LightGBM, opens avenues for future research. Potential applications include multilingual adaptability, cross-media analysis, real-time detection, domain-specific customization, user-centric verification, adversarial robustness, explainability, integration with media platforms, and ethical considerations. Future developments could explore these areas, enhancing the system's adaptability, resilience, and ethical implications. Additionally, assessing long-term societal impact and user behavior changes will contribute to a comprehensive understanding of the approach's effectiveness in mitigating the spread of misinformation.

7 Reference

1. A. A. Deshmukh and S. Govilkar, "Fake News Detection on Datasets," 2022 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2022, pp. 274-279, doi: 10.1109/ICAST55766.2022.10039650.

2. A. Qureshi, D. Megías and M. Kuribayashi, "Detecting Deepfake Videos using Digital Watermarking," 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Tokyo, Japan, 2021, pp. 1786-1793. (**??**)

3. A. Yazdinejad, R. M. Parizi, G. Srivastava and A. Dehghantanha, "Making Sense of Blockchain for AI Deepfakes Technology," 2020 IEEE Globecom Workshops (GC Wkshps, Taipei, Taiwan, 2020, pp. 1-6, doi: 10.1109/GCWkshps50303.2020.9367545.

4. C. C. Ki Chan, V. Kumar, S. Delaney, and M. Gochoo, "Combating Deepfakes: Multi-LSTM and Blockchain as Proof of Authenticity for Digital Media," 2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G), Geneva, Switzerland, 2020, pp. 55-62, doi: 10.1109/AI4G50087.2020.9311067. (**??**)

5. E. Z. Mathews and N. Preethi, "Fake News Detection: An Effective Content-Based Approach Using Machine Learning Techniques," 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbat-

ore, India, 2022, pp. 1-7, doi: 10.1109/ICCCI54379.2022.9741049.

6. Essa, E., Omar, K. & Alqahtani, A. Fake news detection based on a hybrid BERT and LightGBM models. *Complex Intell. Syst.* (2023). https://doi.org/10.1007/s40747-023-01098-0.

7. Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: a highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 3149–3157.

8. H. E. Wynne and K. T. Swe, "Fake News Detection in Social Media using Two-Layers Ensemble Model," 2022 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), Phuket, Thailand, 2022, pp. 411-414, doi: 10.1109/ITC-CSCC55581.2022.9894967.

9. Huafeng Zeng, Qiang Yuan, Li Guo, and Shibiao Xu. 2023. Song popularity prediction model based on multi-modal feature fusion and LightGBM. In Proceedings of the 8th International Conference on Communication and Information Processing (ICCIP '22). Association for Computing Machinery, New York, NY, USA, 28–32. https://doi.org/10.1145/3571662.3571667.

10. K. L. Tan, C. P. Lee, K. S. M. Anbananthen and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," in IEEE Access, vol. 10, pp. 21517-21525, 2022, doi: 10.1109/ACCESS.2022.3152828.

11. M. Kandari, V. Tripathi and B. Pant, "A Comprehensive Review of Media Forensics and Deepfake Detection Technique," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 392-395. (**??**)

12. M. Mayank, S. Sharma and R. Sharma, "DEAP-FAKED: Knowledge Graph based Approach for Fake News Detection," 2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Istanbul, Turkey, 2022, pp. 47-51, doi: 10.1109/ASONAM55673.2022.10068653.

13. M. Onoja, A. Jegede, J. Mazadu, G. Aimufua, A. Oyedele and K. Olibodum, "Exploring the Effectiveness and Efficiency of LightGBM Algorithm for Windows Malware Detection," 2022 5th Information Technology for Education and Development (ITED), Abuja, Nigeria, 2022, pp. 1-6, doi: 10.1109/ITED56637.2022.10051488.

14. M. Weerawardana and T. Fernando, "Deepfakes Detection Methods: A Literature Survey," 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), Negambo, Sri Lanka, 2021, pp. 76-81, doi: 10.1109/ICIAfS52090.2021.9606067. 5

15. Orhan, A. Fake news detection on social media: the predictive role of university students' critical thinking dispositions and new media literacy. *Smart Learn. Environ.* **10**, 29 (2023). https://doi.org/10.1186/s40561-023-00248-8.

16. Sai, M.J., Chettri, P., Panigrahi, R. *et al.* An Ensemble of Light Gradient Boosting Machine and Adaptive Boosting for Prediction of Type-2 Diabetes. *Int J Comput Intell Syst* **16**, 14 (2023). https://doi.org/10.1007/s44196-023-00184-y.

17. Sangjun Lee, Donggeun Ko, Jinyong Park, Saebyeol Shin, Donghee Hong, and Simon S. Woo. 2022. Deepfake Detection for Fake Images with Facemasks. In Proceedings of the 1st Workshop on Security Implications of Deepfakes and Cheapfakes (WDC '22). Association for Computing Machinery, New York, NY, USA, 27–30. https://doi.org/10.1145/3494109.3527189.

18. Songbai Zhu and Guolai Yang. 2023. Research on LightGBM-based fault prediction for electrical equipment in artillery fire control system. In Proceedings of the 4th International Conference on Advanced Information Science and System (AISS '22). Association for Computing Machinery, New York, NY, USA, Article 30, 1–6. https://doi.org/10.1145/3573834.3574499.

19. V. Gupta, R. S. Mathur, T. Bansal and A. Goyal, "Fake News Detection using Machine Learning," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 84-89, doi: 10.1109/COM-IT-CON54601.2022.9850560.

20. Yunxin Liang, Jiyu Wu, Wei Wang, Yujun Cao, Biliang Zhong, Zhenkun Chen, and Zhenzhang

Li. 2019. Product marketing prediction based on XGboost and LightGBM algorithm. In Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition (AIPR '19). Association for Computing Machinery, New York, NY, USA, 150–153. https://doi.org/10.1145/3357254.3357290.

21. Yuxiang Zhang, Jingze Lu, Xingming Wang, Zhuo Li, Runqiu Xiao, Wenchao Wang, Ming Li, and Pengyuan Zhang. 2022. Deepfake Detection System for the ADD Challenge Track 3.2 Based on Score Fusion. In Proceedings of the 1st International Workshop on Deepfake Detection for Audio Multimedia (DDAM '22). Association for Computing Machinery, New York, NY, USA, 43–52. https://doi.org/10.1145/3552466.3556528.

22. Z. Zhang, "Microsoft Malware Prediction Using LightGBM Model," 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Xi'an, China, 2022, pp. 41-44, doi: 10.1109/ICBAIE56435.2022.9985850. 19

23. Zhanbo Li and Xiaoyang Li. 2022. Intrusion Detection Method Based on Genetic Algorithm of Optimizing LightGBM. In Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering (EITCE '21). Association for Computing Machinery, New York, NY, USA, 1366–1371. https://doi.org/10.1145/3501409.3501651.