International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 220-229 www.ijaem.net ISSN: 2395-5252

Propaganda News Classification Using Deep Learning Techniques

Ojasvat Bharadwaj
Lecturer, MVJ College of Engineering, Bangalore, India.

Date of Submission: 05-07-2024 Date of Acceptance: 15-07-2024

ABSTRACT: Whether the information being spread is true or misleading, it can still be done through propaganda. For the purpose of obscuring the understanding of the propaganda targets, bias is frequently used in propaganda. One sort of media that is commonly utilised to spread propaganda is news reports. Propaganda must be identified in news items by text classification, which is essential for stopping its dissemination.
The text classification algorithms Long Short-Term Memory (LSTM), GRU, CONV-LSTM, CONV-GRU, and BERT have all been used extensively. The weakness of LSTM is that it has a large bias when attempting to infer context from the text by word order. In text analysis, Convolutional Neural Networks (CNN) may extract important information using convolutional layers, however, it is insufficient for context extraction. In order to detect propaganda in news articles, this work aims to merge the LSTM, CNN, GRU, and CNN with LSTM and GRU algorithms in text categorization. It has been demonstrated that combining each strategy improves classification performance while requiring less running time. The CONV-LSTM method is the best one out of the five, according to the comparison analysis.

KEYWORDS: News classification, Deep Learning, Filtering.

I. INTRODUCTION

Recent years have seen a surge in research on propaganda news detection, and numerous models have been put forth to detect various forms of deception. There are a few recognised categories of algorithms, common approaches they use, and data types they can analyse because this is a nascent study topic. Even the term "propaganda news" is in dispute because it is open to several interpretations.

There is a variety of mapping studies, surveys, literature reviews, and alternative analysis projects, however, none have offered a comprehensive summary of the classification algorithms that have been studied, and that have been studied the most, exploitation deep learning algorithms and a range of datasets. One of the difficult tasks is to classify and detect news using various deep learning models and algorithms.
(1) Because it is written for a particular purpose and particular group of people and
(2) Because of similar use of context and linguistic features written for normal news as well.

The objective of this work are as follows
1. To provide an overview of the datasets
2. To perform the analysis on the different deep learning techniques
3. Perform the comparison analysis for the present work among the different deep learning techniques
4. Find the algorithm that must detect propaganda news in the fastest feasible way & maximum possible accuracy using neural networks

As said earlier, there aren't enough overviews of the algorithms that work best for the application and those that have already been put to the test. This report's goal is to provide a summary of the various deep-learning models for detecting misinformation. This should make it easier for future researchers to uncover detection models that are comparable to the detection models they are currently creating or to identify the classification algorithms that still need more investigation.

The following research questions are as follows:
1. Why it is required to detect propaganda-based news?
2. Which deep learning model can effectively classify news that is based on propaganda?
3. Do some categorization algorithms perform more accurately and with higher f-scores than others?
4. Does contextual and linguistic features matter for such news classification?

Stroke IC engines are working on camshaft Mechanism. Although the conventional system has proven to be convenient and safe. Its fixed valve timing is necessarily a compromise of combustion stability, fuel economy and maximum torque objectives. Cam is an integral part of an engine as it controls valve actuation which in turn is responsible for supply of air-fuel mixture into the combustion chamber and for the removal of exhaust gases from the combustion chambers.

Even inside the topic articles for this literature study, the term "propaganda news" is employed to denote a good variety of problems and isn't systematically defined. The terms "propaganda news," "false news," "disinformation," "misinformation," "false information," hoaxes, "dishonourable news," "rumour," "clickbait," "conspiracies," "parody," "caustic remark news," and "propaganda" refer to a wide variety of destructive or factual misinformation types. After learning that the phrase was used in academic literature, [2] developed a taxonomy of six essential categories, including satire, parody, fabrication, manipulation, advertisement, and propaganda. The six types can even be further broken down into individuals with a high or low degree of factual accuracy and those who have a high or low level of authoritarian purpose to mislead. [3] Deceptive news, fake news, satirical news, disinformation, misinformation, cherry-picking, clickbait, and rumour are the eight ideas associated with Propaganda news. Each thought is connected with authenticity or the amount of factual truth, the author's goal, and a binary value indicating whether it is news or of an undetermined category. The phrase "propaganda news" is rejected, and incorrect information is instead classified as "misinformation" (false information transmitted without malice) "disinformation" (false information spread with malice), and "malformation" (truths spread with ill intent). While the aforementioned examples all take into account intent, the majority of the material consulted for this study simply addresses the degree of factual accuracy, and the term "propaganda news" is used to refer to both factually wrong news media and social media postings and comments. Regardless of the underlying goal, "propaganda news" is going to be used as a general descriptor of factual incorrectness during this analysis [4].

II. MODEL

At this time, many classification model scenarios will be established. It will be developed a classification model with LSTM, a combination of LSTM and CNN, GRU, a combination of CNN and GRU, and BERT, and the performance of the model will then be compared. To make the final performance comparable, each model will be built easily and with similar hyperparameter settings.

2.1 LSTM modelling

Bidirectional LSTM layers are the type of LSTM employed in the LSTM model. In order to make it simpler for the model to understand the context present in the data, bidirectional LSTM offers the advantage of preserving context in both past and future directions [18]. The figure illustrates the summary of this model.

2.2 GRU Modelling

GRU is the variant of the RNN and is the better than the LSTM. In LSTM it tries to solve the vanishing gradient problem but not as good as the GRU. And GRU is not preferred when there is longer sequences. And GRU uses lesser memory as compared to the LSTM and it is also faster than the LSTM.

2.3 CONV-LSTM Modelling

ConvLSTM utilizes the ability of convolutional layers to understand the complex relationships, so integrating that layer in LSTM gives the ability to represent complex information easily.

Similar to the ConvLSTM here we can also use it along with the GRU to improve the learning of the complex patterns.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 220-229 www.ijaem.net ISSN: 2395-5252

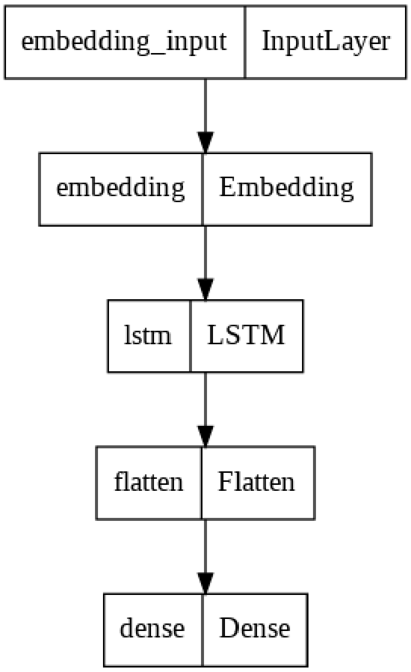

Figure 1 LSTM Modelling

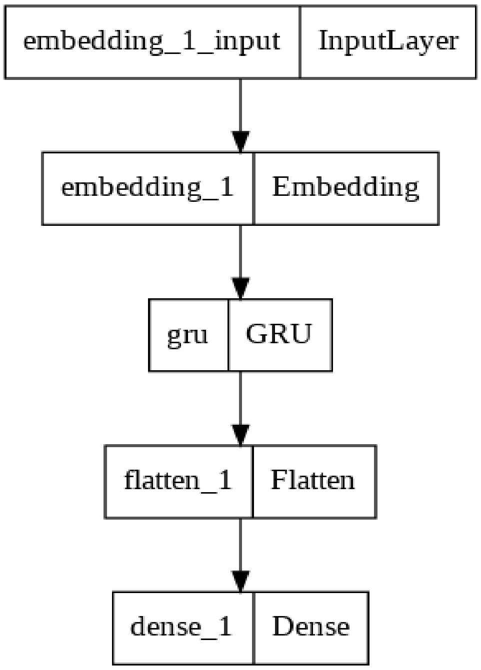

Figure 3 GRU modelling

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 220-229 www.ijaem.net ISSN: 2395-5252

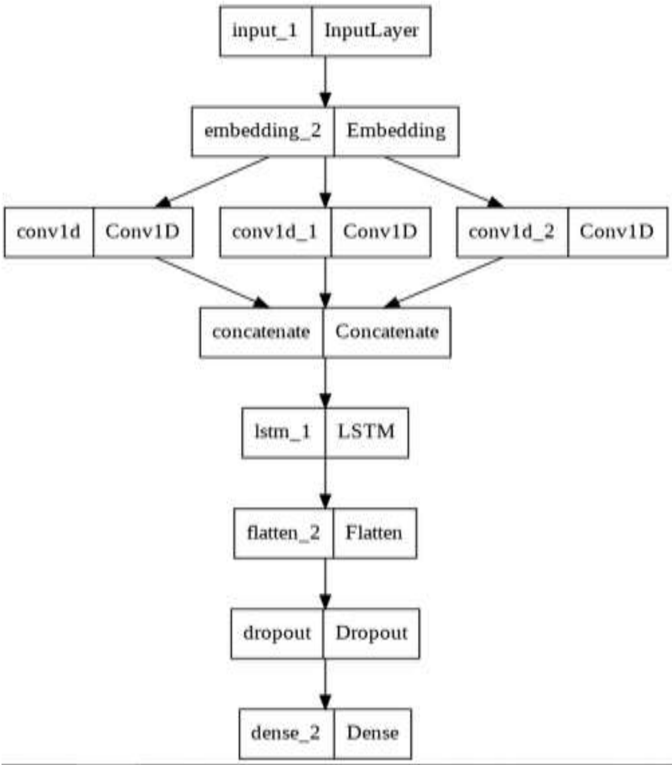

Figure 2 CONV-LSTM Modelling

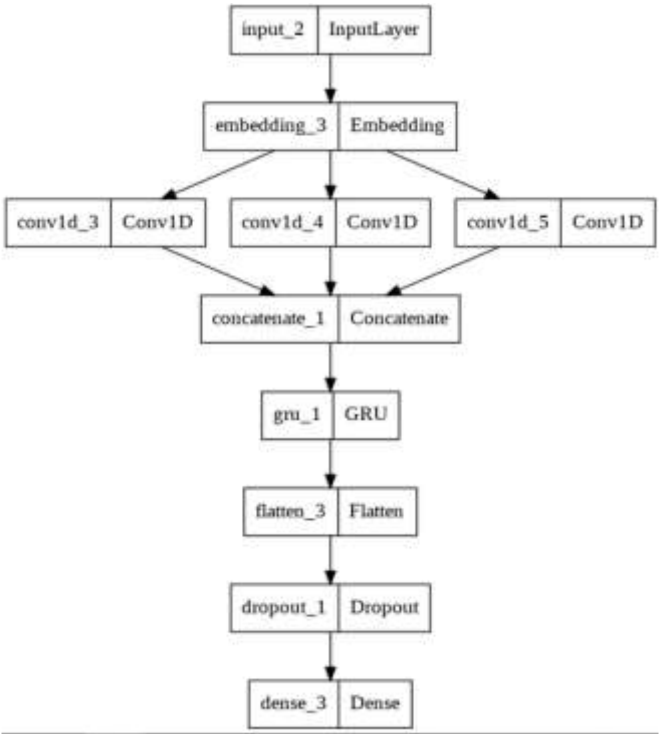

Figure 4 CONV-GRU Modelling

III. LOSS FUNCTIONS

3.1 Mean Squared Error Loss

The loss function’s Mean Squared Error (MSE) is the workspace of fundamental loss functions as it is straightforward to comprehend, simple to use, and generally effective. Just square the difference between the model's predictions and actual data, average it across the entire dataset, and then use that result to determine MSE. No matter which way both the anticipated and actual figures are signed, the result is always positive, and 0.0 is the ideal value.

By dividing the difference between "y" and our prediction by two, we can determine the mean squared error (MSE) loss. We take these new numbers, square them, add them all up to create a total, and then divide this total by y once more. Our final outcome will be this. The formula for calculating mean squared error loss is as follows:

$$MSE = \frac{1}{n}\Sigma\left(y - \hat{y}\right)^2$$

3.2 Squared Error Loss

Each training example's squared error loss is the discrepancy between actual and anticipated values is sometimes referred to as L2 Loss:

$$L = \left(y - f(x)\right)^2$$

The Mean of these Squared Errors, which is the associated cost function (MSE).

3.3 Root Mean Square Propagation

The learning rate in Root Mean Square Propagation is different from the learning rate in Adagrad because it is not the sum of the squared gradients, but an exponential average of the gradients. Root Mean Square Propagation essentially mixes AdaGrad and momentum.

$$\text{cache}_{\text{new}} = \gamma * \text{cache}_{\text{old}} + (1-\gamma) * \left(\frac{\partial(\text{Loss})}{\partial(W_{\text{old}})}\right)^2$$

IV. TRAINING

During the training process, it can be done in 2 ways. One where for each training sample the loss can be calculated and the weights are updated. And in second way after every couple of training samples it is updated instead of updating for every sample. This will reduce the training process as well as also reduces the chance of overfitting. In training process samples are grouped into the group size with Batch. In batch learning weights are only updated when loss for all the samples in the batch is calculated and at end the wight are updated.

This size of the batch can impact the performance of the model. So this parameters is considered as the hyper parameter. If the Batch size is considered very low then model could be over fitting and can not generalize properly. And also results in the slower process of training. And when the batch size is kept too large, it will take very short time but does not fit to data properly and may have poor performance.

- Stochastic Gradient Descent is that Batch Size = 1
- Mini-Batch Gradient Descent is that 1 < Batch Size < Size of Training Set

Epoch

"Epochs" are determined by how often the learning algorithm iterates through the entire set of hyperparameters that makes up the training dataset. Every sample of the training dataset has had its internal model parameters updated once during an epoch. An epoch consists of one or more batches. An example would be the batch gradient descent learning algorithm, which describes a single-batch epoch. One may envision a for-loop that iterates through the training dataset one epoch at a time. A for-loop that is nested inside of the current one iterates through each batch of samples, iterating through the "batch size" number of samples in each batch. Until the model's error is sufficiently decreased, the learning process can go on for hundreds or thousands of epochs. In the literature and tutorials, it is possible to find examples of the number of epochs set to 10, 100, 500, 1000, and higher. A y-axis that shows the model's accuracy or skill is typically present in graphs that show epochs along a line x-axis as time. The term "learning curve" also applies to these graphics. These graphs can be used to evaluate how well the model fits the training dataset and if it was over- or under-trained.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 220-229 www.ijaem.net ISSN: 2395-5252

V. RESULTS

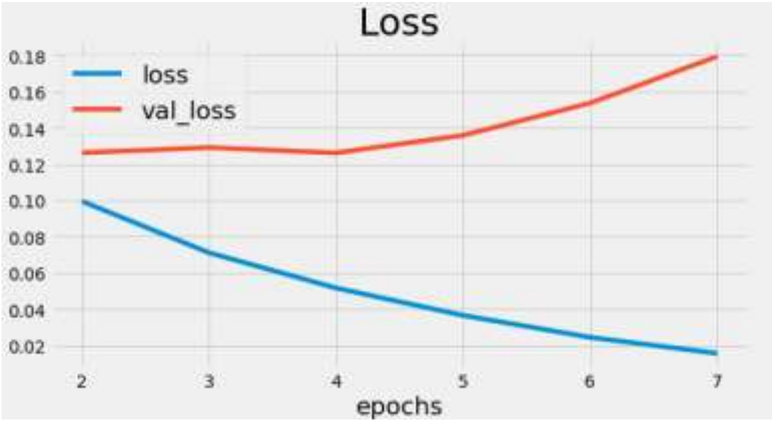

Figure 5 Loss - LSTM Model

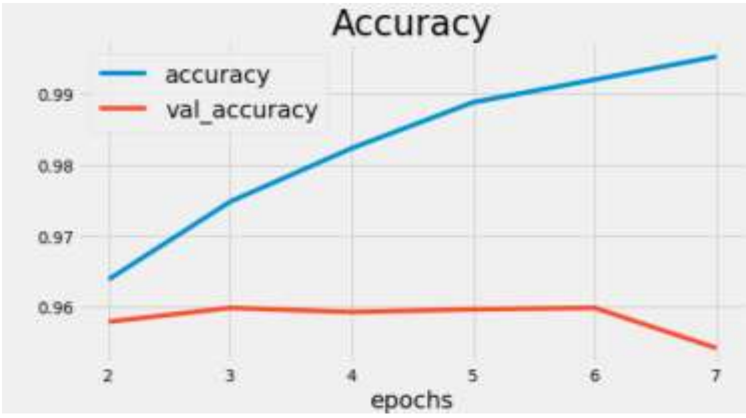

Figure 7 Accuracy - LSTM model

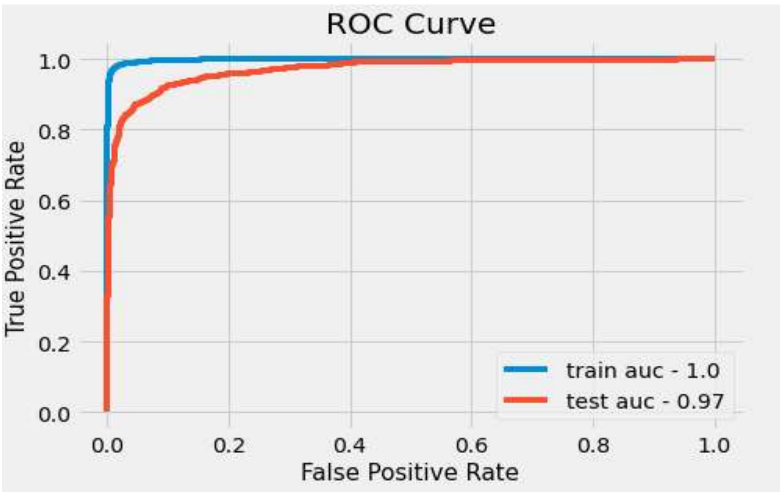

Figure 9 ROC Curve- LSTM Model

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 220-229 www.ijaem.net ISSN: 2395-5252

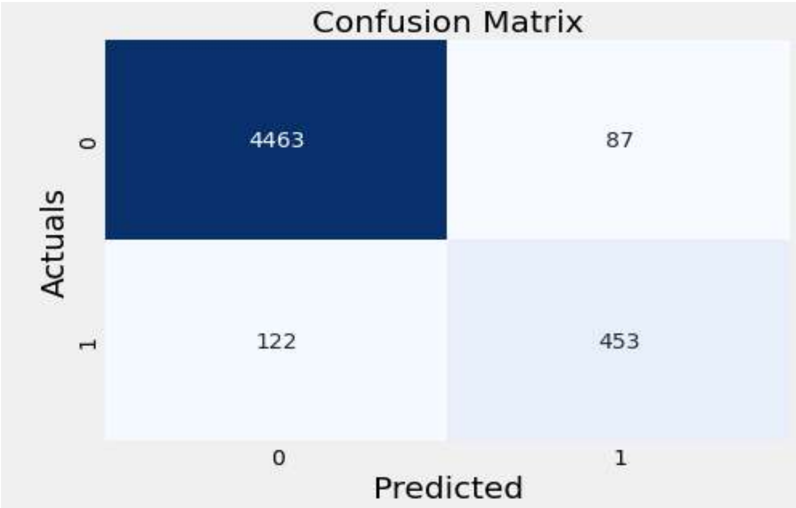

Figure 6 COnfusion Matrix - LSTM Model

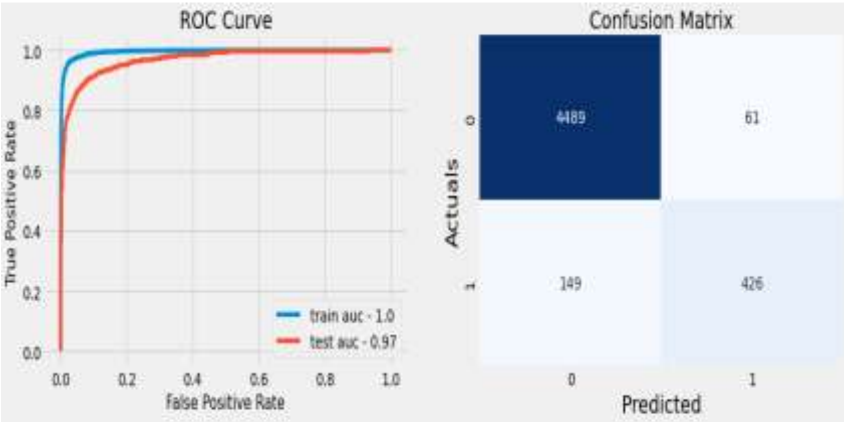

Figure 8 ROC-Confusion- GRU Model

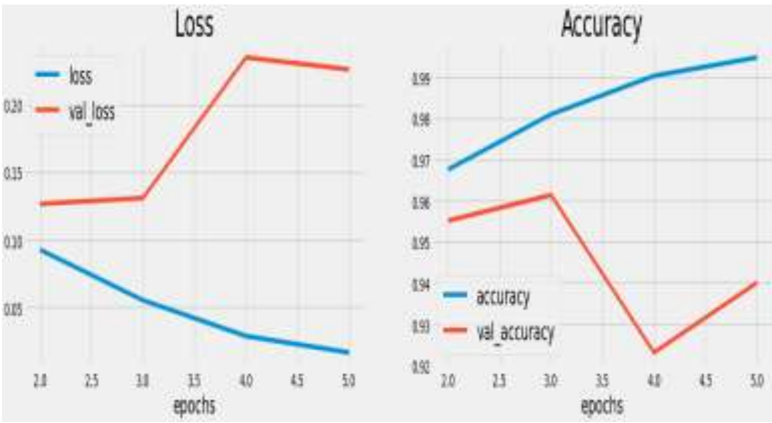

Figure 24 Loss and Accuracy CONV-LSTM Model

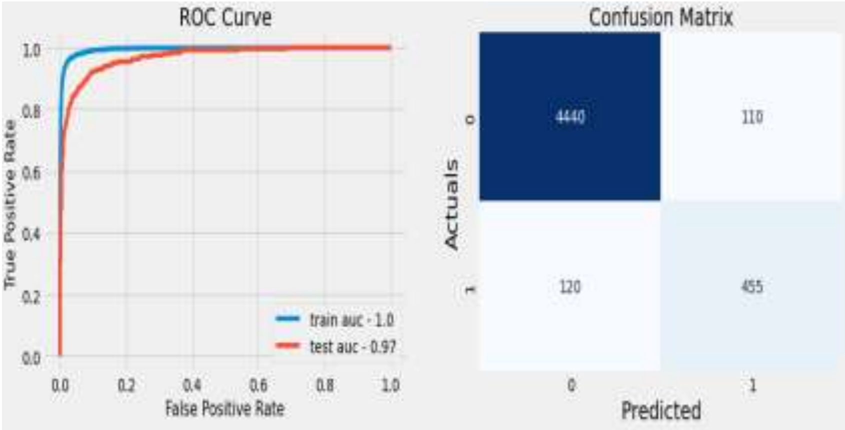

Figure 10 ROC and Confusion - CONV-LSTM Model

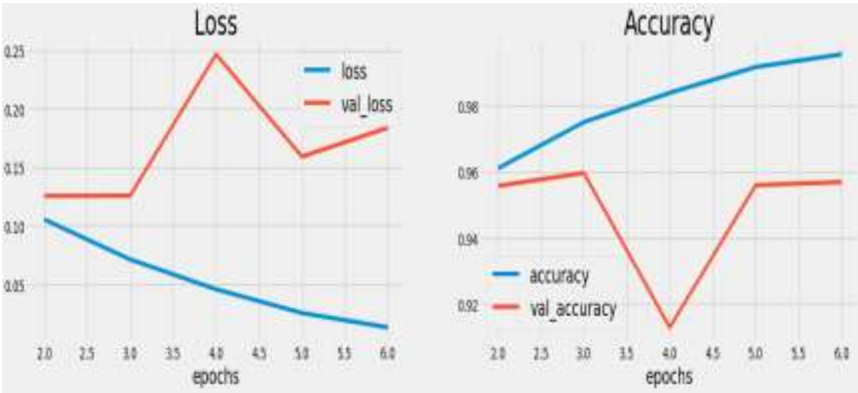

Figure 11 Loss and Accuracy CONV-GRU Model

Comparison Analysis

	model_name	accuracy	auc	not_propaganda_accuracy	propaganda_accuracy	f1	precision	recall	roc
2	CONV LSTM Model	0.9551	0.9701	0.9758	0.7913	0.7982	0.8053	0.7913	0.8836
0	LSTM Model	0.9592	0.9704	0.9809	0.7878	0.8126	0.8389	0.7878	0.8844
3	CONV GRU Model	0.9557	0.9671	0.9822	0.7461	0.7908	0.8412	0.7461	0.8641
1	GRU Model	0.9590	0.9687	0.9866	0.7409	0.8023	0.8747	0.7409	0.8637
4	BERT Model	0.8929	0.8467	0.9868	0.1496	0.2386	0.5890	0.1496	0.5682

Best Model Selection

1. The best model is - **Convolution LSTM Model**
2. The convolution LSTM-based model has beaten all other deep learning models including bert and also machine learning models.
3. The minority class accuracy is 79.82% which is 7% more than the best machine learning model and a little bit more than LSTM model.

VI. CONCLUSIONS

This paper uses a variety of deep learning algorithms to investigate the classification of

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 220-229 www.ijaem.net ISSN: 2395-5252

propaganda news. There aren't many recognised categories of the algorithms, the fundamental tactics they employ, or the kinds of data they can handle because this is a new field of study. The current research demonstrates that the Convolution LSTM model is the most suitable for this kind of application. With this approach, the maximum accuracy possible is 79.13%

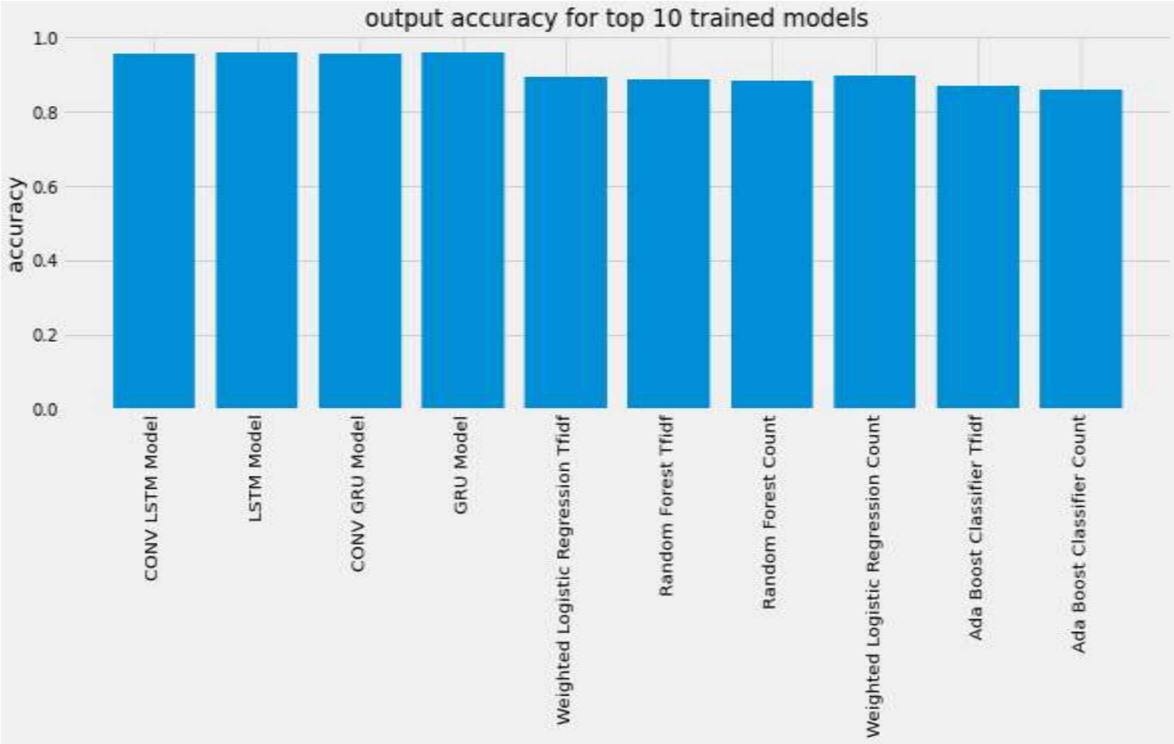

Figure 12 output accuracy for top 10 trained models.

The deep learning algorithms, with a wide range of datasets. Various deep learning models and algorithms with the use of contextual embeddings are used to classify propaganda news. First, the overview of the datasets was provided and then the analysis was performed on the different deep learning techniques.

REFERENCES

[1]. Barata, G., Gama, S., Jorge, J. and Gonçalves, D., 2013, September. Engaging engineering students with gamification. In 2013 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES) (pp. 1-8). IEEE.
[2]. Barata, G., Gama, S., Jorge, J. and Gonçalves, D., 2013, October. Improving participation and learning with gamification. In Proceedings of the First International Conference on gameful design, research, and applications (pp. 10-17).
[3]. Manakkadu, S., Joshi, S.P., Halverson, T. and Dutta, S., 2021, December. Top-k User-Based Collaborative Recommendation System Using MapReduce. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 4021-4025). IEEE.
[4]. Caraban, A., Karapanos, E., Gonçalves, D. and Campos, P., 2019, May. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-15).
[5]. Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J. and Gonçalves, D., 2011, October. Blind people and mobile touch-based text-entry: acknowledging the need for different flavors. In The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility (pp. 179-186).
[6]. Manakkadu, S. and Dutta, S., 2024. Ant Colony Optimization based Support Vector Machine for Improved Classification of Unbalanced Datasets. Procedia Computer Science, 237, pp.586-593.
[7]. Mohammed, T. and Mohammed, A., 2017. Real Time Video Surveillance System for Fire and Smoke Detection Based on Wavelet

Transform. Journal of Zankoy Sulaimani-Part A, 19(2), pp.229-238.
[8]. Manakkadu, S. and Dutta, S., 2024. Efficient Feature Clustering for High-Dimensional Datasets: A Non-Parametric Approach. Procedia Computer Science, 237, pp.576-585.
[9]. Mahmood, S.R., Amin, T.A.M., Ahmed, K.H., Mohammed, R.D. and Karim, P.J., 2024. Multi-Label Feature Selection with Graph-based Ant Colony Optimization and Generalized Jaccard Similarity. Kurdistan Journal of Applied Research, 9(1), pp.38-51.
[10]. Manakkadu, S. and Dutta, S., 2022. ACO based Adaptive RBFN Control for Robot Manipulators. arXiv preprint arXiv:2208.09165.
[11]. Amin, T.A.M., Mahmood, S.R., Mohammed, R.D. and Karim, P.J., 2022. A Novel Classification of Uncertain Stream Data using Ant Colony Optimization Based on Radial Basis Function. Kurdistan Journal of Applied Research, pp.57-70.
[12]. De Zoysa, K., Keppitiyagama, C., Seneviratne, G.P. and Shihan, W.W.A.T., 2007, August. A public transport system based sensor network for road surface condition monitoring. In Proceedings of the 2007 workshop on Networked systems for developing regions (pp. 1-6).