

Enhancing Machine Learning Models in Identifying False Information

Pradeep Kumar, Ravi Joshi, Himanshu Arora, and Gaurav Kumar Soni

Abstract—In today's age, the dissemination of inaccurate and misleading facts on digital systems has ended up a serious problem that poses risks to people, organizations and society as a whole. Fake news spreads quicker than true information due to its bizarre and compelling content material. Fake information may have critical results, undermining the democratic system, undermining belief in the media and growing further divisions. Because of its fast spread and impact on the general public, it's far crucial to have a powerful manner to detect false information. In this paper, we pay attention to the usage of machine learning algorithms which includes SVM, Naive Bayes, and logistic regression, in addition to acquiring knowledge through deep learning models, which includes LSTM & Bi-LSTM, as well as utilising transfer learning with BERT-based approaches. Furthermore, the performance of these models will be evaluated on datasets comprising both Hindi and English languages.

Keywords—False Information; Machine Learning; Deep Learning; Transfer learning

I. INTRODUCTION

WITH the increased utilization of social media and online digital platforms such as Facebook, WhatsApp, Twitter, etc which boats over 5.3 billion active internet users worldwide, the way people consume news has significantly shifted. Surprisingly, according to Forbes nearly 64.5 % of users now receive information from social media rather than traditional media sources such as newsletters, journals, and news wire services [1]. Today, digital news platforms such as blogs, news feeds on social media, and numerous websites have become prominent sources of information offering instant access to news anytime and from anywhere with just one click [2], [3]. However, this paradigm shift in news consumption has unfortunately led to a significant surge in the propagation of fake news.

False or misleading information presented as news, is often created with the intention to deceive or manipulate readers. Fake news can sometimes be portrayed in such a way that it gives an impression of real news, leading others to trust it. It can take many forms, including hoaxes, propaganda, misinformation, disinformation, fabricated stories, distorted facts, and so on. The viral nature of social media amplifies the

reach of fake news, allowing it to propagate rapidly and extensively, potentially influencing public opinion. Fake news thrives in the digital age for a number of reasons, including the decentralized nature of online platforms, a lack of editorial oversight, social media algorithms that frequently favour sensational content over accurate reporting, which amplifies false or misleading information because it tends to get more clicks, likes, and shares for advertising purposes. The impact of fake news is significant and far-reaching. It not only undermines public trust in credible news sources, but also poses a grave threat to global commerce, journalism, and democracy, inflicting substantial collateral damage.

In mid-2016, Craig Silverman, BuzzFeed's media editor, came across a peculiar trend of fabricated stories that appeared to originate from a small town in Eastern Europe. It didn't take long for Silverman and his colleague to delve into the matter, and their investigation revealed the existence of approximately 140 deceptive news websites that were gaining significant traction on Facebook. The young individual involved may or may not have possessed a significant interest in American politics. However, driven by the potential financial gains from Facebook advertising, they sought to spread their fictional content extensively on social media platforms. Since then, the issue of fake news has gained more attention from researchers [4], [5]. There are certain patterns that can be observed in fake news. Such as Sensational language (exaggerated or sensational language), Hyperbolic statements (exaggerated claims or statements that are not supported by evidence), Conspiracy theories (Fake news often promotes conspiracy theories or unfounded claims), Lack of attribution or anonymous sources (Fake news often lacks proper attribution or relies on anonymous sources).

With advancing technology and digital manipulation tools, it has become increasingly easier to create and disseminate deceptive content across different mediums such as tampered images, manipulated audio, edited videos, Clickbait links, etc. This multi-modal nature of false information necessitates the development of comprehensive detection and verification systems that can analyse and authenticate content across different media types.

In this research aims to contribute to the field of identification of fake news by leveraging top features that exhibit a strong association with the target variable. By utilizing these features, we intend to enhance and improve the accuracy and performance of machine learning classifiers such as SVM, NB, and LR. In our research we also proposed dual-layer LSTM architectures in deep learning models, leveraging its ability to capture long-term dependencies, contextual

Pradeep Kumar is with Modern Institute of Technology and Research Centre, Alwar, Rajasthan, India (e-mail: pradeepk0411@gmail.com).

Ravi Joshi is with Apex University, Jaipur, Rajasthan, India (e-mail: raviyoshi93@gmail.com).

Himanshu Arora is with Arya College of Engineering, Jaipur, Rajasthan, India (e-mail: dr.himanshuarora17@gmail.com).

Gaurav Kumar Soni is with Global Institute of Technology, Jaipur, Rajasthan, India (e-mail: gksoni2709@gmail.com).



information and patterns in textual data. By incorporating this architecture, our aim is to enhance the model's capability to understand and interpret sequential information in text, which is crucial for fake news identification. Further, in our study, we conducted a comparative analysis of various BERT based-models such as BERT, ALBERT, DistilBERT, mBERT, and DistilBERT to determine which performs better in specific tasks. Each model comes with its unique architecture and pre-training methodology, allowing us to assess their effectiveness in classifying the given data accurately.

The Hindi and English datasets used in our study were developed by merging several datasets that assist in acquiring news from India and worldwide.

The rest of the paper is organized as follows. In Section 2, a thorough review of previous works in the field is presented, offering a comprehensive understanding of the existing literature. Section 3, the proposed work is introduced, providing an overview of the methodology and outlining the steps involved in the study. Section 4 serves as a summary of the obtained results, highlighting the performance of the proposed approach. Additionally, a comparison is made between the proposed work and other state-of-the-art techniques to provide insights into its efficacy. Finally, in Section 5, the paper is concluded, summarizing the key findings and potentially suggesting avenues for future research.

II. RELATED WORK

This section provides a comprehensive overview of the relevant studies and publications on identifying fake news using various models and algorithms. By exploring the existing literature, this section aims to assess the methodologies employed, evaluate the effectiveness of different approaches, and identify potential gaps or areas for further investigation.

A. Machine Learning Approaches

Some of the conventional machine learning techniques that have undergone thorough research in the identification of fake news include Support Vector Machines (SVM), Naive Bayes, Logistic Regression, Passive Aggressive classifiers, and more. Additionally, the advancement of Natural Language Processing (NLP) techniques has made it easier to detect and stop the spread of misleading information effectively [6].

Machine learning algorithms can be trained in supervised, unsupervised, and Reinforcement manner. Unsupervised learning, however, has only been the subject of a very small number of suggested works. One is suggested by Ying et al. (2019) a reliable collapsed Gibbs sampling technique, a Bayesian network model, and an unsupervised strategy for detecting false news that also evaluates user trustworthiness[7]. Several word-embedding approaches are being combined with ML algorithms to improve the performance of ML models by extracting important features, such techniques include TF-IDF, Word2Vec, FastText, and more.

[8] have employed TF-IDF to transform textual data into numerical representation and four distinct machine learning classifiers (NB, RF, LR, and PA Classifier) to classify text as TRUE or FAKE. NB, RF, and LR were trained on LIAR, and

REAL_OR_FAKE.csv, specifically used for the Passive Aggressive classifier. Furthermore, to split the dataset, 67% for Training and 33% for Testing 3-fold CV is used. while, [9] have utilised three different word embedding techniques namely Doc2Vec, Word2Vec, and FastText along with Linear SVM, SVM-RBF (Radial Bias Function), and Random Forest classifiers for hate speech detection in Hindi-English mixed social media text. Their findings reveal that the fastText features yield the highest accuracy of 85.81% when combined with the SVM-RBF classifier.

B. Deep Learning Approaches

Nowadays, deep learning is preferred over traditional machine learning models because DL models such as RNNs, and CNNs, can effectively capture the sequential and contextual information present in text. Furthermore, pre-trained word embeddings, such as FastText, GloVe, and so on enhances the performance of deep learning models for text classification. In [10] suggested a novel approach for rumour identification in microblogs in a study. The researchers leverage Recurrent Neural Networks (RNNs) to train hidden representations that capture the fluctuation of contextual information of relevant postings over time. The study makes use of two microblog datasets derived from Twitter and Sina Weibo. [11] and [12] have presented models for false news detection that combine CNN and LSTM. Combining CNN with LSTM is an intriguing approach since the model can learn fine-grained features as well as general context, resulting in improved overall performance. In [11] employed FastText for word embedding and presented BCNN (Bi-CNN), LSTM-CNN, Bi-LSTM-BCNN, and APS-CNN (Alternative Pooling Symmetry) to detect clickbait articles among the different categories of fake news in Korean language. [12] proposed seven deep-learning models using CNN and LSTM modifications. On the PolitiFact dataset, CNN+Bi-LSTM ensemble with an attention mechanism achieved the maximum accuracy of 88.78%. [13] have introduced an automatic detection of fake news using Bi-LSTM concatenated and Multi-head LSTM models applied to the FNC-1 dataset. They have utilised Google's pre-trained word embedding model to transform input into vector space.

C. Transfer Learning Approaches

Although Deep learning models are capable of learning complex patterns and representations from raw data, training these models can be computationally intensive and time-consuming, particularly when dealing with complex architectures and a large number of parameters. Achieving good performance often necessitates a substantial amount of labelled data. In contrast, transfer learning offers pre-trained models that have been trained on large corpora in multiple languages. It enhances generalisation, minimising training time and computational resources, and allows fine-tuning on specific tasks using limited datasets. Below, we will discuss some works that have proposed approaches in the field of transfer learning. [14] have proposed BDANN (BERT-Based Domain Adaption Neural Network for multi-modal fake news detection) that focuses on both textual and visual features of news. For extraction of textual features, BERT is used and for Visual features, VGG-19 is used. BDANN is applied to two multimedia datasets Twitter and Weibo.

Using transformer-based approaches, [15] have concentrated on categorising bogus news about COVID-19. The Roberta model performs better with an F1-Score of 0.90 when compared to BERT (0.83) and XLNet (0.86), according to the researchers' examination of models like BERT, Roberta, and XLNet. For identifying fake in the Arabic language,[16] have presented a model consisting of three primary components.

These components include the BERT model for sequence labelling and encodings, an attention mechanism, and JointBERT, which utilises RFC and NER as shared parameters.

Table 1 gives a comprehensive analysis of the related work discussed above.

TABLE I
ANALYSIS OF STATE-OF-THE-ART WORKS PROPOSED IN THE FIELD OF FAKE NEWS DETECTION.

Machine Learning Approaches			
Ref (Year)	Description	Performance	Summary
[7] (2019)	An unsupervised approach with Gibbs sampling for identifying fake news and evaluating user trustworthiness.	Accuracy LIAR dataset: 75.9 % BuzzFeed dataset: 67.9%	Employing an unsupervised approach for fake news detection may yield predictions that don't align with the actual truth.
[8] (2020)	Four distinct ML classifiers incorporate TF-IDF to distinguish between fake and true news. Additionally, 3-fold Cross-validation is used to split data for training and testing.	Accuracy Passive Agg.: 92.7% Naive Bayes: 60% Logistic Reg.: 65% Random Forest: 59%	To deploy the model in a real-world setting, addressing the likelihood of data drift occurring when input data distribution changes over time is necessary.
[9] (2020)	For the identification of hate speech in Hindi-English mixed data. The author utilised FastText, word2Vec, doc2vec with SVM, SVM-RBF, and Random Forest (RF).	Accuracy FastText+SVM-RBF: 85.8% Word2Vec+SVM-RBF: 75.1% Doc2Vec+RF: 64.15%	Instead of only hate and no hate, hate speech can have varying degrees of intensity and should be applied to new trends and words that emerge throughout time.
Deep Learning Approaches			
[10] (2016)	RNN-based model for detecting rumour and learning the hidden representation that captures the variation of contextual information of relevant posts over time.	Accuracy Twitter: 82.7% Weibo: 87.3%	Adding multiple hidden and embedding layers can increase computational complexity and affect the model's scalability when dealing with large amounts of data.
[11] (2019)	BCNN, LSTM-BCNN, BiLSTM-BCNN, and APS-BCNN are presented for clickbait detection in the Korean language.	Highest AUC-Score APS-BCNN: 0.726 BCNN: 0.52	The models achieve meaningful accuracy when the content of the body is irrelevant to the context and accuracy is low when the headline and content are inconsistent.
[12] (2020)	Seven deep learning models which incorporate CNN and LSTM are introduced for fake news detection.	Highest Accuracy CNN+Bi-LSTM with attention mechanism: 88.78%	Due to the comparative character of the investigation, the classification models were unable to discern the semantic change of true news to fake news during its transmission.
[13] (2019)	For the automated identification of fake news, Bi-LSTM concatenated and multi-head LSTM models with Google's pre-trained word embedding model are used.	Accuracy BiLSTM concatenated: 85.3% multi-head LSTM: 82.94%	Multihead LSTM models may struggle to recognise global context, particularly for tasks requiring long-term interdependence.
Transfer Learning Approaches			
[14] (2020)	BDANN for the textual and visual features of news. For textual features, BERT is used and for visual features, VGG-19 is used.	Accuracy Twitter: 83% Weibo: 85.1%	Image noise may have an influence on BDANN's performance. Future studies might focus on determining the relevance of visuals to the accompanying text.
[15] (2021)	Detection of fake news about COVID-19 by employing BERT, ROBERTa, and XLNET.	F1-Score BERT: 0.83 ROBERTa: 0.90 XLNET: 0.86	Training the models using the COVID-19 dataset may bias the models towards one type of news.
[16] (2022)	Author presented Joint BERT with RFC and NER as shared parameters for detecting fake news in the Arabic language.	Accuracy ANS dataset: 66% COVID-19: 85% AraNews: 80%	Since the presented model was trained using Arabic news, it is less generalizable to other languages.

III. PROPOSED WORK

This section outlines the steps and methods for classifying false and legitimate news using machine learning and deep learning models. Figure 1 illustrates the sequential process involved in classifying news as Fake or Real utilizing Machine learning classifiers namely SVM, Naive Bayes, and Logistic Regression.

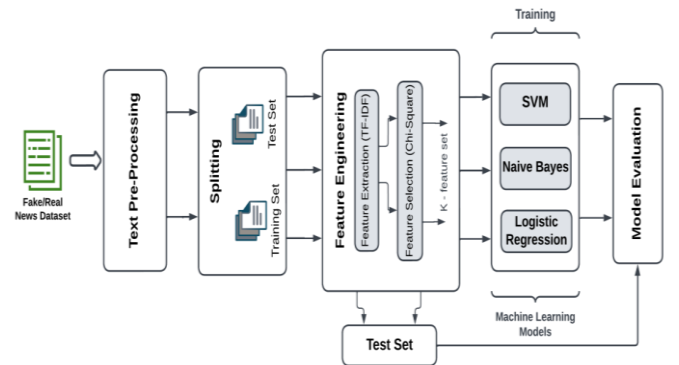


Fig. 1. Machine Learning-based Fake/Real News Classification

On the other hand, Figure 2 provides an overview of the steps employed in the classification process utilising deep learning and Transfer learning models.

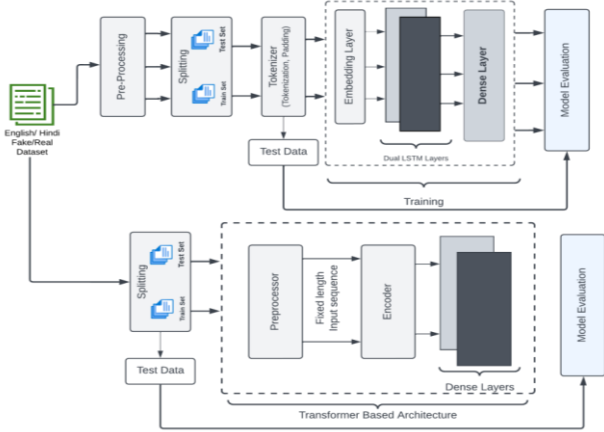


Fig. 2. Deep Learning and Transfer Learning based classification

A. Dataset Collection

The aforementioned models are trained on Dataset I (English) and Dataset II (Hindi) which contain fake and real News. To train the models, we split each dataset into 80% for training and 20% for testing purposes.

The dataset-I was created by combining three different datasets sourced from Kaggle. By combining these datasets, we were able to create a more comprehensive dataset with 50,000 news samples from different categories such as Politics, Medical, international affairs, Sports, etc. Our text classification relies on micro-blogs (headlines) as the primary data source. In addition to Dataset I, which contains text in the English language, we have also created a custom dataset for Hindi Fake/Real news, referred to as Dataset II. Dataset II, The Hindi dataset used in our study was created by merging two different datasets, obtained from Kaggle. By combining these datasets, we created a diverse Hindi dataset with 46,000 news samples. This approach allows us to include a vast variety of news articles from various categories in our analysis. The distribution of Fake (0) and Real (1) news samples is uniform in both the Hindi and English datasets.

B. Data Preprocessing

It is an essential step in text classification problems. In English text, the steps involved in preprocessing are shown in fig 3.

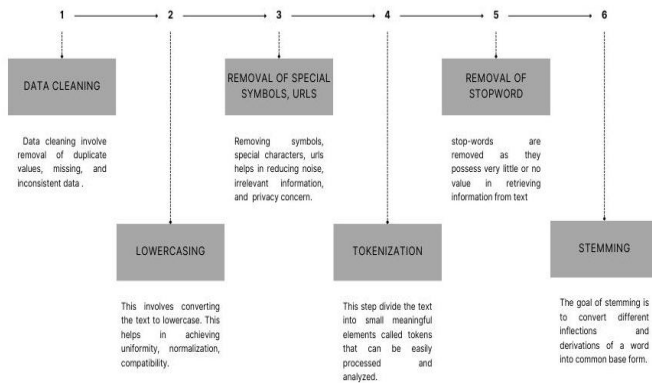


Fig. 3. Steps involve in preprocessing of text

For Hindi Text, the preprocessing steps are the same except for lower casing. Instead of lower casing, normalisation is performed. In Hindi, the text contains various characters and different scripts, which can be normalised with their simple form. The preprocessing steps mentioned earlier are typically used in machine learning architecture. However, in deep learning architectures, the Keras tokenizer and transformer preprocessor were utilised to pre-process the text.

C. Feature Engineering

A crucial aspect of machine learning that plays a significant role in improving model performance and accuracy. By creating meaningful and informative representations of the data, feature engineering enables models to capture underlying patterns and relationships effectively. In this research, we've employed specific feature engineering techniques, namely TF-IDF (Term Frequency-Inverse Document Frequency) for feature extraction and Chi-Square for feature selection.

1) TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF is a widely used technique in natural language processing for quantifying the importance of terms in a document or a collection of documents. It aims to capture the relevance of a term in a document relative to the entire document collection. TF-IDF combines two metrics: TF (Term frequency) and IDF (Inverse document frequency). The higher the TF-IDF score for a term in a document, the more relevant and informative the term is considered to be within that document. Conversely, if a term has a low TF-IDF score, it suggests that the term is less essential or shared across the document collection.

$TF = (\text{Number of repetitions of a term in a document}) / (\text{Total number of terms in a document})$

$IDF = \text{Log}[(\text{Total number of documents}) / (\text{Number of documents containing the term})]$

$$TF-IDF = TF * IDF \quad (1)$$

2) Chi-Square

Statistical method used to identify the most relevant feature in a dataset for predicting a categorical target variable. It measures the association between each feature and target variable by comparing the observed frequency distribution with the expected frequency distribution.

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

χ^2 represents the chi-square symbol, which is a statistical measure.

O_i represents the observed frequency in each category.

E_i represents the expected frequency in each category.

A higher value signifies a stronger dependency of the feature on the target variable. This process effectively reduces the dimensionality of the dataset by focusing on the most relevant features for training.

D. Classifiers

Following the feature extraction and selection the study has employed both Machine learning, Deep learning, and Transfer learning algorithms for classification. For ML-Based classification, the choices include SVM, Naive Bayes, and Logistic regression.

Support Vector Machine (SVM)

A support vector machine (SVM) is a supervised machine learning model that tackles two-group classification problems using classification techniques. SVM seeks an optimal hyperplane that effectively divides the data points into distinct classes, maximizing the margin between the hyperplane and the closest data points from each class, known as support vectors, as illustrated in the figure below. The SVM classifier categorises new text after providing sets of labelled training data for each category. Our research focuses on classifying Fake/Real news, using binary labelled (Fake and Real) datasets. To train the SVM model, we utilize TF-IDF, which assigns weights to words based on their importance in distinguishing between real and fake news. Additionally, the chi-square feature selection selects the most informative words associated with the class label. This strategy helps reduce noise, improve efficiency, and enhance the SVM model's ability to generalise, resulting in improved classifier performance.

Naïve Bayes

We have utilised Naïve Bayes, a probabilistic classifier that is based on Bayes' theorem. Naïve Bayes is also a part of generative learning algorithms for Fake and Real news classification. Naïve Bayes models the joint probability distribution of the input features and the class labels, estimating prior probabilities and the conditional probabilities of the feature given in each class. Despite its assumption of feature independence, Naïve Bayes is efficient and capable of handling high-dimensional spaces. After applying TF-IDF and chi-square for feature extraction and selection, respectively, the model benefits from capturing important words, selecting discriminating features, and utilising a robust classification algorithm, leading to improved performance in classifying fake and real news articles.

Bayes theorem

$$P(Y|X) = (P(X|Y) * P(Y)) / P(X) \quad (3)$$

Naïve Bayes Assumption

$$P(X|Y) = P(X_1|Y) * P(X_2|Y) * \dots * P(X_n|Y) \quad (4)$$

Logistic Regression

A discriminative model in supervised machine learning, is employed for distinguishing between the fake and real news categories. In this approach, TF-IDF (Term Frequency-Inverse Document Frequency) is applied to extract features from the text data, assigning weights to words based on their significance in discerning between the classes. Subsequently, chi-square feature selection is employed to select the most informative words associated with the class labels. By analysing the independent variables, the model estimates the probability of an article belonging to either the fake or real news category. This is accomplished by modelling the log-odd using the formulas given below.

$$\text{Logit}(\pi) = 1/(1+\exp(-\pi)) \quad (5)$$

$$\ln(\pi/(1-\pi)) = \text{Beta}_0 + \text{Beta}_1 * X_1 + \dots + \text{Beta}_k * X_k \quad (6)$$

Where π represents the probability, Beta_0 to Beta_k are the coefficients and X_1 to X_k are the selected independent variables. The aforementioned classifiers are used for classifying English and Hindi news as either fake or Real. Now let's dive into the utilisation of deep learning and Transfer

learning algorithms, which have been employed in conjunction with machine learning algorithms, as discussed below.

LSTM Bi-LSTM

LSTM (Long Short-Term Memory) and Bi-LSTM (Bidirectional Long Short-Term Memory) are recurrent neural network architectures commonly used for sequential data processing tasks such as text classification. LSTM incorporates memory cells and gating mechanisms to selectively retain and update information, allowing it to capture long-term dependencies in the input sequence. Bi-LSTM, on the other hand, performs bidirectional processing by simultaneously considering context from both past and future elements of the sequence, enabling it to capture a more comprehensive understanding of the text. In this study, both Dual Layer LSTM and Bi-LSTM architectures have been employed for the classification of fake and real news. Dual layer LSTM refers to the utilization of two LSTM layers stacked on top of each other. This architectural configuration allows for a more complex representation of the input sequence.

E. Transfer Learning

Machine learning technique that involves leveraging knowledge acquired from one task to tackle another related task. In our study, we have employed several pre-trained models, namely BERT (Bidirectional Encoder Representation from Transformers), ALBERT (A Lite BERT), and DistilBERT (Distilled BERT), for the classification of English news articles as either fake or real. Additionally, for Hindi news articles, we utilised mBERT (Multilingual BERT) and DistilmBERT (Distilled Multilingual BERT).

These models have undergone pre-training using Masked Language Modelling (MLM) and Next Sentence Prediction (NSP) strategies. The BERT family models consist of stacked transformers encoders, and for our research, we specifically utilised the base version of these pre-trained models. Detailed information about the number of Layers (L), Attention heads (A), Hidden size (H), and parameters for each model can be shown below.

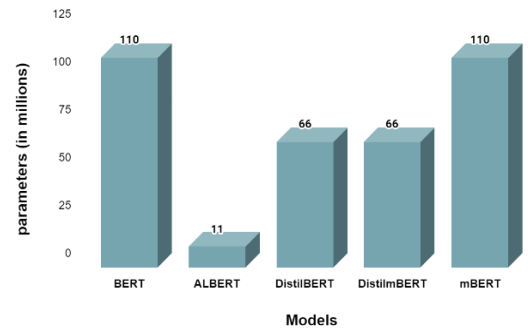


Fig. 4. Models with number of parameters

F. Performance Evaluation

The classification models were evaluated using various metrics, including accuracy, precision, recall, and F1 score.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

By evaluating these metrics, the study was able to gain insights into the classification model's effectiveness in distinguishing between Fake and Real news articles.

IV. RESULTS

After training machine learning classifiers on the training data, we obtained results for English and Hindi data by

utilising feature sets generated through TF-IDF. The performance metrics are depicted as confusion metrics in Fig 5 (a-h) for SVM, Naïve Bayes, Logistic Regression, LSTM, bi-LSTM, BERT, ALBERT, and DistilBERT respectively, applied to English text. Additionally, figure6(a-d) presents the Accuracy, Precision, Recall, and F1 Scores for the models mentioned earlier.

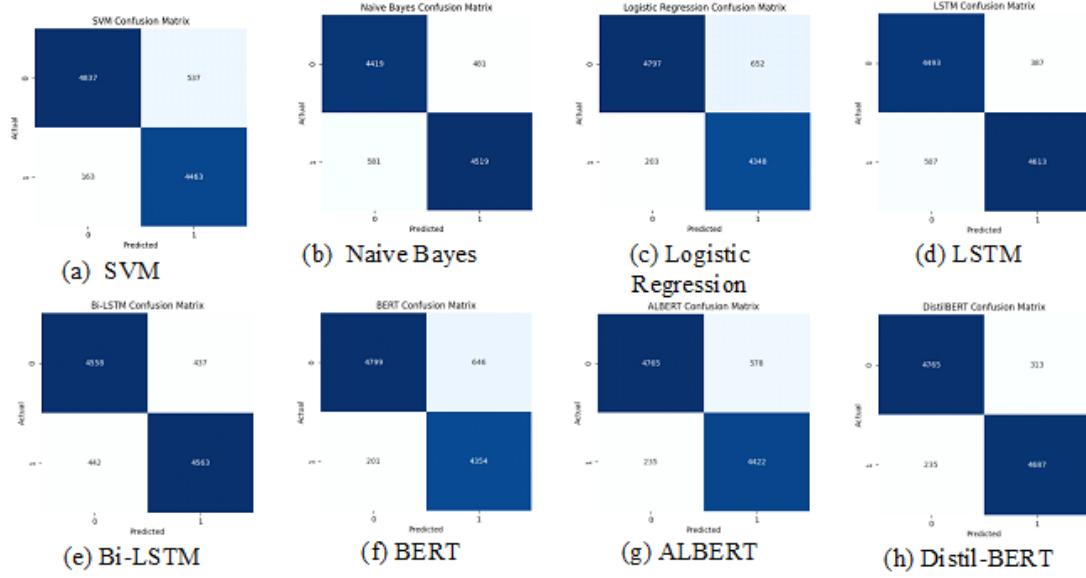


Fig. 5. Confusion matrices (labelled from a to h) depict the performance of diverse machine learning, deep learning, and transfer learning models on an English dataset

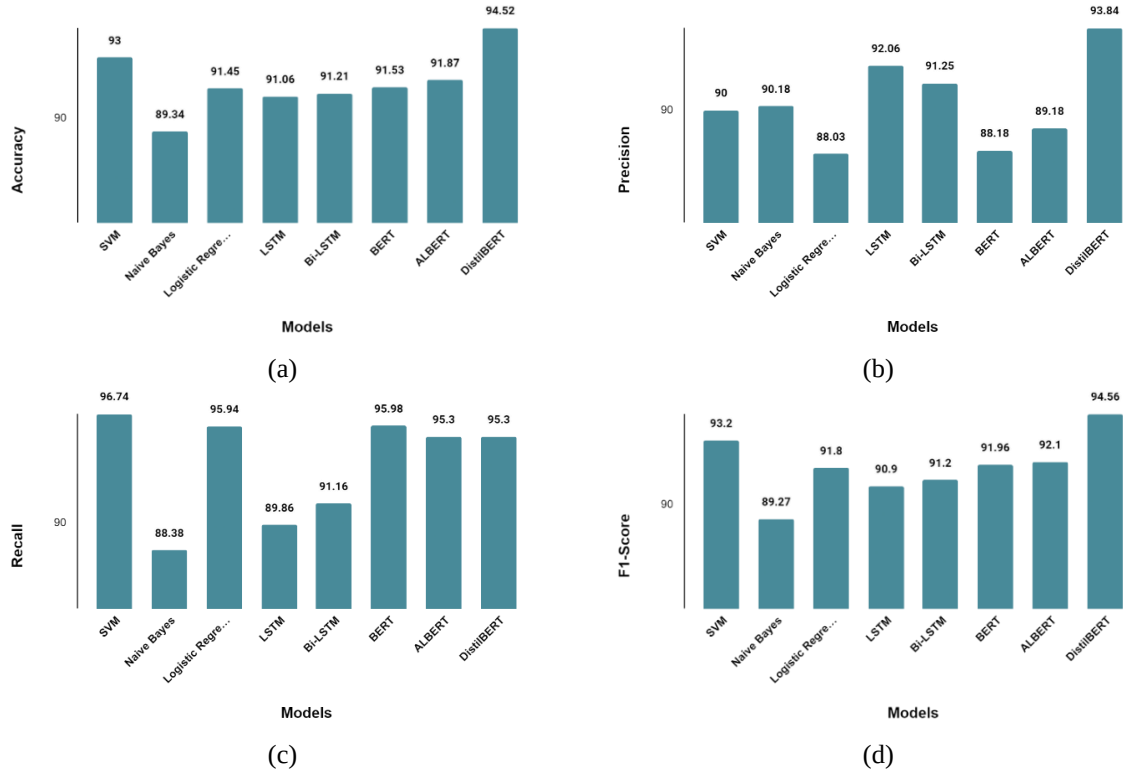


Fig. 6. (a-d) are the outcomes achieved across various evaluation matrices

The findings indicated above were obtained using a feature selection approach, namely by choosing the top 28,000 features using Chi-square from the features collected using TF-IDF from the English dataset. By lowering the dimensionality, this feature selection improves the overall performance and efficacy of machine learning models on the English dataset.

For deep learning and Transfer learning models, we have generated the features using tokenizer and preprocessors for

both the Hindi and English data. Shifting our focus onto the outcomes derived from the Hindi dataset are presented in Fig 7(a-g), showcasing the confusion metrics for SVM, Logistic Regression, Naive Bayes, LSTM, Bi-LSTM, mBERT and Distil-mBERT, respectively. Additionally, figure 8(a-d) displays the other relevant metrics obtained during the evaluation process.

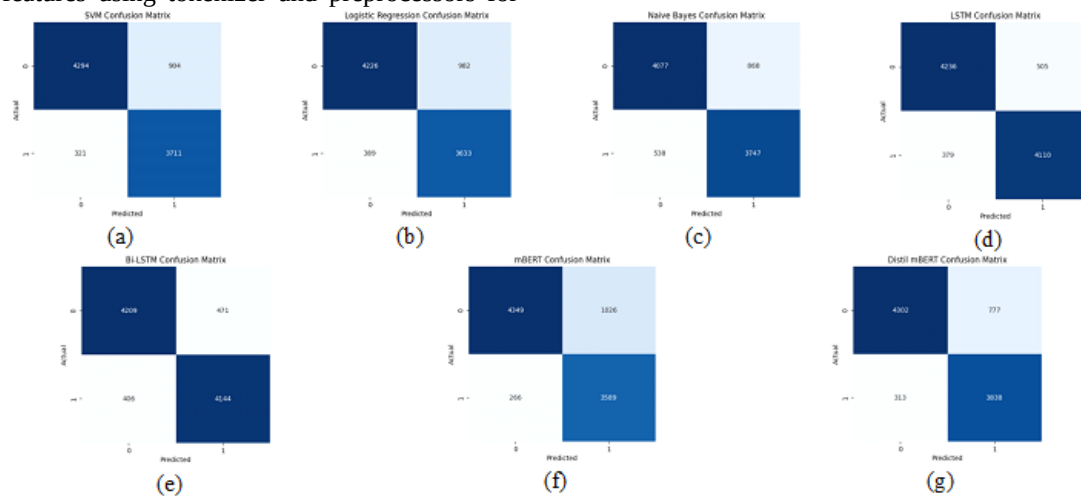


Fig. 7. Confusion matrices (labelled from a to g) depict the performance of diverse machine learning, deep learning, and transfer learning models on Hindi dataset

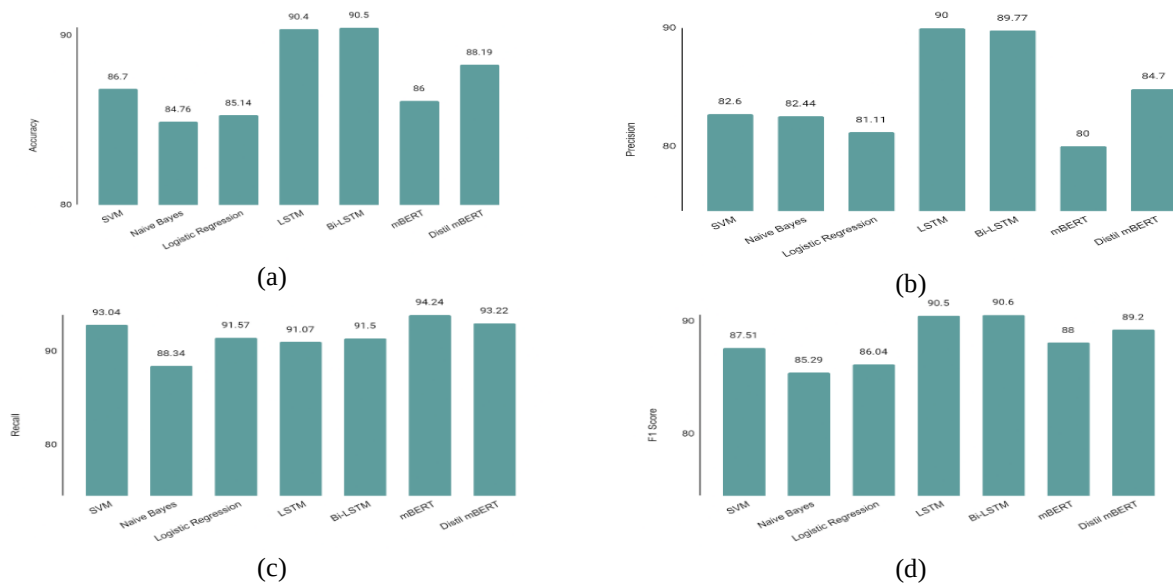


Fig. 8. (a-d) are the outcomes achieved across various evaluation matrices

Results are obtained using the top 30,000 features selected using Chi-square from the features extracted using TF-IDF from the Hindi dataset.

By examining the figures and metrics presented above, we gain valuable insight into the performance of the Hindi dataset. These findings contribute to a comprehensive understanding of the model's effectiveness in handling Hindi Language data, enabling us to provide room for further improvements and optimizations.

Comparison with state-of-the-art works

This section compares the proposed work with other state-of-the-art works to understand its novelty, performance and contribution to the field.

Ahmed (2017) [17] and Mahir (2019) [18] used SVM with TF-IDF in prior studies and obtained an accuracy of 92% and 89.34%, respectively. Khanam (2021) achieved an accuracy of 73% using an SVM and a count vectorizer [19]. However, when tested on the English dataset, our suggested technique

outperforms the earlier research by obtaining an accuracy of 93%.

Granik (2017) [20] and Shaikh (2020) [21] obtained 74% and 84.05% accuracy percentages for the Naive Bayes model. With an accuracy of 89.34%, our work has achieved better results. We also achieved an accuracy of 91.45% when comparing our Logistic Regression model with the TF-IDF and chi-square, outperforming Mahir (2019) and Asiri (2021) [22], who both achieved an accuracy of 69.47% and 87.2%, respectively.

Bi-LSTM and multi-head LSTM were used in Quwasmeh's (2019) research [13], with accuracy rates of 85.3% and 82.94%, respectively. Kumar (2020) attained the maximum accuracy of 88.78% by combining CNN and LSTM with an attention mechanism [12]. However, when comparing the outcomes with our dual-layer LSTM and Bi-LSTM models, we outperformed their efforts with an accuracy of 91.06% and 91.21%, respectively.

Additionally, BERT, ALBERT, and DistilBERT transfer learning models are used in our work. These models have an accuracy of 91.53%, 91.81%, and 94.52%, respectively. This outperforms the work of Shishah (2022) [16], who achieved the greatest accuracy of 85% on the Covid-19 dataset using JointBERT, and Kula (2021), who achieved an F1-Score of 0.90 using RoBERTa [15].

The aforementioned comparison of the proposed work to other current results was done in English news. Let's now contrast our work on the Hindi dataset with another recent research. However, there are very few works that have been proposed on identifying fake news in the Hindi language.

Sreelakshmi (2020) [9] utilised SVM-RBF (Radial Bias Function) with fastText to get an accuracy of 85.81% for mixed Hindi-English tweets, whereas Bharti (2017) [23] used SVM to achieve an accuracy of 79.4% for Hindi tweets. Our study on the Hindi dataset achieved 86.7% when comparing these results and evaluating SVM with TF-IDF and chi-square. However, Bhardwaj (2020) [24] earned an F1 score of 83.98% for the Hindi dataset's coarse-grained assessment of LR, whereas LR received an F1 score of 68.15% for the fine-grained evaluation.

Deep learning models including CNN and LSTM as well as other embedding strategies were suggested by Shailendra (2022) [25], CNN+Bi-LSTM has a word2vec embedding accuracy of 66%, IndicNLPfastText with LSTM has an LSTM embedding accuracy of 61%, and CNN+Bi-LSTM with Facebook's FastText embedding accuracy of 76%. Our dual-layer LSTM and Bi-LSTM have outperformed competitors, with accuracy rates of 90.4% and 90.5%, respectively.

V. CONCLUSION

In this study, we investigated a variety of deep learning, machine learning, and transfer learning models for efficiently identifying false news from actual news in Hindi and English. A custom dataset was constructed that included a variety of news categories rather than favouring just one. We used TF-IDF for feature extraction and Chi-Square for feature selection since manual feature engineering is required for machine learning. Feature extraction using TF-IDF and feature selection using Chi-Square are employed. TF-IDF captures word

importance, while Chi-Square refines feature selection by selecting top features. With an accuracy of 93% in English and 86.7% in Hindi, the SVM model has surpassed the other two models in machine learning. Our dual-layer Bi-LSTM model, which outperforms the LSTM model in deep learning models, has obtained 91.21% in English and 90.5% in Hindi. DistilBERT achieved an accuracy of 88.19% on the Hindi dataset, and DistilBERT achieved 94.52% accuracy in English using the BERT-based technique.

Despite the improvements acquired in the fight against false news, there is still a lot of potential for researchers to expand their methods and their knowledge of the problem. Finding the patterns of false news dissemination and creating early detection techniques are two major areas of research. Researchers can identify common traits and mechanisms that can help identify false news by researching how it travels through digital networks. Furthermore, manipulated photographs and videos may also be used to identify bogus news, opening up a new area for investigation.

REFERENCES

- [1] N. Martin, "How social media has changed how we consume news. Forbes, 2018." 2018.
- [2] I. Ahmad, M. Yousaf, S. Yousaf, and M. O. Ahmad, "Fake news detection using machine learning ensemble methods," *Complexity*, vol. 2020, pp. 1–11, 2020.
- [3] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional neural network with margin loss for fake news detection," *Inf. Process. Manag.*, vol. 58, no. 1, p. 102418, 2021.
- [4] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, 2017.
- [5] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Inf. Process. Manag.*, vol. 57, no. 2, p. 102025, 2020.
- [6] "Natural Language Processing (NLP): 7 Key Techniques," *MonkeyLearn Blog*. Accessed: Jan. 19, 2024. [Online]. Available: <https://monkeylearn.com/blog/natural-language-processing-techniques/>
- [7] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, "Unsupervised fake news detection on social media: A generative approach," in *Proceedings of the AAAI conference on artificial intelligence*, 2019, pp. 5644–5651. Accessed: Jan. 19, 2024. [Online]. Available: <https://aaai.org/ojs/index.php/AAAI/article/view/4508>
- [8] U. Sharma, S. Saran, and S. M. Patil, "Fake news detection using machine learning algorithms," *Int. J. Creat. Res. Thoughts IJCRT*, vol. 8, no. 6, pp. 509–518, 2020.
- [9] K. Sreelakshmi, B. Premjith, and K. P. Soman, "Detection of hate speech text in Hindi-English code-mixed data," *Procedia Comput. Sci.*, vol. 171, pp. 737–744, 2020.
- [10] J. Ma et al., "Detecting rumors from microblogs with recurrent neural networks," 2016, Accessed: Jan. 19, 2024. [Online]. Available: https://ink.library.smu.edu.sg/sis_research/4630/
- [11] D.-H. Lee, Y.-R. Kim, H.-J. Kim, S.-M. Park, and Y.-J. Yang, "Fake news detection using deep learning," *J. Inf. Process. Syst.*, vol. 15, no. 5, pp. 1119–1130, 2019.
- [12] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, "Fake news detection using deep learning models: A novel approach," *Trans. Emerg. Telecommun. Technol.*, vol. 31, no. 2, p. e3767, Feb. 2020, doi: 10.1002/ett.3767.
- [13] E. Qawasmeh, M. Tawalbeh, and M. Abdullah, "Automatic identification of fake news using deep learning," in *2019 Sixth international conference on social networks analysis, Management and Security (SNAMS)*, IEEE, 2019, pp. 383–388. Accessed: Jan. 19, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8931873/>
- [14] T. Zhang et al., "BDANN: BERT-based domain adaptation neural network for multi-modal fake news detection," in *2020 international joint conference on neural networks (IJCNN)*, IEEE, 2020, pp. 1–8. Accessed: Jan. 19, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9206973/>

- [15] S. Kula, R. Kozik, M. Choraś, and M. Woźniak, "Transformer Based Models in Fake News Detection," in *Computational Science – ICCS 2021*, vol. 12745, M. Paszynski, D. Kranzlmüller, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. A. Sloot, Eds., in *Lecture Notes in Computer Science*, vol. 12745, Cham: Springer International Publishing, 2021, pp. 28–38. doi: 10.1007/978-3-030-77970-2_3.
- [16] W. Shishah, "JointBert for Detecting Arabic Fake News," *IEEE Access*, vol. 10, pp. 71951–71960, 2022.
- [17] H. Ahmed, I. Traore, and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," in *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, vol. 10618, I. Traore, I. Woungang, and A. Awad, Eds., in *Lecture Notes in Computer Science*, vol. 10618, Cham: Springer International Publishing, 2017, pp. 127–138. doi: 10.1007/978-3-319-69155-8_9.
- [18] E. M. Mahir, S. Akhter, and M. R. Huq, "Detecting fake news using machine learning and deep learning algorithms," in *2019 7th international conference on smart computing & communications (ICSCC)*, IEEE, 2019, pp. 1–5. Accessed: Jan. 19, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8843612/>
- [19] Z. Khanam, B. N. Alwasel, H. Sirafi, and M. Rashid, "Fake news detection using machine learning approaches," in *IOP conference series: materials science and engineering*, IOP Publishing, 2021, p. 012040. Accessed: Jan. 19, 2024. [Online]. Available: <https://iopscience.iop.org/article/10.1088/1757-899X/1099/1/012040/meta>
- [20] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in *2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON)*, IEEE, 2017, pp. 900–903. Accessed: Jan. 19, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8100379/>
- [21] J. Shaikh and R. Patil, "Fake news detection using machine learning," in *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (ISSSC)*, IEEE, 2020, pp. 1–5. Accessed: Jan. 19, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9358890/>
- [22] N. Islam et al., "Termion: An autonomous model for fake news detection," *Appl. Sci.*, vol. 11, no. 19, p. 9292, 2021.
- [23] S. K. Bharti, K. Sathya Babu, and S. K. Jena, "Harnessing Online News for Sarcasm Detection in Hindi Tweets," in *Pattern Recognition and Machine Intelligence*, vol. 10597, B. U. Shankar, K. Ghosh, D. P. Mandal, S. S. Ray, D. Zhang, and S. K. Pal, Eds., in *Lecture Notes in Computer Science*, vol. 10597, Cham: Springer International Publishing, 2017, pp. 679–686. doi: 10.1007/978-3-319-69900-4_86.
- [24] M. Bhardwaj, M. S. Akhtar, A. Ekbal, A. Das, and T. Chakraborty, "Hostility Detection Dataset in Hindi." *arXiv*, Nov. 06, 2020. Accessed: Jan. 19, 2024. [Online]. Available: <http://arxiv.org/abs/2011.03588>
- [25] P. Shailendra, M. Rashmi, S. Ramu, and R. M. R. Guddeti, "Fake News Detection in Hindi Using Embedding Techniques," in *2022 IEEE Region 10 Symposium (TENSYP)*, IEEE, 2022, pp. 1–5. Accessed: Jan. 19, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9864378/>