International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 15(2023) pp. 59-68

Submitted: 5 Jan., 2023; Accepted: 20 May, 2023; Publish: 9 June, 2023

A Machine Learning Perspective on Fake News Detection: A Comparison of Leading Technqiues

Virendra Singh Nirban[1], Tanu Shukla, Partha Sarathi Purkayastha, Nachiket Kotalwar, and Labeeb Ahsan

[1]Birla Institute of Technology and Science, Pilani, India,
Vidya Vihar, Pilani, Rajasthan 333031,
nirban@pilani.bits-pilani.ac.in

Abstract: **The exponential growth of social media has yielded several advantages, but it has also brought about a major challenge in the form of "fake news", which has become a substantial hindrance to journalism, freedom of expression, and democracy at large. The purpose of this study was to examine the current AI techniques employed for detecting fake news, determine their limitations, and compare them with the latest models. The performance of memory-based and Ensemble methods (LSTM, Bi-LSTM, BERT, Distilled BERT, XGBoost, and AdaBoost) was compared with traditional methods, and the impact of ensemble learning was evaluated. The study aimed to identify appropriate models for fake news detection in order to facilitate a secure and reliable environment for information sharing on social media and ultimately counteract the spread of false information.**
Keywords: Fake news, Machine learning, Ensemble Learning, Artificial Intelligence, Social media

I. INTRODUCTION

The phrase "fake news" has become a buzzword in recent years, however, there is still a lack of agreement on a definitive definition for the term. But, in accordance with recent studies, the definition that has been widely adopted is that "fake news" refers to a piece of information that is presented as an article and contains falsehoods, which can be fact-checked and was created with the deliberate intent of deceiving the reader. The purpose of publishing this false information is to manipulate the beliefs or actions of the reader, making it a significant threat to credible journalism, freedom of expression, and the democratic process as a whole. In the political realm, fake news has been used as a tool for propaganda and manipulation, influencing elections and shaping public opinion. Research has shown that fake news can have a significant impact on election outcomes, spreading false information about candidates and issues (Allcott & Gentzkow, 2017). This threatens the democratic process and can lead to election interference. The spread of fake news can also contribute to political polarization by spreading false information and fueling division within society (Bharali & Goswami, 2018), making it difficult for people to engage in constructive dialogue and find common ground. The proliferation of false information and conspiracy theories related to the COVID-19 pandemic had been facilitated by the utilization of social media platforms (Rocha et al., 2021; Awan et al., 2022). An analysis of the impact of misinformation on health reveals the widespread occurrence of infodemic knowledge, leading to a decrease in trust towards government entities, researchers, and healthcare professionals (Banai et al., 2021). The consequences of misinformation can result in negative psychosocial outcomes such as panic, depression, fear, fatigue, and an increased risk of infection. Furthermore, some findings indicate the potential for fake news to be utilized for covert manipulation of behavior (Bastick ,2021). It posits that existing efforts to address fake news and disinformation are inadequate in safeguarding social media users from this threat, and underscores the significance of this issue for democratic systems. The authors call for an immediate and interdisciplinary effort to examine, safeguard against, and mitigate the dangers of covert, widespread, and decentralized behavior modification through online social networks. The impact of fake news is not limited to the political and social spheres, it can also have a significant impact on financial markets and the economy (Boudoukh et al., 2019). For example, false information about a company's financial performance can lead to a drop in stock prices, affecting investors and potentially leading to economic instability.

Tagging social media articles with flags or labels has been identified as one of the important strategies. In a study with 717 participants, research from Gaozhao (2020) found that people relied heavily on the flags identifying a post as fake irrespective of their political background. Users are not good at differentiating misinformation from genuine information, due to lazy reasoning - reluctance to think critically, and motivated reasoning - thinking on the bases of preconceptions, confirmation bias and conservatism. It also found that the flags used to tag fake news on the internet are highly influential, irrespective of them being expert based or crowdsourced. They don't promote thinking, rather act as a powerful influence. Hence the accuracy of applying flags is of paramount importance in solving the problem. The growing

problem of fake news has prompted a shift towards the use of Artificial Intelligence for its detection and classification. AI-based classification is a field in computer science that strives to emulate human intelligence through the use of specialized hardware and software. To accomplish this, machine learning algorithms are written, trained, and implemented to form correlations between inputs and outputs, allowing the model to make predictions about future states. AI programming focuses on three primary cognitive skills: learning, reasoning, and self-correction. 1. Learning - The main focus here is on collecting data and creating algorithms providing step-by-step instructions to turn raw data into something tangible and accomplish a specific task. 2. Reasoning - The main focus here is on identifying the best algorithm to achieve the desired result. 3. Self-correction - The main focus here is to continuously fine-tune and improve algorithms to provide the most accurate results possible. The success of AI-based classification largely relies on the availability of large amounts of labeled training data. This allows the algorithms to analyze millions of examples and learn to distinguish fake news from authentic information. As AI models continue to evolve, they are expected to become more effective at identifying and flagging fake news, making it easier to combat this growing problem. With the help of machine learning classifiers, fake news in social media can be detected with an accuracy of over 90% (Nistor & Zadobrischi, 2022).

II. METHODOLOGY

A. Objective

To compare the performance of traditional machine learning algorithms with deep learning algorithms and Ensemble Models in detecting fake news.

B. Research Hypothesis

Alternative Hypothesis(H_1) - Deep learning algorithms and Ensemble models have a significantly better performance for fake news detection compared to traditional machine learning algorithms.

C. Research Design

The research technique of content mining has been employed in this study. Two datasets, WELFake and ISOT, were utilized for this purpose. WELFake was generated by combining four major news datasets (Kaggle, McIntire, Reuters, and BuzzFeed Political) and encompasses 72,134 news articles, comprising 35,028 genuine articles and 37,106 false news articles. This method of analysis allows for the extraction of meaningful information from text through the utilization of computational methods to identify patterns from extensive amounts of data. The larger size of the WELFake dataset was implemented to prevent over-fitting of classifiers on smaller datasets and improve the model's training process. The dataset was designed with the aim of training future models utilising real world sources fro detecting fake articles. (Verma et al.); The ISOT dataset was created using genuine sources, and the included fake articles were collected based on flags from Politifact, an organization that verifies the accuracy of political news, and from Wikipedia. The dataset,

ID	Unique ID for a News Article
Title	The Title of a News Article
Text	The Text of the Article (can be incomplete)
Label	Indicates Reliability; 0 for unreliable 1 for reliable

Table 1: Features of the WELFake Dataset

Title	Title of the News Article
Text	Content of the Article
Type	Real or Fake
Date	Date on which the Article was published

Table 2: Features of the ISOT Dataset

comprising over 25,000 news articles, was developed with the aim of enhancing the training process of models used to identify fake news (Ahmed H et al.). The utilization of genuine sources in creating the ISOT dataset provides a more realistic and practical approach to identifying and detecting fake news articles. By including flagged articles, the dataset incorporates potential biases and misinformation that can be present in real-world situations. The training data set used has four features as shown in Table 1.

In our study of fake news detection, we concentrated on the text features and applied word embedding methods to enhance the classification of news articles. The labels for reliable news were assigned as 0 and unreliable news as 1.

We evaluated six AI models, including Naive Bayes, Support Vector Machines (SVMs), Long Short-Term Memory (LSTMs), Bi-LSTMs, BERT, and Distilled BERT to determine their efficacy in detecting fake news. Subsequently, we utilized Ensemble approaches - Random Forest Regressions, XGBoost and AdaBoost to draw an overall comparison between Memory-based techniques and the other existing techniques in use.

D. Measure Design

The models were evaluated using Accuracy, Precision, Recall, F1-score, and Specificity as metrics. The values of these metrics were obtained on a ratio scale, which enabled the comparison of their absolute values.

The proportions of True Positives, True Negatives, False Positives, and False Negatives generated by the models determined the definition of the following metrics:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F - score = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$$

$$Specificity = \frac{TN}{TN + FP}$$

True Positives (TP) : A scenario where the model correctly identifies a true outcome as true.

True Negatives (TN) : A scenario where the model correctly identifies a false outcome as false.
False Positives (FP) : A scenario where the model misidentifies a false outcome as true.
False Negatives (FN) : A scenario where the model misidentifies a true outcome as false.

The evaluation was performed using the sklearn, keras, SVM, and nltk libraries in the Python language on the Google Collab platform.

III. Algorithms

The basic workings of the algorithms used are shown below (Dasaradh, 2020) :

A. Naive Bayes Classifier

The Naive Bayes Classifier is a machine learning technique that uses Bayes' theorem to make predictions. This method assumes that all features are independent of each other, hence the name "Naive." Bayes' theorem states that the probability of an event happening, given that another event has already taken place, can be calculated using the formula:

$$P(\theta|\mathbf{x}) = P(\theta)\frac{P(\mathbf{x}|\theta)}{P(\mathbf{x})}$$

In the case of multiple outcomes (n), the chain rule can be applied to determine the probability of each outcome as below- $\mathrm{P}(\theta, x_1, ..., x_\mathrm{n}) = P(x_1, ..., x_\mathrm{n}, \theta)$
$= P(x_1|x_2, ..., x_\mathrm{n}, \theta).P(x_2, ..., x_\mathrm{n}, \theta)$
$= P(x_1|x_2, ..., x_\mathrm{n}, \theta).P(x_2|x_3, ..., x_\mathrm{n}, \theta)...P(x_\mathrm{n}|\theta).P(\theta)$
Thus the probability can now be measured as -
$\mathrm{P}(\theta|x_1, ..., x_\mathrm{n}) = \frac{P(\theta).\prod_{i=1}^{n} P(x_\mathrm{i}|\theta)}{P(x_1).P(x_2)...P(x_\mathrm{n})}$
This formula forms the basis of the Gaussian Naive Bayes classifier, which is trained on the frequency of words in an article. The independence assumption allows the classifier to make fast and efficient predictions, even when dealing with large amounts of data. By using this method, the classifier can quickly identify articles that contain fake news and distinguish them from credible sources.

B. Support vector machine (SVM)

Support vector machines (SVMs) are a well-established method in the field of machine learning for binary classification problems. The goal of an SVM is to determine a boundary or hyperplane that separates the data points into different classes with maximum margins. The decision boundary is based on the attributes used to classify the data points, with the number of dimensions determined by the number of features involved.
We can show the Cost function for SVM models as:

$$J(\theta) = \frac{1}{2}\sum_{j=1}^{n} \theta^2{}_\mathrm{j}$$

And shown as such:

$$\theta^T x^{(i)} \geq 1, y^{(i)} = 1,$$

$$\theta^T x^{(i)} \leq -1, y^{(i)} = 0.$$

The function here uses a linear kernel, SVMs use by default, but other kernel functions can be used in cases where the data points are not easily separable or are multi-dimensional. We employed the Radial Basis Function (rbf) kernel for training purposes.

C. Neural Networks

Neural Networks are complex models that consist of a series of interconnected layers including the input layer, hidden layers, and the output layer. They utilize simple units known as neurons to make predictions. The two key processes involved in a Neural Network are the Forward Propagation and the Learning Process.
In Forward Propagation, each input x is assigned with a weight ($\mathrm{w_i}$) that represents the strength of the connection between the neurons, and a bias (b) is added to the equation. The calculation for z is represented mathematically as:

$$z = x.w + b$$

To add non-linearity to the model, the result from the above calculation is then passed through a Sigmoid function. The sigmoid function maps the output to a value between 0 and 1 and is defined as:

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}}$$

Here σ denotes the sigmoid activation function, and the output obtained is known as the predicted value ($\hat{y}$).
The optimization of the Neural Network is done using the Gradient Descent algorithm. The algorithm adjusts the weights and biases in proportion to the gradient of the cost function. The result of a single neuron can be expanded to the entire Neural Network with some modifications.

D. Long Short Term Memory Networks (LSTMs)

In a sentence, the relationship between words is crucial for the proper classification of articles. Traditional neural networks, however, have limitations in retaining memories of previous events that could impact future ones. To tackle this problem, Recurrent Neural Networks (RNNs) have been developed. These networks consist of loops that allow them to store and utilize previous events. Among RNNs, Long Short Term Memory Networks (LSTMs) are a specialized version designed to handle long-term relationships between words in a sentence. These models are particularly useful in fake news detection, where understanding the context and interdependence of words can provide important information in determining the authenticity of the news.
The Architecture of LSTMs is shown below in Figure 1. (Socher, 2015).

Input gate: $i^{(t)} = \sigma(W^{(i)}x^{(t)} + U^{(i)}h^{(t-1)})$

Forget gate: $f^{(t)} = \sigma(W^{(f)}x^{(t)} + U^{(f)}h^{(t-1)})$

Output/Exposure gate: $o^{(t)} = \sigma(W^{(o)}x^{(t)} + U^{(o)}h^{(t-1)})$

New Memory Cell: $\tilde{c}^{(t)} = \tanh(W^{(c)}x^{(t)} + U^{(c)}h^{(t-1)})$

Final Memory Cell: $c^{(t)} = f^{(t)} \circ \tilde{c}^{(t-1)} + i^{(t)} \circ \tilde{c}^{(t)})$

Hidden State: $h^{(t)} = o^{(t)} \circ \tanh(c^{(t)})$

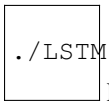

Figure 1 : Architecture of LSTMs.

As shown in Figure 1, the structure of a Long Short-Term Memory (LSTM) network involves five stages in its computation:

- Memory Generation: The process of generating a new memory cell, denoted as $\tilde{c}^{(t)}$, that takes into account both the current input word, $x^{(t)}$, and the previous hidden state, $h^{(t-1)}$.
- Input Gate: This gate determines the importance of the input word, $x^{(t)}$, and the previous hidden state, $h^{(t-1)}$, in the formation of the new memory cell. The input gate produces an indicator,$i^{(t)}$, that reflects this determination.
- Forget Gate: This gate assesses the relevance of the previous memory cell, $c^{(t-1)}$, for the current memory cell formation. The forget gate takes into consideration the input word, $x^{(t)}$, and the previous hidden state, $h^{(t-1)}$, to produce an indicator, $f^{(t)}$.
- Final Memory Generation: The stage of combining the effects of the input and forget gates to produce the final memory cell, $c^{(t)}$. This stage forgets the past memory cell, $c^{(t-1)}$, based on the advice of the forget gate, $f^{(t)}$, and incorporates the new memory cell, $\tilde{c}^{(t)}$, based on the input gate's indicator, $i^{(t)}$.
- Output/Exposure Gate: The purpose of this gate is to control the exposure of the final memory cell, $c^{(t)}$, to the hidden state,$h^{(t)}$. The output gate produces an indicator, $o^{(t)}$, that reflects which aspects of the final memory cell, $c^{(t)}$, should be included in the hidden state. The output gate is used to gate the pointwise tanh of the memory to produce the hidden state.

These five stages allow the LSTM cell to store and process information over time and make predictions based on that information. The gates in the LSTM structure provide the ability to selectively store, forget, and output information, which enables the cell to effectively model long-term dependencies in sequential data.

E. Bi-directional Long Short Term Memory networks(Bi-LSTMs)

A Bidirectional long-short term memory (Bi-LSTM) is a neural network architecture that is designed to capture the relationships between elements in a sequence of data in two directions, i.e., both from the past to the future and from the future to the past. This is accomplished by incorporating extra layers into the model that allow the input to flow in both directions. In comparison to traditional LSTM models, which only consider the relationships in a single direction, Bi-LSTMs offer a more comprehensive view of the relationships between elements in a sequence, enabling them to better capture the long-term dependencies(Verma, 2021).

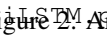

Figure 2: Architecture of a Simple Bi-LSTM Model

In a Bidirectional LSTM model, information flows in both the past to future and future to past directions. This architecture is beneficial when working with sequence-to-sequence problems, making it an advantageous tool in the context of fake news detection.

The Bi-LSTM architecture involves several key components, including the inputs, forward layers, backward layers, activation layers, and outputs. Each of these components plays a crucial role in the functionality of the Bi-LSTM.

- Input: The input in a Bi-LSTM architecture is typically a sequence of vectors, representing either words or subwords. These vectors are then processed by the forward and backward layers.
- Forward Layers: The forward layers are responsible for processing the input in a forward direction, from the beginning to the end of the sequence. These layers consist of a series of LSTM cells, which apply the input gate, forget gate, output gate, and final memory generation operations to the input sequence.
- Backward Layers: The backward layers are similar to the forward layers, except that they process the input sequence in the opposite direction, from the end to the beginning. This allows the Bi-LSTM to capture information about both the context preceding and following each word or subword in the input sequence.
- Activation Layers: The activation layers are responsible for transforming the outputs of the forward and backward layers into a more compact representation. This typically involves applying a non-linear activation function, such as the hyperbolic tangent or sigmoid function, to the output of each layer.
- Outputs: The outputs of a Bi-LSTM are a set of feature vectors, representing the input sequence in a reduced and transformed form. These feature vectors can then be used for various tasks, such as sentiment analysis, named entity recognition, or, as in our case, Fake News Detection.

By incorporating information from both directions, the Bi-LSTM can provide a more comprehensive understanding of the relationships between elements in a sequence, improving the overall performance of the model.

F. Bidirectional Encoder Representations from Transformers (BERT)

Bidirectional Encoder Representations from Transformers(BERT) is a transformer-based architecture used for NLP tasks such as text classification, language translation, and question-answering. The model was introduced by Google in 2018 and has since become one of the most widely used NLP models in industry and research.(Devlin et al., 2018).

The architecture of the BERT model consists of the following components:

- Embedding Layer: The input to the model is a sequence of tokens (words or sub-words). These tokens are first passed through an embedding layer to convert them into dense vectors. The embedding layer is usually initialized with pre-trained weights from a large corpus of text.

- Encoder Layers: The core of the BERT model is its encoder, which is composed of multiple stacked transformer blocks. The transformer blocks consist of a self-attention mechanism and a fully connected feedforward network. The self-attention mechanism allows the model to capture long-range dependencies between tokens.

- Pooling Layer: The output from the final encoder layer is passed through a pooling layer to generate a fixed-length representation for each input sequence. This fixed-length representation is then used as input to a classifier for the NLP task at hand.

- Pretraining: The BERT model is pre-trained on large amounts of unsupervised text data, where it is trained to predict the masked tokens in a sentence given the rest of the sentence. This pre-training allows the model to learn rich representations of the language that can be fine-tuned for specific NLP tasks.

- Fine-Tuning: Once the pre-training is completed, the BERT model can be fine-tuned on smaller labeled datasets for specific NLP tasks such as language translation, question-answering, and in our case, text classification for Fake News Detection . The fine-tuning process involves adding a task-specific layer on top of the pre-trained model and training the model on the smaller labeled dataset.

G. Distilled BERT

Distilled BERT is a smaller, faster, and more efficient version of the original BERT model. The architecture of distilled BERT is similar to the original BERT model but it is designed to have fewer parameters and therefore requires less computational power.
The architecture of distilled BERT is based on the Transformer architecture, which is a type of neural network that is well suited for processing sequences of data such as natural language text. The main components of the distilled BERT model are the encoding layers and the attention mechanism.
The encoding layers are responsible for converting the input text into a fixed-length representation that can be used by the rest of the model. These layers use a combination of linear transformations and activation functions to generate this representation.
The attention mechanism is responsible for allowing the model to focus on different parts of the input sequence at different times. This mechanism uses self-attention, which allows the model to consider the relationships between all elements of the input sequence in a parallel fashion.
In addition to these core components, distilled BERT also uses several other techniques to improve its efficiency and accuracy, such as knowledge distillation and quantization.
Knowledge distillation is a process in which a smaller model is trained to mimic the outputs of a larger model. The smaller model is trained using a combination of the original training data and the outputs of the larger model.
Quantization is a technique that reduces the precision of the model's parameters, which reduces the amount of memory required to store the model and speeds up its inference time.
By combining these techniques, distilled BERT can achieve similar performance to the original BERT model but with much fewer parameters, making it more efficient and easier to deploy in a variety of applications. (Sanh et al., 2019)

H. Ensemble Learning Methods

Ensemble methods have been increasingly used to tackle the challenge of detecting fake news. These techniques combine the predictions of multiple models to achieve improved accuracy and performance compared to individual models. There are several ensemble techniques, including Bagging, Boosting, Stacking, and Blending.
Bagging is a weak learner's model that uses parallel and independent learning from different models and then averages the results to make the final prediction. Boosting, on the other hand, is also a weak learner's model but learns in a sequential manner and adapts to improve the predictions.

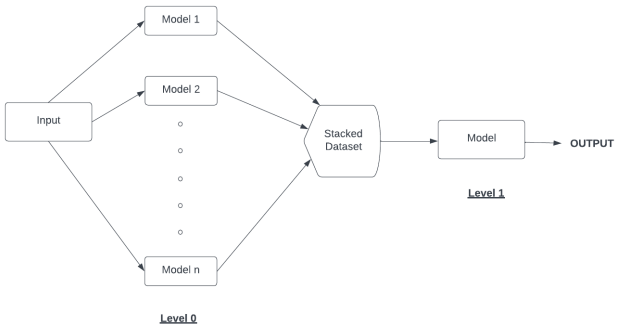

Figure 3 : Concept Diagram of Stacking.

Stacking is a hybrid approach that involves training multiple models in parallel and then combining them by training a meta-model to make predictions based on the predictions of the individual models. This approach is expected to produce more tangible outputs with higher accuracy.
Blending, like stacking, is another ensemble technique that can boost performance and accuracy. It uses a holdout set, separate from the training set, to make predictions and then trains a model using these predictions. (By Great Learning Team, 2022).
Basically, Ensemble learning methods are a class of machine learning techniques that utilize multiple models to obtain improved performance compared to using a single model. AdaBoost (Adaptive Boosting), Random Forest Regression, and XGBoost(Extreme Gradient Boosting) are popular algorithms in the ensemble learning domain.

I. Random Forest Algorithm

Random forest is an ensemble learning technique that combines multiple classifiers to improve the performance and reduce weaknesses of individual models. The algorithm leverages the results of multiple decision trees trained on different subsets of the data, and the outputs are then averaged or voted to produce accurate results.
Random forests are known for their ability to maintain high accuracy even when the dataset has missing values. The algorithm is versatile, and can be applied to various regression and prediction problems, as it requires fewer parameters and can handle complex datasets with high dimensions. It is

Parameters→/ Models↓	Accuracy	Precision	Recall	F1-Score	Specificity
Naïve Bayes	49.84%	70.00%	2.19%	4.24%	99.03%
SVM	85.08%	80.50%	95.27%	87.26%	73.29%
LSTM	85.50%	90.00%	86.40%	88.16%	84.00%
Bi-LSTM	87.46%	93.68%	80.25%	86.45%	94.62%
BERT	93.33%	97.54%	88.78%	92.95%	97.80%
DistilBert	93.33%	94.41%	91.99%	93.18%	94.65%
Random Forest Regression	89.68%	91.19%	88.69%	89.92%	90.76%
XGBoost	75.87%	73.21%	78.07%	75.56%	73.86%
AdaBoost	91.75%	93.35%	90.49%	91.90%	93.09%

Table 3: Comparison among performances of different models (WELFake Dataset)

Parameters→/ Models↓	Accuracy	Precision	Recall	F1-Score	Specificity
Naïve Bayes	50.00%	81.82%	8.41%	15.25%	97.85%
SVM	98.00%	97.83%	98.90%	98.36%	98.17%
LSTM	92.00%	88.18%	97.00%	92.38%	87.00%
Bi-LSTM	96.00%	93.94%	97.89%	95.88%	94.29%
BERT	99.00%	99.99%	98.08%	99.03%	99.99%
DistilBert	96.50%	99.99%	93.27%	96.52%	99.99%
Random Forest Regression	90.50%	91.26%	90.38%	90.82%	90.62%
XGBoost	97.50%	98.96%	95.96%	97.44%	99.01%
AdaBoost	98.50%	97.83%	98.90%	98.36%	98.17%

Table 4: Comparison among performances of different models (ISOT Dataset)

widely used in classification and prediction of univariate and multivariate time series.(Gaurkar S. et. al, 2021).
The implementation process of Random forests involves building a decision tree for each sample set, obtaining the prediction result from each tree, and finally selecting the most voted prediction as the final result. With an increased number of trees in the forest, the accuracy and precision of the solution tends to improve.

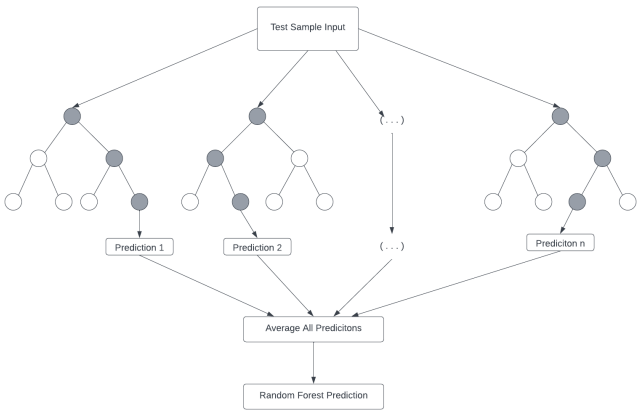

Figure 4 : Concept Diagram of Random Forest Regressions.

J. AdaBoost

Adaptive Boosting (AdaBoost) is a popular ensemble learning method used for binary classification problems. It combines several weak classifiers to form a strong classifier. The algorithm is adaptive in the sense that subsequent weak classifiers are designed specifically to address the mistakes made by previous classifiers.
AdaBoost works by iteratively training a sequence of weak classifiers and then weighting each of the weak classifiers. At each iteration, the weight of each instance is adjusted based on the error made by the previous classifier. Instances that are difficult to classify are given more weight, and instances that are easily classified are given less weight. In this way, subsequent classifiers focus more on instances that have not been accurately classified by previous classifiers.(Freund & Schapire, 1997)
The final prediction of the AdaBoost algorithm is the weighted sum of the predictions of all weak classifiers. In other words, each weak classifier makes its prediction and contributes a weight to the final prediction, with the weights proportional to the classifier's accuracy.
AdaBoost can be applied to any binary classification problem, as long as a weak classifier can be defined. The algorithm has been applied to a variety of problems, including object recognition, speech processing, and biological sequence analysis, among others. The algorithm is relatively fast, robust to noisy data, and is less likely to over-fit compared to other ensemble methods like random forests.
AdaBoost has been thoroughly studied and its theoretical properties have been well established. It has been shown that AdaBoost converges to the Bayes classifier under mild conditions, and it has been proven that the algorithm is capable of achieving a low generalization error rate (i.e., a high accuracy) if the weak classifiers are sufficiently accurate and diverse.

K. XGBOOST

The XGBoost (eXtreme Gradient Boosting) is an open-source software library for gradient boosting trees designed for speed and performance that has been widely used in machine learning and data science. XGBoost is a variation of the gradient boosting trees algorithm that incorporates several enhancements to make it more computationally efficient and scalable, such as regularization, parallel and

distributed computing, and column subsampling.(Chen & Guestrin, 2016).

The architecture of the XGBoost framework is as follows:

- Base Learners: XGBoost is an ensemble of decision trees. Each decision tree is considered as a weak learner that is built one by one to form a strong learner. The objective is to minimize the loss function through the trees.
- Tree Booster: XGBoost implements the tree booster, a highly efficient gradient boosting tree algorithm that minimizes the loss by building trees in a forward stage-wise manner. At each stage, the algorithm updates the gradient by using the gradient information from the previous stage.
- Regularization: XGBoost implements regularization by adding a penalty term to the loss function that is being optimized. The penalty term controls overfitting and the generalization of the model.
- Parallel and Distributed Computing: XGBoost has been designed for parallel computing and supports distributed computing through a combination of parallel and communication optimizations. This makes it highly scalable and enables it to handle very large datasets.
- Column Subsampling: XGBoost implements column subsampling to reduce the complexity of the model and reduce overfitting. At each stage, the algorithm subsamples the columns of the data and builds the trees only with the subsampled columns.

IV. Results and Discussion

A. Comparing Performances of Models

The observations upon training our Models against the WELFake dataset have been presented in Table 3. When comparing the performance of different models, Naive Bayes showed a low accuracy of 49.84% but a high specificity of 99.03%. The highest specificity but lowest accuracy for Naive Bayes indicates that the model has a high ability to correctly identify instances where news is not fake (True Negatives), but its overall performance in correctly identifying both fake and real news is low. This may mean that the model is conservative in its predictions, as it prioritizes avoiding false negatives (identifying real news as fake), but this also results in missing some instances of fake news (false positives).SVM had a better accuracy of 85.08% and a high recall of 95.27%. LSTM had a similar accuracy of 85.50% but a higher precision of 90.00%. Bi-LSTM showed improved results with an accuracy of 87.46% and a precision of 93.68%. BERT and DistilBERT had the highest accuracy at 93.33% with BERT having a higher F1-Score of 92.95% and DistilBERT having a higher specificity of 94.65%. XGBoost had a lower accuracy of 75.87% while AdaBoost showed 91.75% accuracy with a good balance between precision and recall. Random Forest Regression had an accuracy of 89.68% with an F1-Score of 89.92%.

The observations upon training against the ISOT dataset have been presented in table 4. Here, Naive Bayes had a low accuracy of 50.00% but a high specificity of 97.85%. SVM had an impressive accuracy of 98.00% and a high recall of 98.90%, despite being one of the traditional models used. LSTM had a 92.00% accuracy with a recall of 97.00%. Bi-LSTM showed improved results with 96.00% accuracy and a precision of 93.94%. BERT and DistilBERT both had high accuracy, with BERT at 99.00% and a F1-Score of 99.03%, and DistilBERT at 96.50% with a specificity of 99.99%. XGBoost had a good accuracy of 97.50% and AdaBoost showed a similar accuracy of 98.50% with a precision of 97.83%. Random Forest Regression had an accuracy of 90.50% and a balanced F1-Score of 90.82%.

In both the datasets used, SVMs performed quite well for a Traditional Model, but in the end it was observed that in general Memory based techniques and Ensemble models had a more balanced and effective performance. The balance between the parameters for BERT and DistilBERT suggests that both models have relatively strong performance in terms of accuracy, precision, recall, and F1-score. This indicates that the models are able to correctly classify a high number of true positive cases, while also having a low number of false positive cases. This balance in performance implies that these models have a good ability to detect fake news while avoiding false positive classifications. The results, overall, indicate that BERT and DistilBERT performed better in comparison to the ensemble models and also the existing traditional models in terms of all the parameters.

As seen in the work of Aphiwongsophon et al. (2018), who used SVM and Neural Network algorithms and achieved a 99.90% accuracy rate in detecting fake news on Twitter data. Choudhary M et al. (2021) found a similar trend in their research where SVMs and Naive Bayes were identified as the top performers among the reviewed papers, producing the highest outcomes.

Ahmad et al. (2020) and Vijayaraghavan et al. (2020), found that models using other forms of learning methods, such as LSTMs, Bi-LSTMs, and Ensemble Learning Methods, often outperformed traditional models.

In another study by Sastrawan et al. (2021), the effectiveness of combining Bidirectional LSTMs with pre-trained word embedding was tested and observed to be significantly more effective than unidirectional models, achieving better results on multiple datasets. This is further supported by the results of the study by Bahad et al. (2019), which found that the Bi-LSTM model outperformed the unidirectional LSTM model in terms of accuracy and reduced loss.

The study by Ahmad et al. (2020) proposed the use of Ensemble learning methods for identifying patterns in text that differentiate fake news from true news. The dataset used was obtained from the World Wide Web and contained news articles from various domains, not just politics. The textual features were extracted from the articles and used to train models, which resulted in Ensemble learners consistently outperforming individual learners in terms of accuracy in multiple datasets analyzed. This is further supported by the results of the study by Patil, Dharmaraj (2022), which found that Ensemble Models performed much better than traditional models.

In the study carried out by Rai et al. (2022), it was observed that BERT based models outperformed other traditional models used, in mostly all the metrics considered. Kaliyar R. K. et al.(2021) used FakeBERT, combining single-layer deep CNN blocks with BERT and saw a similiar trends as it outperformed existing models with an accuracy of 98.90% in classification. All these observations support the results of our study, which in turn, are in support of our Alternate Hypothesis showing that Deep learning algorithms and Ensemble models performed significantly better, and in our case, Deep learning algorithms were seen to provide the best performance.
Thus, although it was seen that SVMs performed quite well, the observations show that the more intensive we get into Neural Networks and introduce more aspects like memory, the more accurate the models and their results become.

B. Current Methods

The current methods can be broadly classified into - 1. NLP-based approach: Uses Natural Language Processing (NLP) techniques to analyze the text content of news articles and detect patterns that may indicate fake news. 2. Stylometry-based approach: This approach uses computational methods to analyze the writing style of authors and detect inconsistencies or anomalies that may indicate fake news. 3. Graph-based approach: This approach uses graph theory and network analysis to analyze the spread and sources of news, and detect fake news by tracking its flow through social media networks. 4. Deep learning-based approach: This approach uses deep neural networks to learn the features and patterns of fake news and classify news articles as fake or real. 5. Hybrid approach: This approach combines multiple AI methods and uses ensemble learning to achieve better accuracy in detecting fake news. This can also involve combining features together. Seddari et al. (2022) showed a hybrid approach which employs a combination of linguistic and knowledge based features on social media news, which gives improved results over using the features independently.
Facebook is a hub for the production and dissemination of fake news and misinformation (Allcott et al, 2019). Despite implementing various AI methods, including policies and products, the spread of fake news remains a significant technical challenge, particularly regarding image manipulation, such as Deepfakes. To address this issue, Facebook has developed SimSearchNet++, an advanced image matching model that uses unsupervised learning to track picture changes with high precision and recall. Additionally, to detect misinformation, Facebook is implementing AI systems that automatically detect new variants of discredited content and forward them to its fact-checking partners for review, thus limiting the spread of misinformation to an extent (Facebook AI, 2021).
Apart from these methods, social media companies have adopted two primary strategies to combat misinformation. The first approach is to outright block such content, as seen with Pinterest's ban on anti-vaccination content and Facebook's ban on white supremacist content. The second approach, implemented by YouTube, involves providing alternative and accurate information alongside misleading content. This strategy aims to expose users to correct information, which can help to counteract false claims. For instance, when searching for "Vaccines cause autism" on YouTube, users are presented with a link to the MMR vaccine Wikipedia page that debunks such beliefs, in addition to videos posted by anti-vaxxers. This approach seeks to fight misinformation by providing information, rather than solely blocking it.(Yaraghi, 2022)
It is important to note that AI is still not a foolproof solution for fake news detection, and there is a need for continuous improvement and refinement of these methods.

C. Future Directions of Study

The limitations of our study center around the utilization of Supervised Learning algorithms for fake news detection, which necessitates large datasets from trustworthy sources for both model training and testing. This requirement is often challenging to meet, as such datasets are not readily available online and often require significant effort to compile and verify. Therefore, it is recommended to consider adopting an Unsupervised or Semi-supervised learning approach to mitigate this challenge.
It is important to note that relying solely on the outcome of a crowdsourced poll to make decisions regarding the authenticity of news may not always be the optimal approach, as public opinion can be influenced by misinformation and often contributes to the spread of fake news. To address this issue, various metrics such as the presence of a verified account, history of suspensions or bans, frequency of posting, engagement with content, comments, and length of time active on a platform can be utilized to validate credibility.
The detection of fake news has primarily relied on the examination of text content and semantic information. However, incorporating contextual information, such as the source, date, and geographical location, can increase the accuracy of detection. A cross-platform detection system is necessary given the prevalence of fake news on various platforms such as social media, news websites, and messaging apps. The integration of multimedia information, including images, videos, and audio, can also lead to a more robust fake news detection system. Future studies should also examine the psychological factors that contribute to the spread of fake news and develop methods to counteract them. User-centered approaches that consider user behavior and feedback can result in more accurate and personalized fake news detection. The combination of multiple detection methods, rather than relying solely on machine learning or rule-based systems, has the potential to improve the overall performance of fake news detection systems. It is also crucial to develop systems that are robust against adversarial attacks as fake news detection becomes more sophisticated.

V. Conclusion

The rapid spread of fake news and misinformation on social media platforms has necessitated the implementation of effective solutions to tackle this issue. The use of flagging systems has proven to be successful in controlling the spread of false information, however, with the scale of the problem, the use of automated AI-based classification methods has become increasingly necessary.
From the results obtained, we observed that the Naive Bayes

model performed the worst and SVMs performed the best among traditional methods. Memory-based models (BERT and DistilBERT) performed the best overall and showed a significant improvement in performance over traditional models. Ensemble Models had a better and more balanced performance as compared to traditional models, but still were observed to come up short against the Memory-based models in general. The study overall supports our Alternate Hypothesis that "Deep learning algorithms and Ensemble models have a significantly better performance for fake news detection compared to traditional machine learning algorithms."
As for further directions of study, the detection and prevention of fake news is a pressing issue that requires innovative solutions to curb its spread on social media. To address this, researchers are exploring new avenues to refine existing methods and discover more efficient techniques. One of these is the implementation of Unsupervised Learning, which has the potential to reduce the dependence on labeled datasets and allow the models to adapt more swiftly to diverse news articles. Furthermore, advancements can be made by incorporating Multi-Modal approaches into the analysis, incorporating visual and textual elements of the news articles, instead of solely relying on the text. This would provide a more comprehensive understanding of the information being presented, thereby leading to improved accuracy in detecting fake news.

References

[1] Ahmed H, Traore I, Saad S. "Detecting opinion spams and fake news using text classification", Journal of Security and Privacy, Volume 1, Issue 1, Wiley, January/February 2018

[2] Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O.: Fake News Detection Using Machine Learning Ensemble Methods. Complexity, 1–11 (2020).

[3] Allcott, H., & Gentzkow, M.: Social Media and Fake News in the 2016 Election. Journal of Economic Perspectives, 31(2), 211–236 (2017).

[4] Allcott, H., Gentzkow, M., & Yu, C.: Trends in the diffusion of misinformation on social media. Research & Politics, 6(2) (2019).

[5] Aphiwongsophon, S. and Chongstitvatana, P.: Detecting Fake News with Machine Learning Method. 2018 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON),(2018).

[6] Awan, T. H., Aziz, M., Sharif, A., Raza, T., Jasam, T., & Alvi, Y. Fake news during the pandemic times: A Systematic Literature Review using PRISMA. Open Information Science, 6(1), 49–60. (2022).

[7] Bahad, P., Saxena, P. and Kamal, R.: Fake News Detection using Bi-directional LSTM-Recurrent Neural Network. Procedia Computer Science, 165, pp.74-82. (2019).

[8] Banai, I. P., Banai, B., & Mikloušić, I. Beliefs in COVID-19 conspiracy theories, compliance with the preventive measures, and trust in government medical officials. Current Psychology, 41(10), 7448–7458. (2021).

[9] Bastick, Zach. "Would You Notice if Fake News Changed Your Behavior? An Experiment on the Unconscious Effects of Disinformation." Computers in Human Behavior, vol. 116, Elsevier BV, Mar. 2021, p. 106633. Crossref,

[10] Bharali, Bharati, and Anupa Lahkar Goswami. "Fake News: Credibility, Cultivation Syndrome and the New Age Media." Media Watch, vol. 9, no. 1, SAGE Publications, Mar. 2018. Crossref,

[11] Chen, T., & Guestrin, C. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2016).

[12] Dasaradh, S. K.: "A Gentle Introduction To Math Behind Neural Networks." medium.com (accessed July 1, 2022).

[13] Devlin, J., Chang, M., Lee, K., & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv: Computation and Language. (2018).

[14] Facebook AI.: "Here's how we're using AI to help detect misinformation." Facebook AI. (n.d.). (accessed December 2, 2021)

[15] Freund, Y., & Schapire, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55(1), 119–139. (1997).

[16] Gaozhao, D. (2020). Flagging Fake News on Social Media: An Experimental Study of Media Consumers' Identification of Fake News. SSRN Electronic Journal. Published.

[17] González, S., García, S., Del Ser, J., Rokach, L. and Herrera, F.: A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. Information Fusion, 64, pp.205-237.(2020).

[18] Great Learning Team.: "Ensemble learning with Stacking and Blending." mygreatlearning.com (accessed July 1, 2022).

[19] Kaliyar, R. K., Goswami, A., Narang, P. (2021). FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. Multimedia Tools and Applications, 80(8), 11765–11788.

[20] Li, D., Guo, H., Wang, Z. and Zheng, Z.: Unsupervised Fake News Detection Based on Autoencoder. IEEE Access, 9, pp.29356-29365.(2021).

[21] M. Choudhary, S. Jha, Prashant, D. Saxena and A. K. Singh, "A Review of Fake News Detection Methods using Machine Learning," 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 2021, pp. 1-5.

[22] Nirban, V. S., Shukla, T., Purkayastha, P. S., Kotalwar, N., Ahsan, L. (2023, January 1). The Role of AI in Combating Fake News and Misinformation. Lecture Notes in Networks and Systems. pg. 690-701 (WICT 2022).

[23] Nistor A, Zadobrischi E. The Influence of Fake News on Social Media: Analysis and Verification of Web Content during the COVID-19 Pandemic by Advanced Machine Learning Methods and Natural Language Processing. Sustainability. (2022).

[24] Patil, Dharmaraj. Fake News Detection Using Majority Voting Technique.(2022).

[25] Rai, N., Kumar, D., Kaushik, N., Raj, C., & Ali, A. Fake News Classification using transformer based enhanced LSTM and BERT. International Journal of Cognitive Computing in Engineering, 3, 98–105. (2022).

[26] Rocha, Y., de Moura, G., Desidério, G., de Oliveira, C., Lourenço, F. and de Figueiredo Nicolete, L.: The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. Journal of Public Health,(2021).

[27] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv: Computation and Language. (2019).

[28] Sastrawan, I., Bayupati, I. and Arsa, D.: Detection of fake news using deep learning CNN–RNN based methods. ICT Express,(2021).

[29] Seddari, Noureddine & Derhab, Abdelouahid & Belaoued, Mohamed & Halboob, Waleed & Al-Muhtadi, Jalal & Bouras, Abdelghani.: A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection on Social Media. IEEE Access. 1-1 (2022).

[30] Singh, S. P.: "Understand Stacked Generalization (blending) in depth with code demonstration." iq.opengenus.org. (accessed July 1, 2022).

[31] Socher, R. https://cs224d.stanford.edu/lecture_notes/LectureNotes4.pdf. Stanford Lecture Notes (2015).

[32] Tandoc, E., Lim, D. and Ling, R.: Diffusion of disinformation: How social media users respond to fake news and why. Journalism, 21(3), pp.381-398.(2019).

[33] Verma, P., Agrawal, P., Amorim, I. and Prodan, R.: WELFake: Word Embedding Over Linguistic Features for Fake News Detection. IEEE Transactions on Computational Social Systems, 8(4), pp.881-893.(2021).

[34] Vijayaraghavan, S., Wang, Y., Guo, Z., Voong, J., Xu, W., Nasseri, A., Cai, J., Li, L., Vuong, K., & Wadhwa, E.: Fake News Detection with Different Models (Version 1). arXiv.(2020).

[35] Yaraghi, N. How should social media platforms combat misinformation and hate speech? Brookings. (2022).