Fuzzy Systems and Soft Computing
ISSN : 1819-4362

FAKE NEWS DETECTION IN DISASTER MANAGEMENT USING MACHINE LEARNING TECHNIQUES

Vijila J, Department of Information Technology, University College of Engineering Nagercoil,Tamil Nadu,India. vijilaebenezer@gmail.com
Sahaya Vasanthi S, **Benila Grace B V,** Department of Civil Engineering,University College of Engineering Nagercoil,TamilNadu, India sahaya85@gmail.com

ABSTRACT
Fake news is false or misleading information presented as legitimate news, created deliberately to misinform or deceive readers. The rapid advancement in online communications and the embrace of social media platforms like WhatsApp and Facebook triggered ascend in the propagation of false news in recent years. Recently as the covid'19 related false information spreading faster and cause confusion that create many undesirable consequences. The fear of unknowingly consuming fake news has also created an environment of mistrust and doubt. As a result, a fake news detecting system is urgently needed. This study suggests an ensemble boosting method for machine learning that combines the AdaBoost boosting algorithm with Support Vector Machines (SVM). The model is created to potentially identify the fake news via system conditioning with four datasets ISOT, Kaggle, News Trends and Reuters and successfully validated by utilizing same four datasets. It outperforms existing traditional algorithms with heightened accuracy by 1.52%, recall by 4.02%, average precision-recall score by 2.22% and F1 score by 1.84% and precision value decreased by 0.19% on test data. Misinformation concerning casualties, infrastructure damage, or emergency response operations can lead to needless panic, resource misallocation, and delays in help distribution, making this problem especially crucial in disaster management.

1. INTRODUCTION
COVID-19 is a disaster due to its effects on everyday life, the economy, and public health. It causes high mortality, disrupts the economy, and overwhelms the healthcare system. Panic resulted from the quick dissemination of hoax news via online, news websites, and edited videos. False statements regarding government initiatives, illnesses, and remedies were examples of misinformation. Certain ethnic groups were immune, 5G distributes the virus, and vaccinations contain microchips, to name a few examples of bogus news. This false information led to social instability, vaccine reluctance, and dread. In order to stop additional harm, it is imperative to combat fake news.
With the rapid advances in electronic communication and the World Wide Web (WWW), a mass shift in how people consume news. With these changes, the circulation of false information has also increased exponentially. False or misleading information presented as a genuine news story is known as fake news. Almost all the fake content is simply created to distract people and set off mistrust among the readers by changing their mentality [1]. The mass dissemination of misinformation can have grave consequences, in many domains such as politics, health, science and economy. This is observed very prominently whenever elections take place and more recently when the coronavirus pandemic first started.World Health Organization (WHO) has announced Covid-19 to be an International concern public health emergency. Through global lockdown it is observed that there is an increase in 25% of users those who are engaging social media activity [2]. Web-based communities like Instagram, Twitter, Facebook and instant messaging applications such as WhatsApp have become a primary source of instantly knowing what is happening around the world. Twitter is the platform that spread fake news over 1.5M daily active users [3].
In the era of internet-based life, spread of fake news is faster [4] the stories often relate to topics that are trending on social media. The stories usually have an outrageous headline designed to click on it. Very often it is noted that the fake news has more views and engagement than actual news. Editors and journalists who curate the online news content are thus in dire necessity of novel mechanisms that

can assist them in speeding up the verification process for the questionable content found in social media. So, it is very essential to create an approach that can effectively notice false information.
Several machine learning-based ensemble techniques is available for noticing false news on social media networks such as Twitter and Facebook [5, 6]. Due to aggressive use of digital channels, false news has attained a great deal of attention. The study by Gundapu & Mamidi [7] (2021) consider an ensemble of three Pre-trained language models like BERT, ALBERT and XLNET to notice false news on the social media platforms and got more generalized model with a higher frequency. Rubin et al. [8] proposes a 5-feature classification model to identify fake news. Kesarwani et.al [9] (2021) develops two classifier models composed of the mixture of several machine learning methods and the accuracy of these algorithms was examined, and it was discovered that SVM outperformed KNN, Random Forest, and Logistic Regression in one dataset while the Logistic Regression algorithm outperformed the other algorithms in the second.
Yu et al. [10] develops an algorithm for detecting social media spammer using semi supervised learning algorithm. Ahmad et.al [11] (2020) extensively study different ensemble ways like as voting, utilizing a different types of machine learning strategies to bag and improve the classifier and evaluated their performance on four real world datasets. Rao & Seshashayee [12] (2020) explore a straightforward approach for identifying false stories using Scikit-Learn classifier. Nigam et al. [13] uses the combination of Naïve Bayes and Expectation Maximization (EM) algorithm for fake news classification. Vijjali et.al [14] (2020) develops a two-stage automated pipeline using transformer-based models for detection of fake news relating to the COVID’19 disease and for fact checking.
Ozbay & Alatas [15] (2020) combines the techniques of text analysis and supervised machine learning algorithms. An experimental valuation of the classification representations has been executed through publicly available datasets and the accuracy, precision and F-measure were compared in terms of mean value. Decision Tree, CVPS, ZeroR and WIHW algorithms showed best mean values according to the four metrics used. For extracting the features Kaur et.al [16] (2020) uses Hashing-Vectorizer (HV), Term Frequency - Inverse Document Frequency (TF-IDF) and Count-Vectorizer (CV). A new multi-level voting ensemble model is suggested and twelve machine learning (ML) models are reviewed to retrieve the best model after considering the benchmarking standards. Mahabub [17] (2020) evaluates the Ensemble Voting Classifier's (EVC) application and contrasts it with alternative classifiers. To detect bogus news, this method suggests using a perceptive detection system based on EVC.
Umer et.al [18] (2020) uses deep learning concept for false news stance detection and uses a hybrid neural network design using Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN). It also uses Principle Component Analysis (PCA) and Chi-Square to decrease the dimensional space of the feature embeddings prior to classification. The significant downside of this work is computationally more expensive to train data sets in deep learning. Kaliyar et.al [19] (2020) conduct various experiments with a tree-based Ensemble Machine Learning architecture (Gradient Boosting) using enhanced parameters joining both content and contextual features. A multi-class fake news dataset was used and Distinct ML algorithms were employed, such as Naïve Bayes, KNN and Decision Trees. The Gradient Boosting model achieved a high accuracy of 86%, demonstrating its accuracy for multi-class textual classification problems.
Most of the aforesaid methods are tested on single dataset and found that the existing method is achieved relatively low accuracy of individual algorithms and high processing time when deep learning or neural networks are used. To address the issues, by first introducing the boosting algorithm AdaBoost along with SVM to enhance the accuracy, F1 score, recall and precision. Similarly, this suggested method uses the compilation of four datasets such as ISOT, News Trends, Kaggle and Reuters for identifying the fake news. Hence, the fake news can be discovered efficiently and Obtained more precise results by 1.52%, recall by 4.02%, average precision-recall score by 2.22% and F1 score by 1.84% and precision value decreased by 0.19% on trial data while related to remaining methods.

2. PROPOSED WORK

In the proposed work, machine learning ensemble technique is explored for fake news classification. The performance of the SVM algorithm is enhanced by combining it with the boosting technique AdaBoost.

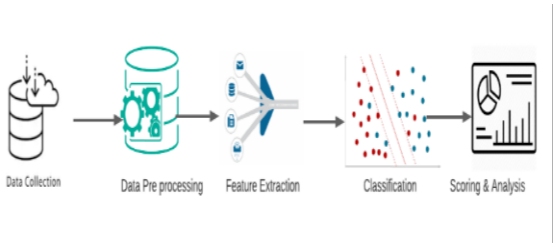

Figure 1 Architecture of proposed system

SVM has been chosen due to its reported high accuracy when considered with other classical ML algorithms for text classification. Figure.1 represents architecture of the proposed work. Data Collection, Data Pre-processing, Feature Extraction, Classification, and Scoring & Analysis are its five fundamental modules. Figure 2 displays the detailed architecture diagram for the suggested system.

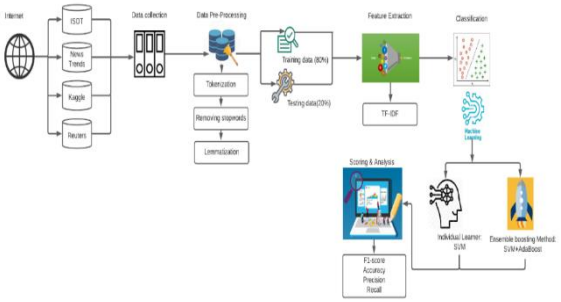

Figure 2 Detailed architecture of proposed system

During data collection, four different datasets such as ISOT [11], Kaggle [12], News Trends [13] and Reuters [14] are used which contains a labelled collection of real and false news stories. Data collection is conducted on these four different datasets to generate a unique dataset of 2000 records, it will be utilized for model testing and training. Pre-processing is done on the data for text normalization. During pre-processing punctuation is removed, text is converted to lowercase and tokenization is done to obtain an array of word tokens. Stop-word removal is carried out on the tokens to eliminate the very common words in English which are irrelevant as training features.

After pre-processing, the features are taken using TF-IDF vectorization. Here, a score is considered for both token based on its uniqueness, frequency as well as relevance to the particular news article text and to every record on whole dataset and a input matrix is obtained. The resulting vectorized dataset acts as coordinates for the testing and training phase. The data set allocated for testing and training has an 80:20 ratio.Initially SVM algorithm training is done. Support Vector Classifier is used with a “linear” kernel and the features in the vectorized training set are fitted into the model. On training, an optimal hyper plane is obtained which classify the news as “REAL” or “FAKE”. The vectorized testing set is then used to assess the F1 score, accuracy, recall, precision and other performance metrics of the algorithm.

Then the boosting algorithm AdaBoost is used with SVM as the base learner. The amount of estimators is set as 50 and the learning frequency is 1.0. Similarly, the “linear” kernel is applied for the Support Vector the model and classifier are trained by fitting the training vectors and finding the optimal hyper plane after multiple iterations of SVM occurs. In each iteration the weight of the misclassified features is increased and that of easily classified features is decreased. This is done by the AdaBoost classifier. Then the model is tested using the testing dataset to check the precision, accuracy, F1 score, recall and other performance metrics of the ensemble system.

Scoring is done for both models and performance of SVM used with boosting algorithm AdaBoost will be compared with that of SVM on its own based on metrics such as exactitude, accuracy, F1 tally, recall and average precision-recall score. Hence an calculation of the ensemble model will reveal how it fares against individual learners and whether or not this methodology acts effectively for distinguishing between phony and authentic news.

2.1 DATA COLLECTION

Social media sites like Facebook, Instagram, Twitter, and others are various of the main providers of fake news articles. In this work, the datasets such as ISOT, Kaggle, News Trends and Reuters which are accessible for downloading with columns such as title, text, author, subject and published date. The ISOT, News Trends, Kaggle and Reuters consists of 44,898, 44,676, 6,335 and 19,968 news articles respectively, labelled as either fake or real news.

Table 1 Statistics of collected datasets

Datasets	Total Articles	Real Articles	Fake Articles
ISOT	500	250	250
Kaggle	500	213	287
News Trend	500	255	245
Reuters	500	268	232

From each of these datasets, 500 news articles are manually collected and then concatenated to produce a unique dataset of about two thousand (2000) datasets. Table 1 shows the statistics of four collected datasets.

2.2 PRE-PROCESSING

In the pre-processing phase, the unstructured data will be converted into structured data. Data pre-processing is the crucial first step while creating any machine learning model. It is the system of taking the raw data, cleaning it and making it suitable for processing by a machine learning algorithm. Lemmatization, stop-word removal, and tokenization are the pre-processing phases.

2.2.1 Tokenization

Tokenization is the process in which textual data is divided into meaningful fragments or pieces. These pieces are often termed tokens and they remove every single punctuation from the text [14]**.** During this process, a distinctive ID of type integer is indexed to each word and then the frequency rate of each token is tallied and then normalization occurs [15]. An array of text tokens is obtained in this process.

2.2.2 Removing Stop words

Language-specific terms that are not especially important are called stop-words and do not carry any information when used on their own. Stop-words include conjunctions, prepositions, and pronouns used in the language. The English language has around 400-500 of such words. Some illustrations of stop lyrics in the English language are *a, an, and, am, but, does, on, once, until, too, when, where, what, any,* etc. [15]. During this step of pre-processing the English stop-words are eliminated during this phase and the resulting tokens don't contain these less relevant words.

2.2.3 Lemmatization

Lemmatization is the mechanism in which a word is converted to its root form. The words in context is observed and suitably it is changed to its basic form. To lemmatize, an instance of Word Net Lemmatizer () is created and the lemmatize() function is called on every single word token. Figure 3 depicts how lemmatization occurs on a word.

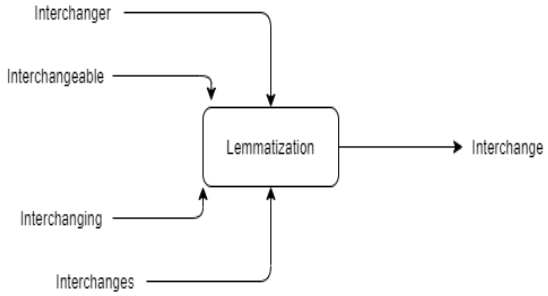

Figure 3 Lemmatization

2.2.4 Data Splitting

After pre-processing, the whole dataset is divided into 80:20 for testing and training the dataset using *train_test_split* function provided in Python.

2.3 FEATURE EXTRACTION

The dataset's characteristics are extracted using TF-IDF. This is a weighing matrix normally used for measuring the significance of a word. This is basically the sum of the count and weight. It is employed to evaluate and measure a word's relevance to an article within a group of articles. This is accomplished by multiplying the frequency of a word's occurrence in a news article (tf) by the number of cases it occurs in the dataset (idf). Equation (1) represents the TF-IDF score.

$$TF.IDF(w,d,D) = TF(w,d).IDF(w,D) \quad (1)$$

where, w → Word

d → Document

D → Document Set

2.3.1 Term Frequency (TF)

TF is measured by the number of occurrence in which a word occurs in a document or division of the article by the total count of words in the document. The TF is calculated using Equation 2.

$$TF(w,d) = \log\big(1 + freq(w,d)\big) \quad (2)$$

where, w → Word

d → Document

2.3.2 Inverse Data Frequency (IDF)

IDF is calculated by dividing the total number of documents or articles in the dataset N by the number of documents that contain the term w. Equation 3 is utilized to compute IDF.

$$IDF(w,D) = \log\left(\frac{N}{count(d \epsilon D : w \epsilon d)}\right) \quad (3)$$

where, w → Word

d → Document

D → Document Set

N→ Document Count

2.4. CLASSIFICATION MODULE

The refined dataset obtained after the pre-processing and extraction of features stages is then brought into the classification stage for detecting fake news articles. Support Vector Machine (SVM) is the machine learning model utilized here, in conjunction with AdaBoost, a boosting algorithm, and ensemble methodology is employed for classification.

2.4.1 Support Vector Machine (SVM)

SVM operates on the minimization principle and called as discriminative classifier [15]. It determines the most suitable decision boundary between vectors that belong to a group, class or category and vectors that do not belong to it. Vectors are a set of numerical values which represent a set of coordinates in some space.

The feature vectors obtained after TF-IDF feature extraction process are mapped in a two-dimensional space as a group of coordinates. The SVM algorithm then performs classification and identifies the right hyper-plane that segregates two classes (**fake or real**) very well by optimizing the space between the hyperplane and the support vector. The gap between the vectors and the hyper-plane is called as margin. SVM aims to maximize this margin distance. The ideal hyper-plane is the one with the largest margin between the support vectors. Figure 4 displays the vector charting and the connection between the ideal hyperplane and support vectors.

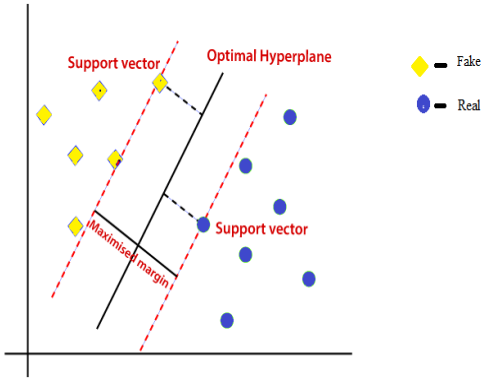

Figure 4 Graphical representation of SVM

During training the optimal hyper plane is found from an infinite number of options which classified the training feature vectors into “real” and “fake” categories and during testing time this hyper plane is used on testing feature vectors to evaluate how the result of SVM algorithm actually performs.

2.4.2 Adaptive Boosting (AdaBoost)

AdaBoost is a recognized boosting framework that works by combining many weak classifiers or base classifiers to evolve a single strong classifier. It is selected to enhance the efficiency of any algorithm. The Adaboost classifier first assigns a higher weightage to feature vectors which are tough to deal with and a lower weightage to the features that may be more effortlessly processed. This process happens

recurrently until the classifier is able to more accurately classify the learning data. Adaboost can be utilised alongside any classifier to improve it and produce a more accurate model

2.4.2.1 AdaBoost Algorithm steps:

The steps that are required to boost SVM using AdaBoost is as follows:

Step 1: Give each observation a same weight.

First, use Equation 4 to give each entry in the dataset the same weights.

$$Weight(xi) = \frac{1}{N} \quad (4)$$

where, N= The quantity of records

Step 2: Use stumps to classify random samples.

Fit the model by drawing replacement-based random samples from the initial data with probability equal to the sample weights.

Step 3: Determine the Total Error

The sum of the weights of the incorrectly classified records is the total error. There will always be a total inaccuracy between 0 and 1. Equation 5 is utilized in its computation.

$$Total\ Error = Weight\ of\ misclassified\ records \quad (5)$$

0 represents perfect stumps (correct classification) and 1 represents weak stump (misclassification).

Step 4: Calculating the implementation of stump (α)

The implementation of stump is calculated using Equation 6.

$$Performance\ of\ stump(\alpha) = \frac{{}^{1}/_{2}\ln\left[1-TE\right]}{TE} \quad (6)$$

where, ln → natural log

TE → Total Error.

Step 5: Update Weights

The weight for all misclassified records are updated using Equation 7.

$$New\ Weight = Weight * e^{(performance)} \quad (7)$$

Equation 8 is used to update the weight for all correctly classified records.

$$New\ Weight = Weight * e^{-(performance)} \quad (8)$$

Step 6: Weights are updated iteratively.

Step 7: Final forecasts

Equation (9) is to make the final forecast.

$$Final\ prediction/Sign(Weighted\ sum) = \sum(\alpha_i * (Predicted\ value\ at\ each\ iteration) \quad (9)$$

Every repetition involves these computations of the SVM algorithm when it is used with AdaBoost. The number of iterations can be specified in the 'n' estimators parameter when using the AdaBoost Classifier function provided by sklearn. When SVM with AdaBoost model is run, the specified number of iterations of the base learner occurs, and each time the misclassified records i.e., the vectors which were incorrectly classified by SVM during testing are assigned major priority than ones which were correctly classified.

2.5 SCORING AND ANALYSIS

Scoring and analysis is done to ascertain the efficacy of the suggested framework. The projected and actual values are displayed in Figure 4. Actual values can be either true or false and they are the real and expected outcome for any input given. Predicted values can be either positive or negative and they are the observed outcomes are the result of ML model. It is very helpful for calculating recall, precision, and accuracy and also crucial for generating the confusion matrix.

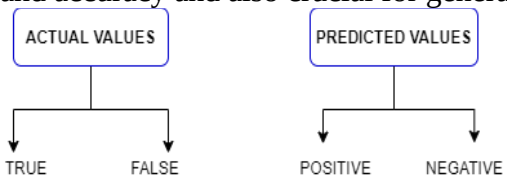

Figure 4 Actual and Predicted Values

The table which consists of four distinct groupings of predicted and actual values as represented in Figure 5. True Positive (TP)

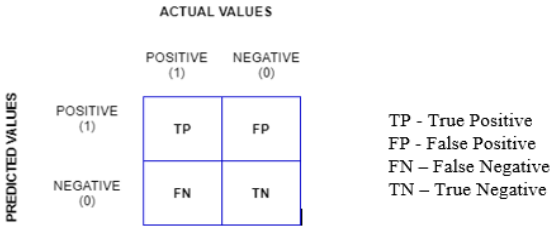

Figure 5 Different combinations in confusion matrix

Accuracy (A) is calculated to catch the percentage of accurate predictions or to know the classification made by the exemplary for the testing records. It is computed by dividing the total number of forecasts by the sum of the real positive and real negative outcomes. Equation 10 displays the accuracy calculation formula.

$$A = \frac{True\ positive + True\ negative}{True\ positive + False\ positive + True\ negative + False\ negative} \quad (10)$$

The model's positive predictive value is measured using precision (P). It can be written as the sum of real positive and false positive values divided by the ratio of real positive predictions. The formula for calculating precision is shown in Equation 11.

$$P = \frac{True\ positive}{True\ positive + False\ positive} \quad (11)$$

The model's sensitivity to relevance is known as recall (R). It is the sum of true positive and false negative values divided by the ratio of true positive values. Equation 12 is utilized in the computation of recall.

$$R = \frac{True\ positive}{True\ positive + FalseNegative} \quad (12)$$

F1-score tells us about model performance for positive and negative classes. It is used when the data set is not balanced. The score conveys the extensiveness of the proposed model. F1-score is computed as the harmonic average of precision and recall, which provides a balance between the two. F1-score is calculated using Equation 13.

$$F1 = 2 * \frac{precision.recall}{precision + recall} \quad (13)$$

3. RESULT AND ANALYSIS

3.1 DATA COLLECTION MODULE

The fake news detection datasets, News Trends, Kaggle, Reuters and ISOT are modified to contain 500 of the original records each to keep the size manageable. The Python Pandas package is used to import the CSV data files.The individual datasets are then linked into one unique dataset containing 2000 records total. Each news article is labelled as “REAL” or “FAKE”. The data is shuffled to reduce bias and unused columns are dropped. There are real and fake news articles. Figure 6 shows the first 5 rows of the dataset obtained after collection.

	title	text	label
0	Special election to replace Conyers to be held...	WASHINGTON (Reuters) - Michigan Governor Rick ...	REAL
1	Videos on the Pacific Crest Trail Association ...	Posted on October 30, 2016 by Graywolf Publish...	REAL
2	Hillary Clinton Tops "Islamist Money in Politi...	Hillary Clinton Tops "Islamist Money in Politi...	FAKE
3	UKIP MEPs Steven Woolfe & Mike Hookem reported...	UKIP MEPs Steven Woolfe & Mike Hookem reported...	FAKE
4	House tax panel chair to urge longer-lasting i...	WASHINGTON (Reuters) - The head of the U.S. Ho...	REAL

Figure 6 Sample Result obtained from Data Collection Module

3.2 DATA PREPROCESSING MODULE

Textual data needs to be refined and encoded to numerical values before feeding them into any ML model. The punctuation is removed and the text is modified to lowercase form. This is followed by tokenization, which creates an array of word tokens from the text sentences.

	title	text	label	removed_punc	tokens	stop_tokens	lem_words	clean_text	target
0	Special election to replace Conyers to be held...	WASHINGTON (Reuters) - Michigan Governor Rick ...	REAL	WASHINGTON Reuters Michigan Governor Rick Sny...	[washington, reuters, michigan, governor, rick...	[washington, reuters, michigan, governor, rick...	[washington, reuters, michigan, governor, rick...	washington reuters michigan governor rick snyd...	0
1	Videos on the Pacific Crest Trail Association ...	Posted on October 30, 2016 by Graywolf Publish...	REAL	Posted on October 30 2016 by Graywolf Publishe...	[posted, on, october, 30, 2016, by, graywolf, ...	[posted, october, 30, 2016, graywolf, publishe...	[posted, october, 30, 2016, graywolf, publishe...	posted october 30 2016 graywolf published oct ...	0
2	Hillary Clinton Tops "Islamist Money in Politi...	Hillary Clinton Tops "Islamist Money in Politi...	FAKE	Hillary Clinton Tops Islamist Money in Politic...	[hillary, clinton, tops, islamist, money, in, ...	[hillary, clinton, tops, islamist, money, poli...	[hillary, clinton, top, islamist, money, polit...	hillary clinton top islamist money politics li...	1
3	UKIP MEPs Steven Woolfe & Mike Hookem reported...	UKIP MEPs Steven Woolfe & Mike Hookem reported...	FAKE	UKIP MEPs Steven Woolfe Mike Hookem reported ...	[ukip, meps, steven, woolfe, mike, hookem, rep...	[ukip, meps, steven, woolfe, mike, hookem, rep...	[ukip, meps, steven, woolfe, mike, hookem, rep...	ukip meps steven woolfe mike hookem reported f...	1
4	House tax panel chair to urge longer-lasting i...	WASHINGTON (Reuters) - The head of the U.S. Ho...	REAL	WASHINGTON Reuters The head of the US House o...	[washington, reuters, the, head, of, the, us, ...	[washington, reuters, head, us, house, represe...	[washington, reuters, head, u, house, represen...	washington reuters head u house representative...	0

Figure 7 Result obtained from Data Pre-Processing Module

Lemmatization is performed on the tokens to get root words. Then stop-word removal is performed. The label "FAKE" is set as target "0" and "REAL" is set as "1". Next separate the dataset into sets for training and testing. 80% of the dataset is used for training, and 20% is used for model testing. Figure 7 shows a sample of the dataset after removal of punctuation, tokenization of text, stop-word removal, lemmatization of tokens, creation of clean text and conversion of the labels into target values '0' and '1'.

3.3 FEATURE EXTRACTION MODULE

An instance of TfidfVectorizer is first initialised. The TfidfVectorizer is used to perform the pre-processed text into TF-IDF features. Then fit and transform the vectorizer on the training dataset to obtain feature vectors in the form of matrix. The importance of each word in the text is calculated. Figure 8 shows the extracted features. The count of features is 40195 and the values of each feature act as vector coordinates during classification using SVM.

```
[[0.        0.        0.        ... 0.        0.        0.       ]
 [0.        0.        0.1164288 ... 0.        0.        0.       ]
 [0.        0.        0.        ... 0.        0.        0.       ]
 ...
 [0.        0.        0.        ... 0.        0.        0.       ]
 [0.        0.        0.        ... 0.        0.        0.       ]
 [0.        0.        0.        ... 0.        0.        0.       ]]
(1600, 40195)
[[0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]]
(400, 40195)
```

Figure 8 Result obtained from Feature Extraction module

3.4 CLASSIFICATION MODULE

3.4.1 SVM

Classification is done by initializing the SVC classifier with kernel set as "linear" and C=1.0. Then fit this on training set (training feature vectors) and y_train. Next, we supply the test set and use the accuracy_score() function to calculate the accuracy. Figure 9 tells the result obtained from classification using SVM.

```
SVM model fitted..
Accuracy: 0.817500
Confusion matrix :
 [[153  39]
 [ 34 174]]
Precision: 0.816901
Recall: 0.836538
F1 score: 0.826603
Average precision-recall score: 0.90
```

Figure 9 Result of Classification using SVM

The obtained results are accuracy comes with 0.817500, the precision is 0.816901, the recall score is 0.836538, F1 score is 0.826603, and the average precision-recall score is 0.90. The confusion matrix shows 153 true positive results, 174 true negative results, 39 false negative results and 34 false positive results.

A precision-recall curve (PR curve) curve is the mapping of precision in the y-axis and recall in the x-axis. The curve shows the trade-off involving the precision and recall values for a certain threshold value. If the area under the curve(AUC) is high then the algorithm is said to have great recall and precision values, in which case the high precision represents having a small false positive rate, and high recall represents a small false negative rate. This indicates that the classifier model is generally providing correct outcomes (high precision), and also giving mostly positive outcomes (high recall). Figure 10 shows the PR curve for the classifier SVM which shows that (AUC) area under the curve is high. This indicates a good recall and precision score.

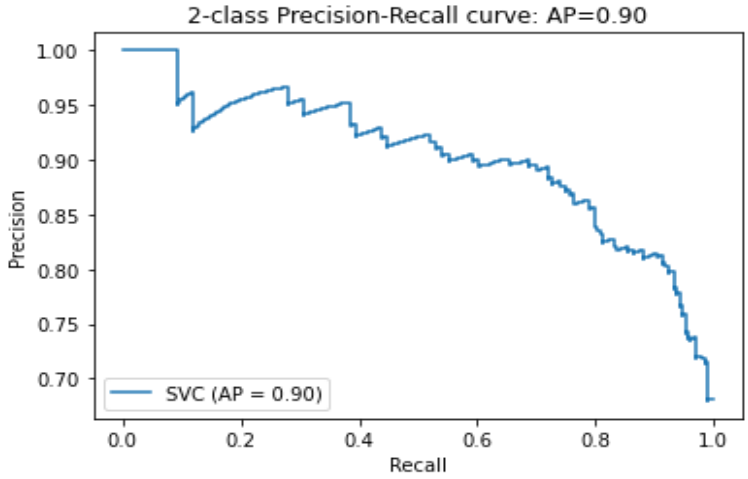

Figure 10 Precision-Recall curve using SVM

3.4.2 SVM with AdaBoost

AdaBoost is initialized using AdaBoostClassifier with base learner as SVC with kernel set to “linear”, the n_estimators = 50 and learning rate as 1.0. This classifier model is fit on training feature vectors and y_train. Next, we assess the model on the test set and compute the accuracy. Also evaluate the precision, recall values and produce a graph for analysis.

Figure 11 shows the result of classification using the ensemble boosting model, where AdaBoost is utilized with SVM as base learner. The result obtained using SVM with AdaBoost are accuracy comes out as 0.830000, the precision is 0.815315, the recall score is 0.870192, F1 score is 0.836186, and the average precision-recall score is 0.92. The confusion matrix shows 151 true positive results, 181 true negative results, 41 false negative results and 27 false positive results.

```
SVM+Adaboost model fitted..
Accuracy: 0.830000
Confusion matrix :
 [[151  41]
 [ 27 181]]
Precision: 0.815315
Recall: 0.870192
F1 score: 0.841860
Average precision-recall score: 0.92
```

Figure 11 Result of Classification using SVM and AdaBoost

Figure 12 shows the Precision-Recall curve for the classification algorithm SVM with AdaBoost which shows a high area under the curve.

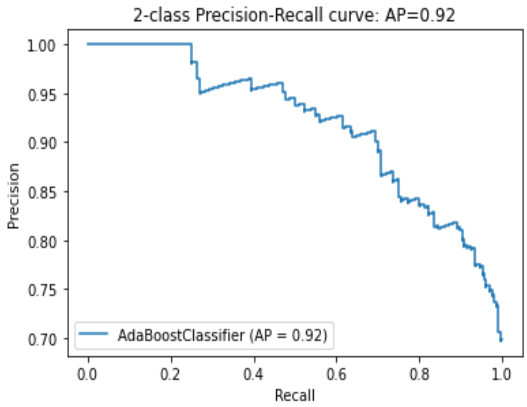

Figure 12 Precision-Recall curve using SVM with AdaBoost

The findings achieved from the two models are summarized in Table 2 which shows that when the boosting technique, AdaBoost is used with SVM algorithm an improvement is observed in the model's accuracy, recall, average precision-recall and F1 score.

Table 2 Analysis of performance metrics using SVM and SVM+AdaBoost

Performance Metrics	SVM	SVM + ADABOOST
Accuracy	0.817500	0.830000
Precision	0.816901	0.815315
Recall	0.836538	0.870192
Average Precision-Recall	0.90	0.92
F1 – Score	0.826603	0.841860

Figure 13 shows the analysis of performance metrics using SVM and SVM+AdaBoost. It clearly shows that the accuracy increases by 1.52%, precision decreases by 0.19%, recall increases by 4.02%, average precision-recall score increases by 2.22% and F1 score increases by 1.84%.

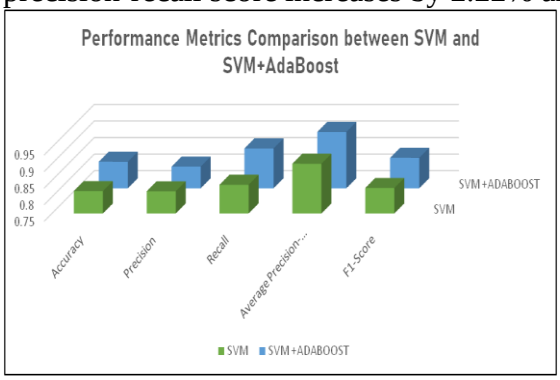

Fig.13 Graphical representation of comparison of performance metrics using SVM and SVM + AdaBoost

4.CONCLUSION

Presently, it has become overly complex for readers or even regular social media users to acquire truthful and dependable information because of enormous amounts of false news in circulation. In this analysis, a system is proposed to discover fake news by using ensemble classification method and a supervised ML algorithm. The algorithm used here is SVM algorithm. Along with SVM an AdaBoost is used as a boosting algorithm. This algorithm is shown to enhance the performance of SVM in metrics such as accuracy, recall, average precision-recall and F1 score. Precision had a slight decrease in value compared to SVM which has no discernible effect on the model's accuracy in classifying the relevant

outcomes. However, since the F1 score of the model is greater this illustrates that the trade-off between precision and recall is better for SVM with AdaBoost. The suggested system exhibits encouraging outcomes for a model that approaches fake news identification broadly. In future, the current system may be enhanced and made better by looking into different base learners with ensemble algorithm for better results. Evolutionary weight optimization techniques and existing feature extraction approach may also be explored for improving the performance. Models may be created to run on real-time datasets so fake news can be detected at earlier stages. By preventing misleading information that leads widespread panic, improper resource allocation, and delays in emergency response, this system is an essential tool in disaster management. The suggested model can help authorities make well-informed decisions, improve public trust, and enable effective disaster relief operations by guaranteeing that only verified information is shared during crises. This will ultimately lessen the outcome of manipulated information on infrastructure and human lives.

REFERENCES

1. **Gundapu, S., & Mamidi, R. (2021)**. Transformer based automatic COVID-19 fake news detection system. *arXiv preprint arXiv:2101.00180*. https://doi.org/10.48550/arXiv.2101.00180.
2. **Kesarwani, A., Chauhan, S. S., Nair, A. R., & Verma, G. (2021)**. Supervised machine learning algorithms for fake news detection. In *Advances in Communication and Computational Technology: Select Proceedings of ICACCT 2019* (pp. 767-778). Springer Singapore. DOI:10.1007/978-981-15-5341-7_58
3. **Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020).** Fake news detection using machine learning ensemble methods. *Complexity*, *2020*(1), 8885861. **https://doi.org/10.1155/2020/8885861**
4. **Tuteja, A., Verma, A., & Badholia, A. (2020).** Investigating Fake News Detection Using Machine Learning. *Solid State Technology*, *63*(4), 1410-1421.
5. **Vijjali, R., Potluri, P., Kumar, S., & Teki, S. (2020).** Two stage transformer model for COVID-19 fake news detection and fact checking. *arXiv preprint arXiv:2011.13253*. https://doi.org/10.48550/arXiv.2011.13253
6. **Ozbay, F. A., & Alatas, B. (2020).** Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: statistical mechanics and its applications*, *540*, 123174. DOI: 10.1016/j.physa.2019.123174
7. **Kaur, S., Kumar, P., & Kumaraguru, P. (2020).** Automating fake news detection system using multi-level voting model. *Soft Computing*, *24*(12), 9049-9069. https://doi.org/10.1007/s00500-019-04436-y
8. **Mahabub, A. (2020).** A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers. *SN Applied Sciences*, *2*(4), 525. https://doi.org/10.1007/s42452-020-2326-y
9. **Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G. S., & On, B. W. (2020).** Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access*, *8*, 156695-156706.
10. **Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G. S., & On, B. W. (2020).** Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access*, *8*, 156695-156706.
11. **Prakasha, S., Firdosh Parveen, S., & Suma, G. C.** Disaster Management with Intelligent Education Utilizing Deep Learning and Social Media to Achieve Environmental Sustainability While Countering False News.
12. **Zair, B., Abdelmalek, B., & Mourad, A. (2022).** Smart Education with Deep Learning and Social Media for Disaster Management in the pursuit of Environmental Sustainability While avoiding Fake News.
13. **Azim, S. S., Roy, A., Aich, A., & Dey, D. (2020).** Fake news in the time of environmental disaster: Preparing framework for COVID-19.
14. **Singh, D. K., Shams, S., Kim, J., Park, S. J., & Yang, S. (2020, January).** Fighting for Information Credibility: AnEnd-to-End Framework to Identify FakeNews during Natural Disasters. In *ISCRAM* (pp. 90-99).

15. **Domala, J., Dogra, M., Masrani, V., Fernandes, D., D'souza, K., Fernandes, D., & Carvalho, T. (2020, July).** Automated identification of disaster news for crisis management using machine learning and natural language processing. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 503-508). IEEE.
16. https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php
17. https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset
18. https://bit.ly/2zVRLxK
19. https://bit.ly/2BmqBQE