Detection of Fake News using Machine Learning

Sakshi Neeraj
Dept. of CSE
Amity University
Noida, India
sakshi1neeraj@gmail.com

Leena Singh
Dept. of CSE
Amity University
Noida, India
leenasafya@gmail.com

Sudhanshu Tripathi
Dept. of IT and Engineering
Amity University
Tashkent, Uzbekistan
stripathi1@amity.edu

Nidhi Malik
Dept.of CSE
The NorthCap University
Gurgaon, India.
nidhimalik@ncuindia.edu

***Abstract*—Regulation of false news causes challenges in the dissemination of accurate information and leads to misconceptions that endanger national cohesion and individual tranquility. Because information fabricates how the folk encapsulates the world, it is prime to combat this unauthentic news. People establish their own ideas in addition to basing critical judgments on these news articles, thus inaccurate news can exert a disastrous impact on the culture of a society making segmentation of publication pieces as spurious or authentic highly important. Numerous academicians are endeavoring to spot spurious news, and Machine Learning has been shown to be fruitful. In this paper, varied Machine Learning Algorithms are utilized to generate models to classify a particular news piece as authentic or fake. Python was applied as the scripting language throughout this development. Individual Machine Learning Algorithms such as K Nearest Neighbors and Decision Trees along with in-built ensembled classifiers (Random Forest, Gradient Boosting) and custom ensembled models (Stacking, Maximum Voting Classifier) are pattered for the purpose. With the avail of the appropriate models and tools, the challenging work of detecting false news may be made simple. This paper has been able to achieve an accuracy of 91.5% in classifying news as true or fake by stacking three individual Machine Learning Models namely, K Nearest Neighbors, Support Vector Classifier and Logistic Regression into a custom-ensembled model.**

Keywords—Natural Language Processing (NLP), Natural Language Toolkit (NLTK), Term Frequency-Inverse Document Frequency (tf-idf) Vectorizer, In-built and Custom ensembled Machine Learning (ML) Models, Support Vector Classifier (SVC), Logistic Regression (LR), K- Nearest Neighbors (KNN)

I. INTRODUCTION

Inaccurate or deceptive information that is reported as news is termed fake news. Central objective of fake news is to tarnish someone's reputation or to benefit through advertisements. The label "fake news" was first used in the 1890s, a time when exaggerated and flashy newspaper stories were ubiquitous. Misinformation has always been disseminated throughout history. Contemporary research has evidenced that false news is proliferating at an astounding level, leading to its widespread popularization. The propagation of anti-vaccination content during the Covid 19 catastrophe is a glaring example of this. Internet is one of the most vital breakthroughs, and plenty of people have access to a variety of social networking channels. Each user possesses the liberty to disseminate material or communicate a narrative via the web tenets. The content on these venues does not behold any authenticity. Therefore, some individuals attempt to distribute false information via these channels. These stories may be disinformation about a person, group, political party, or society.

A person cannot discern all of these false reports on his own. Therefore, there is a need for ML classifiers that can automatically identify these phony news[1]. By suggesting a method that can accurately categorize phony news, this paper seeks to make the fake news spotting process efficient. Without being overtly taught to do so, ML algorithms create a model using sample data, also referred to as training data, in order to make predictions or judgments[2]. This domain has proved to be useful in areas such as biomedical informatics, from drug development to epidemiological modelling and disease detection from medical imagery[3,4]. The use of ML models like K Means Clustering and KNN in real-world situations like movie recommender systems is also extensive in the contemporary era[5]. Deep Learning Convolution Model, a subset of ML has also put forth numerous sentiment analysis plans helpful in recognizing the true crux of a text piece[6]. The greater accessibility of datasets to the research community and improvements in architectures for modelling scientific events have improved the application of ML in many sectors and one such is fake news spotting.

The primary idea behind this paper is to gather real and fake news articles over time, provide them to ML algorithms for the creation of classification models, and then use these models to determine whether a news story is real or phony. It is crucial to carry out the data pretreatment step of NLP before supplying the dataset to any ML model to make the text easier for ML to understand. NLP is basically the evaluation and synthesis of biological language and speech using evolutionary computation in a machine[7]. In this paper, NLP is performed as follows: special characters, hyperlinks and stop words are removed from the text through a process called text cleaning, sentences are tokenized to words and these words are further lemmatized. Additionally, training and testing portions of the entire dataset are separated. The next step entails text vectorization, which involves converting a sentence's words to numbers so that an ML system can use them, tf-idf vectorization model is used to achieve this goal. Further, the preprocessed dataset is used to produce various ML models for identifying fake news. In general, methods to classify a news piece as true or fake have been suggested using either individual classification ML models or in-built ensembled models, however, in this paper, custom ensembled models have also been constructed. A custom ensembled model has been constructed by stacking KNN, SVC and LR, another custom ensembled model named Maximum Voting Classifier has also been constructed by appending these three individual models.

II. LITERATURE SURVEY

The propagation of false information across all online platforms has been facilitated by the development of the world wide web. In terms of the current content-based analysis of conventional approaches, identifying fake news is regarded as one of the difficult challenges[8].

Previously done research elucidated many automated tactics for spotting bogus uploads and counterfeit articles. There are several dimension false news detection features ranging from accessibility of chatbots to distribute false information or clickbait to facilitate the proliferation of misconceptions. Apparently multiple social media networks such as Facebook, encourage post sharing which disburses erroneous news pieces. There has been a lot of work done already to identify fake information. Novelists in [9] have outlined a multitude of detecting methods. As diagnostic approaches, the authors presented linguistic semantics, clustering, and forecasting. The accuracy achieved by these lied between 63 to 70% only. The authors of [10] have divided each post as a Boolean segmentation activity. The author conducted DMOZ and Twitter API to manually acquire data sets. On the data sets, the following methods were employed: Random Forest, XG Boost, Naive Bayes, Decision Trees, SVM, and neural networks. The writers of [11] observed the guidelines, procedures, and algorithms used for categorizing fake and manufactured news articles. The authors of the publication discuss the FakeDetector, automated fake news inference model. It uses textual categorization as a foundation and erects a model to concurrently segment phony news pieces from the true ones. This model was made up of the two primary components of FakeDetector: portrayal feature acquisition and credibility tag inference. The editors of [12] have highlighted click baiting as a manifestation of tabloidization. They have characterized click baiting as a tactic of disseminating misinformation quickly online. The authors have addressed efficient tactics for automatically identifying bait thumbnails as a deceptive ploy. The authors utilized content clues, including lexical and semantic level analyses. As a result of the recent deployment of deep learning research and applications, numerous research projects have used Deep Learning techniques, such as convolutional neural networks for phony news spotting[13].

In an editorial, Facebook was cited as performing on two aspects to combat the transmission of erroneous information. The first is sabotaging economic impulses because the preponderance of fake news is driven by money. Second erects the implementation of products to barricade the proliferation of erroneous information [14,15]. Here are some of the precautions that Facebook has taken: Ranking Advances: The frequency of misleading news information is decreased by News Feed rankings. Determine what is valuable and what is not for reporting purposes. Their community deems fake news to prevent its professing. Although they are still in the alpha stage and have not yet been made available to beta users, WhatsApp introduced security features and bogus news demarcations to combat the dissemination of phony information. Suspicious Link Detection to warn users is a function that WhatsApp performs by labelling links that it knows will take users to a fraudulent or alternate website or source of news with a red asterisk. A message may also be prohibited if it is routed on from a single device for 25 times[16].

In 2020, a study reported that while ensemble learners' accuracy was 81.5%, individual learners, such Linear Support Vector Machine(SVM), had an accuracy of 47.75%. When the dataset was changed, a pattern emerged, with individual learners' accuracy falling to 80% and ensemble learners' accuracy rising to 93.5%. Individual algorithms often obtained an accuracy of 85%, while ensemble students typically reached an accuracy of 88.16%. The Wang-Bi-LSTM algorithm, with an accuracy of 62%,had the lowest results[17].

In 2017 Tim Berners-Lee, who established the World Wide Web, himself asserted that fake news constituted one of the most significant and worrisome new internet concerns that need to be addressed[18,19]. Over the past few years, fake news has proliferated more and more, for instance 100 bogus articles had been hosted around the 2016 US presidential election alone[20]. This paper attempts to utilize various Machine Learning algorithms to produce classification models to categorize news as true or fake.

III. DESIGN AND IMPLEMENTATION

This paper revolves around a seven-step methodology as indicated in Fig.1

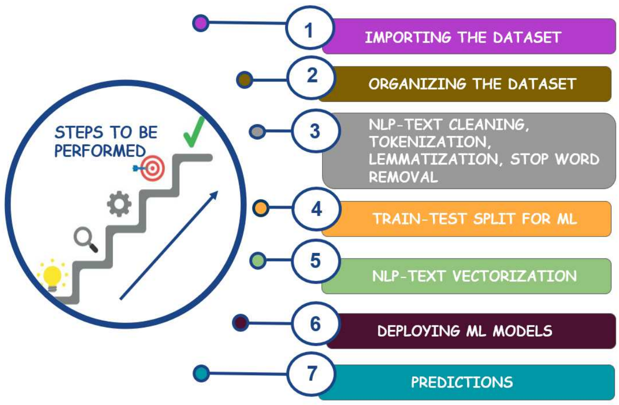

Fig. 1. Methodology of the proposed model

Two distinct datasets, 1 for true news and 1 for fake news pertaining to the years 2016-17 are used. Each dataset constitutes 4 columns – Subject of news, Date of publishing, Title, and the text of the news. Fig.2 is a snapshot of the true news dataset imported.

	title	text	subject	date
21407	Mata Pires, owner of embattled Brazil builder ...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017
21408	U.S., North Korea clash at U.N. forum over nuc...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017
21409	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017
21410	Headless torso could belong to submarine journ...	COPENHAGEN (Reuters) - Danish police said on T...	worldnews	August 22, 2017
21411	North Korea shipments to Syria chemical arms a...	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews	August 21, 2017
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017

Fig. 2. True news dataset

Next is the stage of data organization entailing the systematic orientation of the imported datasets into an easily comprehendible format. In this research, a new column pertaining to the nature of news, value 1 for true news and 0 for fake news is generated and the two datasets are further merged after removing the unimportant columns like subject and date of publication. Fig.3 shows the final dataset.

	title	text	class
10231	AUDIT: Obama's IRS 'Misled' Americans to Get T...	Soooo the IRS lied to Americans to prod them...	0
13470	Kremlin: U.S. sanctions aimed at turning busin...	MOSCOW (Reuters) - The Kremlin said on Thursda...	1
22875	SYRIA: British and American Presence Directly ...	US paratrooper on security duty during a miss...	0
2240	Watch NBC's Andrea Mitchell Get BULLIED Out O...	If one thing has become abundantly clear, it s...	0
17190	FAMILY THREATENED AT GUNPOINT FOR DISPLAYING C...	Nothing says tolerance like putting a loaded g...	0

Fig. 3. Organized dataset

The assessment and synthesis of organic language and speech using evolutionary computation in a machine is what is NLP. Here, the column "title" of the news is taken as the independent variable for the construction of the classification model, thus, it is imperative to execute these data pretreatment step of NLP on the title of the news so that the classification ML model is able to get trained on the basis of the "title" of the news:

A. *Data cleaning:* Text cleaning is performed in which any special characters or hyperlinks are replaced with white spaces.

B. *Tokenization*: A word based tokenization is performed which converts a sentence into its constituent words separated by commas.

C. *Lemmatization:* The aim of lemmatization is to convert a group of words to a similar base form so that the meaning of the word can be inferred easily. For instance, the word "going" is changed to "go" and also "gone" is changed to "go" to signify that the intended meaning of both the words is relatively in the same context.

D. *Stop word removal*: Words which do not imply any significant meaning to the text like "a" are removed so that more focus can be put forward to the information pieces which are more important for model training.

Further, the whole dataset is split into training and testing sections, here "80% of the data is taken for training purpose and the rest 20% is used for testing the accuracy of the model".

Next comes the step of text vectorization which is a step in which a sentence's words are changed to numbers so that they may be supplied to an ML algorithm. In this paper, tf-idf vectorization model is employed for this purpose. The framework of tf-idf vectorization entails computing the tf-idf values for each word in the fed strings of data and then converting that data into a vector of numerical value equivalent to this tf-idf score. Fig.4 is the mathematical equation for the computation of tf-idf scores.

$$tf-idf\ score = \left(\frac{count\ of\ a\ word\ in\ the\ sentence}{number\ of\ words\ in\ the\ sentence}\right) * \log\left(\frac{number\ of\ sentences}{number\ of\ sentences\ containing\ that\ word}\right)$$

Fig.4. tf-idf score

The preprocessed dataset is then used to produce various ML models for classifying the news as real or fake and the independent variable for classification is taken to be the "title" of the news. Supervised learning algorithms used for model generation are as follows:

A. *K-Nearest Neighbors:* KNN is a non-parametric, classifier that employs vicinity to anticipate the category of a particular data point. It is stated that a point lies in the category in which majority of its neighboring points on Cartesian plane lie. A 3-Fold hyperparametric tuning was conducted during which it was observed that the most optimum value of K for model generation was "25".

B. *Logistic Regression:* This algorithm uses the sigmoid function to calculate the likelihood of an occurrence, and if the result is less than a threshold number, the category is said to be false; otherwise, true.

C. *Decision Tree:* These are a sort of tree frameworks which mimic flowcharts. Intermediate nodes elucidate tests on attributes and each branch is the result and the terminal nodes are the class labels.

D. *Support Vector Classifier:* This classifier disbursements data points to a high-dimensional feature mapping, aiding classification of input points not linearly detachable. A hypothetic hyperplane in the Cartesian frame divides the data points into their respective categories. This tactic is grounded on the idea that the probability of two points lying in the same category is proportional to their vicinity on the cartesian frame.

E. *AdaBoost Classifier*: It is a recursive estimator that commences by linking a classifier on the initial corpus, then fits replicas of the classifier on the same fed set while altering the weights of incidences of misclassification in to boost the efficacy of subsequent classifiers. Decision Tree is the default base estimator from which the boosted ensemble is constructed.

F. *Gradient Boosting Classifier*: This approach constructs an iterative model in a sequential phase manner; it permits the optimization of any differentiable loss function, allowing the results at each stage to improve successively, leading to a better accuracy at the end.

G. *Random Forest Classifier*: This algorithm is generated by consolidating a collection of decision trees, and each tree is fed with data sample extracted from a training set called the bootstrap sample. The output given by the majority of the trees constituting the forest is taken as the final output of the model.

H. *Stacking Classifier*: This custom ensemble learning technique combines various categorization models that the user has chosen to produce a single "super" model. Here, K-NN and SVC are employed as base models, and the predictions produced by these base models are then used by a meta model, in this case, LR to make the final prediction. Fig.5 gives a pictorial representation of the stacking classifier.

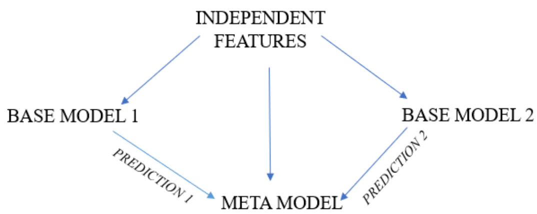

Fig.5. Stacking Classifier

I. *Maximum Voting Classifier*: Three models, K-NN, SVC and LR are appended to generate this custom ensembled model. All the three models make their predictions and the output of the majority amongst these is deemed to be final result.

IV. Review and Results

The main parameters of assessment of a model's functioning elucidated in this section are accuracy, bias, and variance of the model. Along with these, a K-fold cross validation score is also assessed using 10 number of folds.

Amongst the deployed models for classification, the highest accuracy was proposed by the custom ensembled model generated by stacking KNN, SVC and LR. Accuracy professed by the model accounted for 91.50%, bias was 0.10%, variance was 8.49% and cross validation score with 10 folds amounted to be equal to 92%. A detailed analysis of the scores proposed by all the models is given in Fig.6.

Model	Accuracy	Cross Validation Score	Bias	Variance
K Nearest Neighbors	87.30%	87.26%	11.35%	12.70%
Logistic Regression	91.20%	91.52%	3.27%	8.79%
Decision Tree	80%	84.72%	0%	19.99%
Support Vector Classifier	91.20%	91.52%	3.27%	8.79%
Ada Boost Classifier	78.10%	81.06%	16.82%	21.89%
Gradient Boosting Classifier	80.20%	82.53%	14.20%	19.79%
Random Forest Classifier	87.70%	90.02%	0%	12.30%
Stacking Classifier	91.50%	92%	0.10%	8.49%
Maximum Voting Classifier	91.40%	91.80%	1.40%	8.59%

Fig.6. Scores of each model

A confusion matrix for the stacking classifier constructed on the actual testing data and the predicted values of testing data shows that out of a total of 1000 test samples of news articles available out of which 457 were true and 543 were false, "416 were accurately predicted to be true (True Positives), 499 were accurately predicted to be false (True Negatives), 41 were false but were predicted to be true (False Positives) and 44 were true but were predicted to be false (False Negatives)". Fig.7 is a pictorial representation of the confusion matrix.

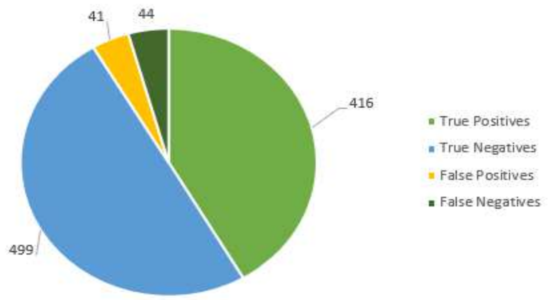

Fig.7. Confusion matrix pictorial representation

Out of 1000 entries, 915 were accurately predicted which accounted for 91.5% while 85 were miss-classified.

A classification report for the stacking model was built on a support of testing data which had 543 instances of fake news and 457 instances of true news. Fig.8 shows a graphical analysis of this classification report.

	News	Precision	Recall	f1-score	Support
	Fake News	92%	92%	92%	543
	True News	90%	91%	91%	457
Macro Average		91	91	91	1000
Weighted Average		92	92	92	1000

Fig.8. Classification Report

Whenever a model predicts a news piece as true, it is 90% correct and whenever it predicts it to be false, it is 92% correct. The model detects 91% of true news and 92% of false news.

V. Conclusion

This paper was designed to manifest and isolate original news from counterfeit news pieces so that fake news is not broadcasted on the internet and is not circulated among the folks. Transmission of inaccurate news pieces causes serious impact on the cultural well-being of a society and has been a core subject of concern for quite a while. This paper aimed to shield society from the circulation of inaccurate information by making use of varied ML frameworks, which incorporate diverse algorithms for categorization of news articles as true or fake. Steps for NLP like text cleaning, tokenization, lemmatization, and vectorization, have been used for data pre-processing preparatory to building the classifier. A custom-ensembling ML approach created by stacking three distinct algorithms—KNN, LR, and SVC, showed an accuracy of 91.5%, a bias of 0.1%, variance of 8.49% and a cross validation score with 10 folds amounting to 92%.

References

[1] A. Ahmed, A. Aljabouh, P.K. Donepudi, and M.S. Choi, "Detecting Fake News Using Machine Learning : A Systematic Literature Review", arXiv: 2102.04458, February 2021.

[2] J.H. Friedman, "Data Mining and Statistics: What's the connection?, Computing Science and Statistics", vol. 29(1), pp. 3–9, 1998.

[3] I. Baltruschat, H. Nickisch, M. Grass, T. Knopp, A. Saalbach, "Comparison of Deep Learning Approaches for Multi-Label Chest X-ray Classification", vol. 9, Nature Publishing Group, Berlin, Germany, 2019.

[4] R. Siddiqi, "Automated pneumonia diagnosis using a customized sequential convolutional neural network", 3rd International Conference on Deep Learning Technologies, Xiamen, China, pp. 64–70, July 2019.

[5] N. Kaur and A. Solanki, "Sentiment Knowledge Discovery in Twitter Using CoreNLP Library", 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2018.

[6] N. Kaur and A. Solanki, "Sentiment Knowledge Discovery in Twitter Using CoreNLP Library", 8th International Conference on Cloud Computing, Data Science & Engineering, 2018.

[7] G. Guida, G. Mauri, "Evaluation of natural language processing systems: Issues and approaches", Proceedings of the IEEE, vol. 74 (7), pp. 1026– 1035, July 1986.

[8] M.C. Buzea, S. Trausan-Matu, T. Rebedea, "Automatic Fake News Detection for Romanian Online News", 2022.

[9] S.B. Parikh and P.K. Atrey, "Media-Rich Fake News Detection: A Survey", IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) pp. 436-441, April 2018.

[10] S. Helmstetter, and H. Paulheim, "Weakly supervised learning for fake news detection on Twitter", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) pp. 274-277, August 2018.

[11] J. Zhang, L. Cui, Y. Fu, and F.B. Gouza, "Fake news detection with deep diffusive network model", 2018.

[12] Y. Chen, N.J. Conroy, and V.L. Rubin, "Misleading online content: Recognizing clickbait as false news". Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection pp. 15-19, November 2015.

[13] G. Hinton, S. Osindero, and Y. The, "A fast learning algorithm for deep belief nets. Neural Computation", 2006.

[14] Goldsborough, Reid, "Understanding Facebook's News Feed". Tech Directions, Vol. 76 (9), May 2017.

[15] M. Garrahan and T. Bradshaw, R. Waters, "Harsh truths about fake news for Facebook, Google and Twitter", Financial Times, January 2017.

[16] N. Kumar, S. Sharma, “Survey Analysis on the usage and Impact of Whatsapp Messenger”, Global Journal of Enterprise Information System.

[17] A. Ahmad, Y. Muhammad, Y. Suhail, A. Muhammad Ovais, "Fake News Detection Using Machine Learning Ensemble Methods", 2020.

[18] J. Swartz, "The World Wide Web's inventor warns it's in peril on 28th anniversary”, *USA Today*, March 2017.

[19] M. Bleigh, "The Once And Future Web Platform", TechCrunch, May 2014.

[20] H. Allcott, M. Gentzkow, "Social media and fake news in the 2016 Election", Journal of Economic Perspectives, vol. 31(2), pp. 211–236, May 2017.