

Fake News Detection in Machine Learning Hybrid Model



P. Chandana, K. Sree Vijaya Lakshmi.

Abstract: Now a day's prediction of fake news is somewhat an important aspect. The spreading of fake news mainly misleads the people and some false news that led to the absence of truth and stirs up the public opinion. It might influence some people in the society which leads to a loss in all directions like financial, psychological and also political issues, affecting voting decisions during elections etc. Our research work is to find reliable and accurate model that categorize a given news in dataset as fake or real. The existing techniques involved in are from a deep learning perspective by Recurrent Neural Network (RNN) technique models Vanilla, Gated Recurrent Unit (GRU) and Long Short-Term Memories (LSTMs) by applying on LAIR dataset. So we come up with a different plan to increase the accuracy by hybridizing Decision Tree and Random Forest.

Keywords: Deep Learning, Decision Tree, Random Forest.

I. INTRODUCTION

The fake news of these days causes numerous problems, from satirical posts to false news and government misinformation in some media. Counterfeit news and nonappearance of trust in the media are creating issues with gigantic outcomes in our overall population. Clearly, an intentionally misleading story is "fake news" yet of late gabbling the discussion of online life is changing its definition. Some other directly use the term to pardon the real factors counter to their supported points of view. The criticalness of disinformation inside American political talk was the subject of significant thought, particularly following the American president political race. The term 'fake news' became standard discourse for the issue, particularly to depict honestly mistaken and misdirecting articles circulated generally to get money through online visits. In this paper, it is seemed to make a model that can exactly foresee the likelihood that a given article is phony news. Facebook was the subject of much assessment despite media thought. They have quite recently executed a component to flag fake news on the site when a customer see's it; they have moreover said uninhibitedly they are managing to isolate these articles in a motorized way. Obviously, it's definitely not a straightforward task. A given computation must be politically impartial – since counterfeit news exists on the two pieces of

the deals – and besides give equal leveling to genuine news sources on either end of the range. Also, the subject of validness is an irksome one. Regardless, in order to handle this issue, it is essential to have an understanding on what Fake News is. A while later, it is required to examine how the strategies in the fields of AI, regular language taking care of help us to perceive counterfeit news.

II. RELATED WORK

Sherry et.al, [1], suggested deep learning algorithms for false news in online text. In this, they used a method called RNN model (vanilla, GRU) and LSTM techniques. The advantage of this model is the GRU is best among all the results and the disadvantages is the technique vanilla causes failure to solve complex tasks like changing original information format which meant that the essential memory material could not be retained. Monther Aldwairi et.al, [2], had suggested finding counterfeit news in online life systems. For this they use clear and carefully chosen title and post features to correctly distinguish false posts and the target benefits from clickbait's. This show breaks down the commonness of phony news given the advances made conceivable by the rise of person to person communication locales in the correspondence. The point of the work is to make an answer that clients can use to recognize and sift through destinations that contain bogus and deluding data. Automatic Online Fake News detection combining content was proposed by Macro L.DellaVedova et.al. [3]. In this paper, they at first propose a novel ML fake news recognizable proof methodology which, by joining news substance and social setting features, beats existing strategies in the composition, growing their viably high precision. Mykhailo Granik et.al, proposed a strategy for distinguishing counterfeit news, called the naive bayes classifier [4]. This paper shows a basic methodology for counterfeit news discovery utilizing guileless Bayes classifier. This methodology was actualized as a product framework and tried against an informational index of Facebook news posts. They accomplished arrangement exactness of around 74% on the test set which is a not too bad outcome thinking about the overall straight forwardness of the model. This outcomes might be improved in a few different ways that are portrayed in the article too. Akshay Jain et.al, introduced a title called Fake News Detection [5]. They used naive bayes classification algorithm. In this paper clarifies, each news story named as "Genuine" and "Counterfeit" were labeled as 1 and 0 separately in order to apply Naïve Bayes classifier. Right off the bat they create our jargon by utilizing the pack of-words idea. Next, when they change our preparation information, fit our model, make forecasts on the change test information, and process the AUC score for the title.

Manuscript received on April 02, 2020.

Revised Manuscript received on April 21, 2020.

Manuscript published on May 30, 2020.

* Correspondence Author

P.Chandana*, CSE, VR Siddhartha Engineering College, Vijayawada, India. Email: popuri.chandana97@gmail.com

K.Sree Vijaya Lakshmi, CSE, VR Siddhartha Engineering College, Vijayawada, India. Email: vijaya@vrsiddhartha.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Fake News Detection in Machine Learning Hybrid Model

Feng Wang et.al, presented the title A First Step to Fight Fake News over Online Social Media [6]. They utilized assistance calculation for the order of vector machines in this. LDA, a probabilistic topic model, to get the subjects of fake, certified and hybrid news corpus independently. The goal of LDA point showing on fake and real news is to grasp the difference or similarity of subjects between named fake and certified news.

Kai Shu et.al, suggested a title called Fake News Tracker: a resource for gathering fake news, identifying, visualizing [7]. In this article, they collected the information from the phony sites and kept an eye on Twitter for any social responsibilities applicable to the news things.

Sherif Shad et.al, [8] proposed a strategy titled As Online False News Identification, using N-Gram Research and Machine Learning Techniques. In this, they use Decision Tree, K Nearest Neighbor, Support Vector Machine calculations and the investigations were run utilizing 5-crease cross approval; in every approval around the data set 80 percent for preparation and 20 percent for testing are isolated. A title called Fake News Detection Using a Deep Neural Network [9] was introduced by Rohit Kumar Kaliyar et al. In this paper, they detect fake news using Naive Bayes Model, Decision Tree, CNN, K Nearest Neighbor algorithms. And, the Naive Bayes Model exhibits high precision.

Anjali Jain et.al proposed a brilliant framework for counterfeit news identification utilizing AI [10]. This paper presents a model and the strategies for distinguishing bogus news. The creator attempted to total the news with the assistance of Machine learning and regular language handling and later decided if the news is genuine or counterfeit utilizing Support Vector Machine. The after effects of the proposed model are contrasted with current models. The model being recommended functions admirably and decides the precision of tests up to 93.6 percent of exactness.

III. METHODOLOGY

Fake news detection is done on the taken dataset by applying feature extraction techniques

❖ CountVectorizer

Machine learning algorithms applied on the above extracted features

- ❖ DATA SPLITTING
- ❖ Random Forest
- ❖ Decision Tree
- ❖ Hybrid model

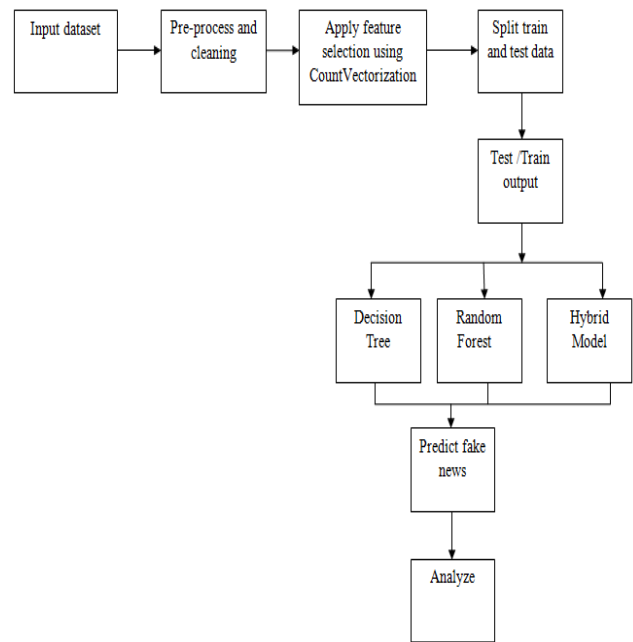


Fig. 1. System Architecture.

MODULES

CountVectorizer

CountVectorizer transforms a collection of text documents to matrix token count. This execution uses `scipy.sparse.csr` matrix making a sparse representation of the counts. The CountVectorizer offers a simple way that not only tokenize a set of text documents and create a vocabulary of known words but also to encode new documents using that vocabulary. Creates an instance of the class CountVectorizer. In the event that a string is passed to check stop list and the comparing stop list is as of now returned as the main acknowledged string esteem. In the event that a rundown is intended to contain stop words, which will all be discarded from the subsequent tokens. The result bag of words or vocabulary matrix is give as train and test input. Then the three machine learning algorithm is applied to arrive different result and compared.

Data Splitting

The dataset is split into train set and test set in the ration of 80% and 20% respectively. The landslide trigger is considered in numeric value range from 1 to 8 as Y-train value. The X-train values are all other numeric column considered for training.

Decision Trees

Decision Trees are a non-parametric, managed strategy for grouping and relapse learning. The objective is to develop a model that predicts the estimation of an objective variable by learning essential principles of choice got from the information qualities. In decision tree, leaves speak to class names, inside hubs speak to highlights and the branches speak to the estimation of the element in the interior hubs. Scikit-learn bundle gives a DecisionTreeClassifier which is the usage for a decision tree.

Decision Tree functions accordingly

- ❖ Position the best dataset attribute at the tree root.
- ❖ The DecisionTreeClassifier is a class equipped for playing out an informational collection multi-class characterization.

Random Forest Classifier

- Given that the preparation dataset contains N cases. Sub-tests are picked haphazardly with replacement from those n cases. Those irregular sub-tests chose from the preparation dataset are utilized to build singular trees.
- Expecting input factors are k, a number m is chosen so that $m < k$. At every hub m factors are arbitrarily chosen from k factors. To part the hub is picked the split which is the best of these m factors. The estimation of m stays unaltered when the forest is developed.
- Each tree is developed to a most extreme size without pruning.
- The class of the new thing is assessed dependent on the quantity of votes earned from the entirety of the decision trees joined.

Hybrid Model

We created hybrid model utilizing the ML algorithms.

- LIAR Dataset is taken as an input
- Pre-process and cleaning
- Apply CountVectorizer for data transformation.
- The combined model works based probabilities of Random Forest(RF).
- The probabilities from Random Forest are added to train data and fed to Decision Tree (DT) algorithm.
- Then, Decision Tree probabilities are identified and fed to test data.
- Finally the values were predicted by Hybridizing RF and DT.

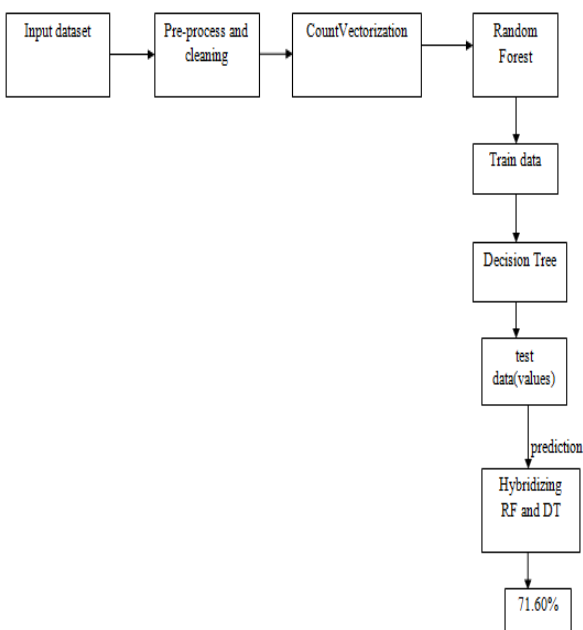


Fig 2.Hybrid Model.

By hybridizing the Decision Tree and Random Forest we got 71.60 percent accuracy.

IV. RESULTS AND DISCUSSION

We have implemented Fake news detection on the taken dataset by applying feature extraction techniques CountVectorizer. The extracted features are trained and predicted using three algorithms of machine learning, namely Decision Tree, Random Forest and Hybrid Model. The following table shows the results arrive from our implementation model for N-count Vectorizer feature extraction and prediction models.

Table 1.Experimental analysis of N-count Model

Algorithm	Accuracy
Decision Tree	69.30
Random Forest	69.80
Hybrid Model	71.60

V. CONCLUSION

As of late, double dealing identification in online audits and phony news has a significant job in the rules of industry requirement owing to the possible impact of counterfeit surveys might have on buyer conduct and buying choices. Scientists utilized profound having to learn from enormous dataset to increase in learning and along these lines get the best results by using word embedding for remove features or signals that perceive relations between words in syntactic and semantic. In this study, we had introduced a discovery model for counterfeit news, utilizing CountVectorization as the extraction method which was used to achieve its highest precision. The extracted data fed to the hybrid model and compared with the Random Forest and Decision Tree algorithms. We can conclude that the Hybrid Model achieved better accuracy than exclusive RF and DT Algorithms.

REFERENCES

1. Girgis S, Amer E, Gadallah M. "Deep learning algorithms for detecting Fake News in Online Text". In2018 13th International Conference on Computer Engineering and Systems (ICCES) 2018 Dec 18 (pp. 93-97). IEEE.
2. Aldwairi M, Alwahedi A. "Detecting Fake News in Social Media Networks". Procedia Computer Science. 2018 Jan 1;141:215-22.
3. Della Vedova ML, Tacchini E, Moret S, Ballarin G, DiPierro M, de Alfaro L. "Automatic online fake news detection combining content and social signals". In2018 22nd Conference of Open Innovations Association (FRUCT) 2018 May 15 (pp. 272-279). IEEE.
4. Granik M, Mesyura V. "Fake News Detection using Naive Bayes classifier". In2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON) 2017 May 29 (pp. 900-903). IEEE.
5. Akshay Jain, AmeyKasbe, "Fake News Detection". IEEE 2018.
6. Xu K, Wang F, Wang H, Yang B. "A first step towards combating Fake News over Online Social Media". In International Conference on Wireless Algorithms, Systems, and Applications 2018 Jun 20 (pp. 521-531). Springer, Cham.

Fake News Detection in Machine Learning Hybrid Model

7. Shu K, Mahudeswaran D, Liu H. "FakeNewsTracker: a tool for Fake News Collection, Detection and Visualization". Computational and Mathematical Organization Theory. 2019 Mar 1;25(1):60-71.
8. Ahmed H, Traore I, Saad S. "Detection of Online Fake News using N-gram analysis and Machine Learning techniques". In International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments 2017 Oct 25 (pp. 127-138). Springer, Cham.
9. Kaliyar RK."Fake News Detection Using A Deep Neural Network". In2018 4th International Conference on Computing Communication and Automation (ICCCA) 2018 Dec 14 (pp. 1-7). IEEE.
10. Jain A, Shakya A, Khatter H, Gupta AK. "A Smart System for Fake News Detection Using Machine Learning". In2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) 2019 Sep 27 (Vol. 1, pp. 1-4). IEEE.

AUTHORS PROFILE



P. Chandana, Studying Master of Technology, Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada. My Areas of interest are Machine Learning, Data Structures and Algorithms and Software Testing.



K. Sree Vijaya Lakshmi, M.Tech,(Ph.D), LMISTE, working as an Assistant Professor in the Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada since March 2010 and worked as a Programmer for 15 years in same Department and College from 1995 to 2010. Pursuing Ph.D. from JNTUK, Kakinada. Awarded a Certificate of Merit from DOEACC Society (Department of Electronics Government of India and Computer Society of India) on 31st July 2002. 7 papers were published in International Journals, 4 papers were presented in International Conferences and one paper was presented in National Conference. All these presented papers were published as book chapters in proceedings of those Conferences.