

Research Article

Fake news detection on twitter using machine learning algorithm

Shivani Suresh Nikam Prof.Rupali Dalvi

Department of Computer Engineering, Marathwada Mitra Mandal's College of Engineering, Pune, Savitribai Phule Pune University, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Social media networks have emerged as platforms for communication and information sharing. People all round the world trust heavily on news printed via social media to learn about regarding current events and activities. As a result, many users have started to exploit social media by broadcasting misleading news for financial and political purposes, which has an undesirable impact on society. Social media services like facebook and twitter have recently become a large and continuous supply of daily news. These platforms serve hundreds of millions of users and provide many services such as content creation and publishing. Not all published news on social media is reliable and accurate. Many people try to publish fake and erroneous news in order to influence public opinion. Our aim is to find a reliable and correct model that classifies a given article as either true or false.

Keywords: Fake news; Social media; natural language processing; machine learning

Introduction

Now a days, social medium has become a vital means that of large-scale data sharing and communication in all occupations, as well as promoting, journalism, promotion, and more. This change in consumption behaviors is due to some novel features such as mobility, free, and inter-activeness. However, the low value, fast access, and speedy dissemination of data of social media draw an outsized audience and change the wide propagation of fake news, i.e., news with intentionally fake data. For instance in 2016 fake news stories proclaiming that pope francis has supported donald trump for us president. When the Pakistani defense minister mistakenly believed a fake news article, he threatened a nuclear war with Israel. Fake news in India has led to episodes of violent behavior between castes and religions, intrusive with public policies. It usually spreads through the smartphone instant messenger whatsapp that had two hundred million monthly active users within the country as of Feb 2017. On Nov eight, 2016, Asian country established a 2,000-rupee currency bill on an equivalent day because the Indian 500 and 1,000 rupee note termination. Fake news went viral over WhatsApp that the note came equipped with spying technology that tracked bills one hundred twenty meters below the globe. Government minister Arun Jaitley refuted the falsities, however not before that that they had thought news retailers. These examples clearly demonstrate that pretend news stories are problematic not just for the believability of online journalism, however additionally because of their

damaging real-world consequences, leading to violence or influencing election results. Therefore, it becomes more and more necessary for policy manufacturers to manage and discourage the formation of fake news, for online business to detect and prevent pretend news, and for voters to protect themselves from fake news. Fake news on social media presents distinctive challenges. First, fake news is deliberately written to misinform readers, that makes it nontrivial to find simply supported content. Second, social media information is large-scale, multi-modal, largely user-generated, typically anonymous and noisy. Third, the consumers of social medium arrive from completely different backgrounds, have disparate preferences or desires, and use social media for various purposes. Finally, the low cost of making social media accounts makes it simple to make malicious accounts, like social bots, cyborg users, and trolls, all of which can become powerful sources of proliferation of fake news. Despite the importance of the problem, our understanding of fake reports remains restricted. As an example, we would like to understand why people produce fake news, who produces and publishes it, how fake news spreads, what characteristics distinguish fake news from legitimate news, or why some people are a lot of susceptible to fake news than others. Therefore, we tend to propose to know fake news with disciplines like journalism, psychology, and science, and characterize the distinctive characteristics for its detection. Establishing a enhanced understanding of fake news will allow us to come up with algorithmic solutions for detecting fake news and managing it before fake news is widely distributed.

Literature Survey

Daniel (Yue) Zhang et al. [1], considered the source reliability, report credibility, and a source's historical behaviors to effectively address the misinformation spread and data sparsity challenges in the truth discovery problem. Author used K-means clustering algorithm for categorization of news.

Hadeer Ahmed et al. [2], proposed features extraction techniques -Term Frequency and Term Frequency-Inverted Document Frequency and apply various machine learning algorithms Stochastic Gradient Descent, Support Vector Machines, Linear Support Vector Machines, K- Nearest Neighbour and Decision Trees. This model achieves its highest accuracy when using unigram features and Linear SVM classifier. The highest accuracy score is 92%.

Ghaith Jardaneh et al. [3], proposed features extracted using Tweepy API, and apply machine learning algorithms Random Forest, Decision Tree, AdaBoost, and Logistic Regression.

The experimental evaluation shows that system can filter out fake news with an accuracy of 76%.

Mykhailo Granik et al [4], proposed Naive Bayes classifier for identifying fake news. This approach was implemented as a software system and tested against various data set of Facebook etc. which provided an accuracy of 74%. This paper did not consider punctuation errors, leading to a low accuracy.

Shlok Gilda et al [5], proposed feature extraction techniques- term frequency-inverse document frequency of bi-grams and probabilistic context free grammar. Classification algorithms - Support Vector Machines, Stochastic Gradient Descent, Gradient Boosting, Bounded Decision Trees, and Random Forests. TF-IDF of bi-grams fed into a Stochastic Gradient Descent model identifies non-credible source with an accuracy of 77.2%, with PCFGs having slight effects on recall.

Akshay Jain et al [6], proposed feature extraction techniques- TF-IDF. Classification algorithm- Naive Bayes. This paper did not consider punctuation errors, leading to a low accuracy.

Rohit Kumar Kaliyar et al [7], proposed features extraction techniques -Term Frequency (TF) and Term Frequency Inverted Document Frequency (TF- IDF), word2vec, chi2. Machine learning models like Naive Bayes, K nearest neighbors, Decision tree, Random forest and Deep Learning networks like Shallow Convolutional Neural Networks (CNN), Very Deep Convolutional Neural Network (VDCNN), Long Short-Term Memory Network (LSTM), Gated Recurrent Unit Network (GRU), Combination of Convolutional Neural Network with Long Short-Term Memory (CNN-LSTM)

and Convolutional Neural Network with Gated Recurrent Unit (CNN LSTM).

Amitabha Dey et al [8], proposed features extraction techniques -Term Frequency (TF) and Term Frequency.

Machine learning algorithm -k-nearest neighbor algorithm.

Terry Traylor et al [9], used Textblob, Natural Language, and SciPy Toolkits to develop a novel fake news detector that uses quoted attribution in a Bayesian machine learning system as a key feature to estimate the likelihood that a news article is fake. The resultant process precision is 63.333% effective at assessing the likelihood that an article with quotes is fake.

Proposed Methodology

Affected intellect perception is an rising investigation area with few open datasets. Data correctness on Cyberspace, abnormally via web-based system media, is an indisputable important concern, yet web-scale communication hampers, capability to admit, assess and right such communication, or expected "counterfeit news," Current in these phase. We build up a unequivocal NLP based classifier to separate among fake and honest news story. Fake news is a incidence which is having a significant effect on our social life, in specific in the party-political world. Fake news credit is an emerging investigation area which is ahead interest but complex some challenges due to the incomplete quantity of possessions (i.e., datasets, published literature) available.

A. Architecture

The system works on previously trained Machine Learning algorithms. Several machine learning algorithms have been trained by providing a data set of both fake and genuine news.

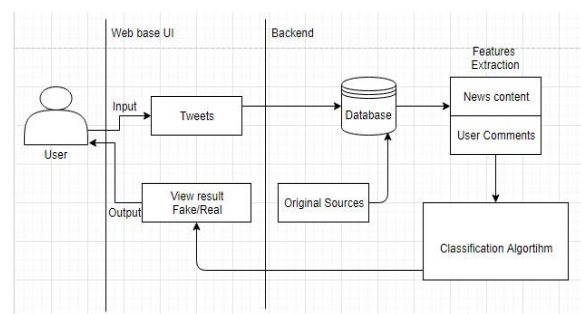


Figure no 1. System Architecture

The summary of overall technique is as follows. The system consists of two major components: a web-based user interface and a backend which integrates our fake news detection model. The web-based interface provides users with explainable fact-checking

of news. A user can input the tweet The backend consists of several components:

- (1) a database to store the pre-trained results as well as a crawler to extract hidden news and its comments;
- (2) the detection module, which gives the detection result of news.

Text Pre-Processing

It is a technique that is used to translate the raw information into a clean data set. Each and every time the data is collected from different sources it is collected in fresh format which is not achievable for the study. For achieving improved results from the realistic model in Machine Learning developments the format of the data has to be in an accurate manner. So in data pre-processing is required because of the presence of unformatted real world data.

Tweet Segmentation

Segmenting act of breaking up a sequence of series into parts such as phrases, words, keywords, symbols and other features called tokens. Tokens or words are separated by whitespace, punctuation marks or line breaks. White space or punctuation marks may or may not be included depending on the need.

3. Features Selection The initial incentive for feature assortment is that the social data often cover many dissimilar features that are problematic to deal with this eye, and most of the features are ended except for exact tasks. to deal with this problem, Apply feature removal methods. Feature selection is often favored over extraction; [8] because the selected features have more comprehensible and useful they select the three chief features first is Physical Features Structural features capture Twitter-specific possessions of the tweet stream, including tweet capacity and activity deliveries. Second is User features capture possessions of tweet authors, such as connections, account ages, friend/follower counts, and Twitter verified rank and third topographies Content topographies measure textual aspects of tweets, like schism, subjectivity, no of commentaries and contract.

A. Algorithms

In machine learning, Naive Bayes classifiers are a family of easy "probabilistic classifiers" based on applying Bayes' theorem with powerful (naive) independent assumptions between the features. Naive Bayes classifiers are extremely scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Support Vector Machines (SVM) are an arrangement of related supervised learning techniques operated for grouping and classification.

Algorithm : for detecting News is fake or not:

- 1) Input: D dataset, user-comments features
- 2) Output: Classification of News
- a) for each application news_id in D do
- b) Get user-comments features and stored on vector x for news_id
- c) x.add (Get_Features(news_id));
- d) end for
- e) for each application in x vector do
- f) Fetch first feature and stored in b, and other features in w.
- g) $hw, b(x) = g(z)$ here $z = (wTx + b)$
- h) if $(z \geq 0)$
- i) assign $g(z)=1$;
- j) else $g(z)=-1$;
- k) end if
- l) end for

Contribution

In our proposed system we will find the users who spread fake news after classification of news.

Result and Discussions

To train the system, dataset collected from two different sources Kaggle and Reuters websites. In the preprocessing phase, non-semantic words such as prepositions, conjunctions, and pronouns also known as stop words are removed from the document. The redundant data in the form of textual strings are removed from the document using a regular expression (regex) in the next step. The regex and pandas library has been used to perform the pre-processing task. Regex library has been used in python to define a search pattern using a sequence of characters, whereas dropna method from pandas is used for cleaning the missing values in python DataFrame. We have used random_state function to select the entries from the dataset which is further used to split training and testing data points as it is used to split the data randomly. The data retrieved are then split into two parts, training (0.67) and testing (0.33) sets. The feature extraction phase is then carried out to retrieve meaningful features from the textual data. In this phase, the features are extracted from the articles. In order to extract numerical features from a textual document, tokenization, counting and normalization are done. During tokenization, each word is given a unique integer ID, following which occurrence of tokens is counted and then normalization of such tokens takes place. The processed dataset retrieved after pre-processing and feature extraction phase is then fed to the classification phase for the identification of fake news. After the classification of news we will find the users who spread fake news and we will assign them creditability score.

The performance measures for binary classifiers applied with the help of a confusion matrix defined by four cells as shown in Table 1, where

Table no 1. Confusion Matrix

Predicted Actual ↓	→ Fake	Real
	Fake	TP (a)
Real	FN (c)	TN (d)

- 'a' counts the predicted document as 'Fake' when actually it is 'Fake,' known as true positive (TP) rate.
 - 'b' counts the predicted document as 'Real' when actually it is 'Fake,' known as false positive (FP) rate.
 - 'c' counts the predicted document as 'Fake' when actually it is 'Real,' known as false negative (FN) rate.
 - 'd' counts the predicted document as 'Real' when actually it is 'Real,' known as true negative (TN) rate.
- The performance measure has been evaluated from the above confusion matrix cells. The measures computed from the matrix are precision as represented by Eq. (1), recall by Eq. (2), specificity by Eq. (3), accuracy by Eq. (4), F1-score by Eq. (5) as shown below. The precision metric helps to calculate the proportion of news article that are predicted fake and actually also belongs to the fake news article category.

$$\text{Precision (Pr)} = \frac{a}{a + b} \quad (1)$$

The recall metric helps to calculate the proportion of news article that are predicted fake but actually belongs to both fake and real articles.

$$\text{Recall (Re)} = \frac{a}{a + c} \quad (2)$$

The specificity metric helps to calculate the proportion of news article that are correctly predicted as a real news article known not to be fake. Specificity is measured as inverse of recall metric.

$$\text{Specificity (Sp)} = \frac{d}{d + b} \quad (3)$$

Accuracy is the performance measure and it is simply a ratio of correctly predicted observation to the total observations. Accuracy is only measured as a strong metric when both false negative and false positive values are closer to each other else the metric is not considered as a good performance measure.

$$\text{Accuracy (Acc)} = \frac{a + d}{n}, \text{ where } n = a + b + c + d > \quad (4)$$

F1 Score is the weighted average of Precision and Recall.

Therefore, this score takes both false positives and false negatives into account.

$$\text{F1-score (F1)} = \frac{2 \times \text{Pr} \times \text{Re}}{\text{Re} + \text{Pr}} \quad (5)$$

Here, a, b, c, and d represent the TP, FP, FN and TN rates, respectively. predicted document as 'Real' when actually it is 'Fake,' known as false positive (FP) rate, c counts the predicted document as 'Fake' when actually it is 'Real,' known as false negative (FN) rate, d counts the predicted document as 'Real' when actually it is 'Real,' known as true negative (TN) rate.

The predictions made by the classification models are evaluated in this phase based on their performance metrics.

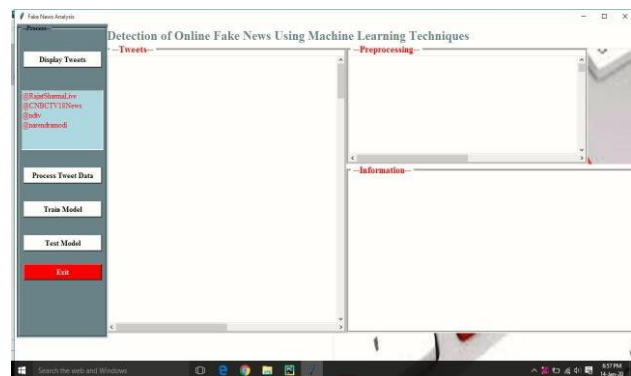


Figure no. 2 Enter tweets



Figure no 3. Display tweets

Conclusions

With the increasing popularity of social media, more and more people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has strong negative impacts on individual users and

broader society. In this paper, we will develop a system which identify fake news on twitter. In our proposed system we will use naive bayes algorithm for categorization of tweets.

References

- [1]. Daniel (Yue) Zhang, Dong Wang, Nathan Vance, Yang Zhang, and Steven Mike, "On Scalable and Robust Truth Discovery in Big Data Social Media Sensing Applications", 2019 IEEE transaction on big data.
- [2]. Hadeer Ahmed, Issa Traore, and Sherif Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques", Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments. ISDDC 2017.
- [3]. Ghaith Jardaneh, Momen Buzz, Hamed Abdelhaq, Douglas Johnson, " Classifying Arabic Tweets Based on Credibility Using Content and User Features", 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT).
- [4]. Mykhailo Granik, Volodymyr Mesyura, "Fake News Detection Using Naive Bayes Classifier", 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON).
- [5]. Shlok Gilda, "Evaluating Machine Learning Algorithms for Fake News Detection", 2017 IEEE 15th Student Conference on Research and Development (SCORED).
- [6]. Akshay Jain, Amey Kasbe, "Fake News Detection", 2018 IEEE International Students Conference on Electrical, Electronics and Computer Sciences.
- [7]. Rohit Kumar Kaliyar, "Fake News Detection Using A Deep Neural Network", 2018 4th International Conference on Computing Communication and Automation (ICCCA).
- [8]. Amitabha Dey, Rafsan Zani Rafi, Shahriar Hasan
- [9]. Parash, Sauvik Kundu Arko and Amitabha Chakrabarty, "Fake News Pattern Recognition using Linguistic Analysis", 2018 Joint 7th International Conference on Informatics, Electronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR).
- [10]. Terry Traylor, Jeremy Straub, Gurmeet, Nicholas Snell, "Classifying Fake News Articles Using Natural Language Processing to Identify In- Article Attribution