Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/140

FAKE NEWS DETECTION IN HINDI NEWS USING A PASSIVE CLASSIFIER

[1]Noor Mohd., [1]Kireet Joshi, [1]Rohan Raj, [2]Dibyahash Bordoloi

[1]Department of Computer Science & Engineering Graphic Era Deemed to be University, Dehradun, Uttarakhand, India
[2]Department of Computer Science & Engineering, Graphic Era Hill University, Dehradun, Uttarakhand India, 248002

ABSTRACT

During this modern time when people have such easy access to social media and internet, they have become very connected and sometimes it takes less than a minute for any piece of information to get viral. Consequently, fake news is spreading very easily and quickly on those platforms which lead to inappropriate actions and consequences. Sometimes this fake information is spread by people intentionally to create chaos among the people. The aim of this paper is to create a model wherein it classifies each piece of information provided to it into two categories, either non-hostile, fake, offensive or defamation. In order to accomplish the project's objective, we have suggested the algorithm term frequency - inverse document frequency. It is a statistical technique for determining how pertinent a word is in a group of documents. It is widely used for information retrieval and summarization. This method is used to calculate the term frequency of each word and further classify the words according to the occurrence in each particular type of document in the training dataset. Furthermore, the testing dataset uses this information to find the nature of the text document. This method is proposed for the natural language “Hindi” because of the unavailability of such fake news classifiers in this language.

Keywords: Fake News Detection, Machine Learning, TF-IDF, Passive Classifier

INTRODUCTION

Before the internet was ever a thing, fake news existed. A common definition of fake news is an item that is purposefully false in order to mislead readers and form a false impression in their minds that would cause them to act in an improper manner. When people read these erroneous articles and messages, their brains generate an opinion that is wholly incorrect, and this false knowledge prevents them from thinking and acting in the right way. The goal of this project is to develop a system that will allow users to confirm whether a news story is authentic or not by comparing it to other messages. These days, messages like WhatsApp forwards travel like wildfire, therefore a fake news classifier appears vital in this situation. This model concentrates on the false news classifier model for the "Hindi" language in particular so that non-Hindi speakers can also utilize this model. Since there aren't many models for natural languages, my goal is to develop one for "Hindi" speakers and listeners. The technique utilized for this is term frequency - inverse document frequency. In coding, passive aggressive classifiers are also employed. In the past, humans have

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/140

gotten their information from a variety of reliable sources that were required to abide by a strict set of standards [1]. With no underlying principles or rules, several new techniques for disseminating and sharing information have been suggested as a result of the internet's accessibility [2, 3]. The news that has been purposefully picked with the intent to deceive readers is referred to as fake news [4, 5]. Fake news is spread, often to suit political or financial interests [6]. They publish this fake news, as well as using fraudsters or bots to spread information more quickly, and they employ various deception techniques to evade being noticed [7]. Tech firms like Google, Facebook, and Twitter have, at best, made an effort to address this specific issue. These initiatives haven't done much to address the issue, though, since the corporations have shifted to not paying the people connected to these sites the money they would have made from the increased traffic.

LITERETURE SURVEY

2.1 RELEVENCE

As social media and the internet became more popular, significant problems arose. Social media is a terrific instrument for spreading free, unrestricted information at an exponential rate, but it also provides the ideal environment for creating and spreading this false information.

As more people use the internet daily and are regularly exposed to new information and stories, fake news occasionally becomes fairly convincing and disseminates quickly. Recent academic study demonstrates that experts in a variety of professions have been deeply troubled by the problem of fake news.

For instance, fake news is now more than just a marketing problem, according to some authors [8]. The role of the information technology (IT) department is also increasingly being seen as a part of the issue

In the past, it was assumed that the two departments indicated above would be in charge of handling any repercussions from the spread of false information about an organization. The practice has changed from what it was a few years ago, when those who engaged in this kind of gimmicks were only interested in increasing internet content, to a situation where hackers are now involved.

Some content creators, in particular, have resorted to including content with hazardous code as part of what is given on their web pages, enabling visitors to these sites to click the links and unintentionally download the virus. Too much time has passed since false information first surfaced.

Study of the many models made it clear that while several were made for the purpose, they were largely for the English language and not for the native languages used by the majority of Indians

Six different algorithms, including Naive Bayes, Decision Trees, support vector machines, neural networks, random forests, and XG Boost, are used in paper [9]

The techniques employed are highly complex and require a thorough knowledge of the subject.

We have used term frequency inverse document frequency for the implementation to make the working process simpler. We have also employed the passive aggressive classifier in addition to it.

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/140

The second study cited [10] makes use of logistic regression, naive bayes, and SVM. They compared the models they employed and made an effort to determine which algorithm was the most successful. The naive bayes algorithm demonstrates, a maximum accuracy of 83%.

Furthermore, fake news is bad for society as a whole in addition to the detrimental effects it has on individuals. The "balance of the news ecosystem" [11] might be disrupted by false information because of how extensively it is spread. For instance, "the most popular false news was substantially more frequently distributed on Social Media" [11] during the 2020 US Presidential election than the "most authentic news networks".

PROPOSED METHODOLOGY

The paper's ability to accurately categorize user-provided news as unfriendly, false, offensive, or defamatory is one of its key goals. We are utilizing the tf-idf and passive aggressive classifier algorithm. Term frequency inverse document frequency is referred to as tf-idf. It can be summed up as determining how pertinent each word in a corpus or series is to a text.

The frequency of a phrase in the corpus balances the rise in meaning brought on by a word's increased usage in the text (data-set). The letter "d" denotes the number of times a given term (abbreviated "t") appears in a document.

It follows that it is obvious that a term becomes even more relevant when it occurs in the text.

Each distinct phrase used in the document has a corresponding item with the value "term frequency." tf(t,d) = count of instances of given words in document / number of instances of given words in document (1)

Document frequency, a statistic somewhat similar to tf, is used to assess the meaning of the text over the whole dataset collection. The sole distinction is that TF represents the frequency counter for the phrase in document "d," whereas df apply for the number of times the term "t" appears in document set "N".

Simply put, there are DF papers that contain the word. df(t) = total occurrences of the instances of words in a document (2)

The main purpose of idf (Inverse Document Frequency) is to assess a document's applicability. The search's primary goal is to find the pertinent records that meet the requirement. Because tf values each term equally, it is impossible to determine a phrase's significance in a document by looking at its frequency alone. In order to determine the phrase's document frequency, the first step is to count the number of documents that contain the phrase.

$$df(t) = N(t) \quad (3)$$

where df(t) is the probability of the term in documents and N(t) is the number of documents containing the term t.

$$N/\ df(t) = \log(idf(t)) \quad (4)$$

One of the finest measures for figuring out how important a phrase is to the text in a document sample is tf-idf. According to the term frequency (tf) and reverse document frequency (tf-idf)

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/140

balancing approach, each word in a document is given a weight. It is assumed that the words with greater weight ratings are more significant.

$$\text{tf-idf}(t, d) = \text{tf}(t, d) * \text{idf}(t) \quad (5)$$

Algorithms that are passive-aggressive are frequently used in large-scale learning. Online machine learning techniques make use of sequential input data and update the machine learning model one step at a time, in contrast to batch learning, which uses the entire training dataset all at once. This is quite useful when there is a lot of data and it takes a while to train the entire dataset due to the size of the data.

We first import the pandas and sklearn packages that we will need for our solution. The complete dataset is divided into two columns, x and y, where x represents the text (which is the news story) and y is the label (the nature of the article). The complete x and y dataset is divided into a training set and a testing set, with the training set making up 80% of the dataset and the testing set making up the remaining 20%. Then we compute the corresponding values of the weightage of the words using the tf-idf vectorizer.

With a max df value of 0.7, we ignore phrases that appear in more than 70% of documents. The regularisation parameter, or parameter "C," of the passive aggressive classifier has a default value of 1.0 in this instance. The training data can be iterated over a maximum of 50 times. Sending the text data for term frequency-inverse document frequency verification is the initial step. The information obtained in this stage is then used to classify passive aggressive behaviour, and the end result of this procedure is what we are left with. The model's accuracy is also determined.

WORKFLOW

The workflow of the project implementation is shown below. The steps are shown one after another and in the chronological order. The packages are imported and the dataset is loaded. 80% of the dataset is loaded in the training set and 20% of the dataset is loaded in the testing set. The input text is provided to the model and then the nature of the input text is provided in the output.

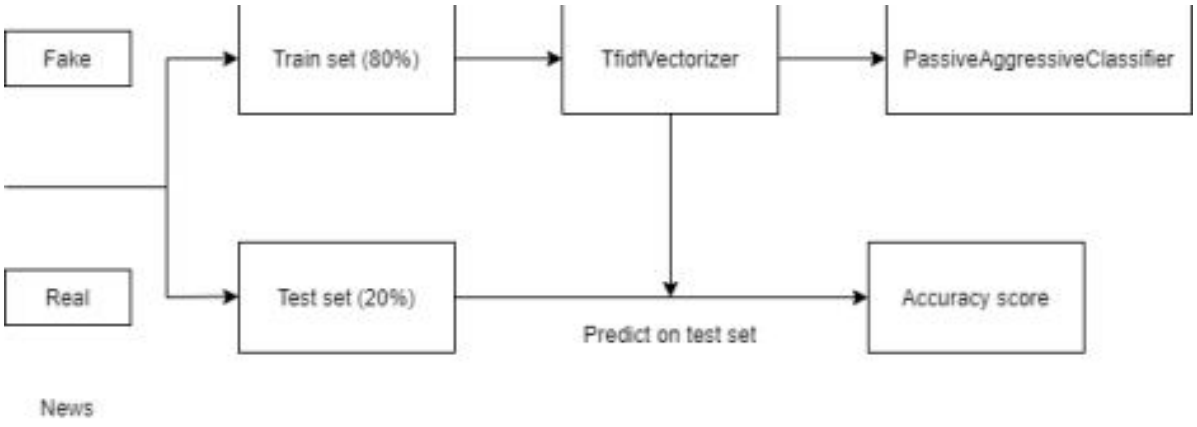

Figure 1: Training and Testing Categorization

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/140

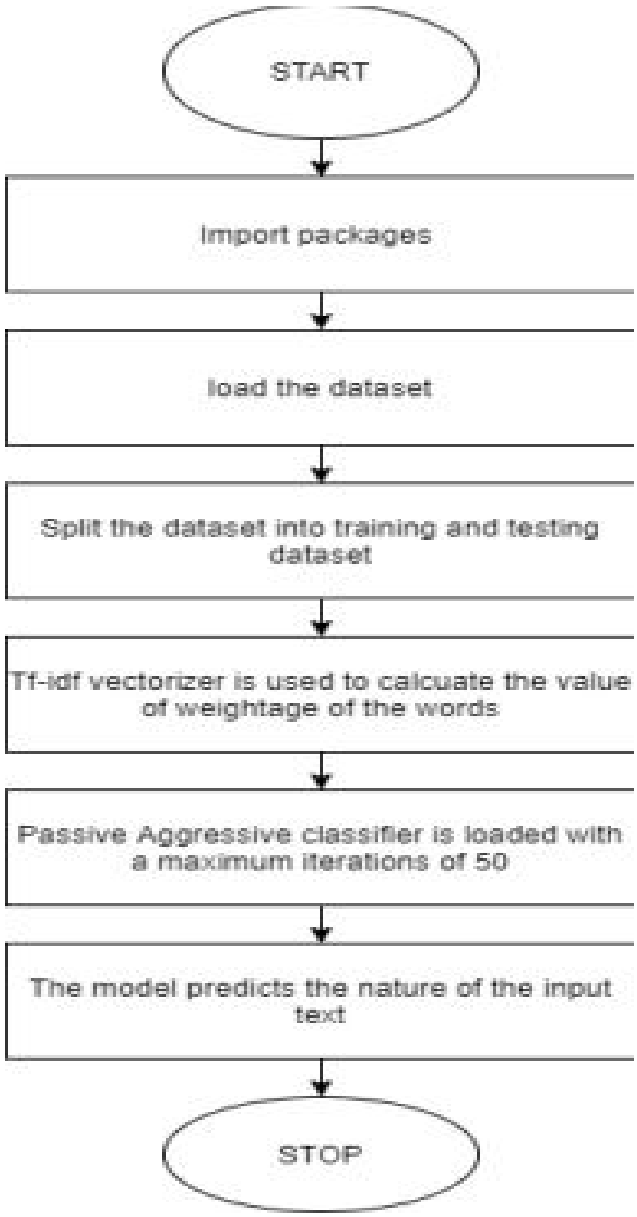

Figure 2: Workflow of the Project Implementation Process

ALGORITHM PROPOSED

The algorithm used to achieve the aim and final result comprises majorly of tf-idf and the use of passive aggressive classifier.

STEP 1. We import the packages pandas and sklearn and relative packages for processing

STEP 2. We apply the our use dataset using pandas

```
dataframe = pd.read_csv('constraint_Hindi_Train - Sheet1 (1).csv')
```

STEP 3. The two columns of the dataset are divided into x and y and it define the as: x column is

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/140

the text and y is the label of the text i.e. the nature if the text article whether it is non-hostile, fake, defamation and offensive.

STEP 4. The dataset is splitted into training and testing dataset. 80% is the training dataset and 20% is the testing dataset.
x_train,x_test,y_train,y_test = train_test_split (x,y,test_size=0.2,random_state=0)

STEP 5. A matrix of tf-idf vectorizer is created by converting a group of raw documents. tfvect = TfidfVectorizer(max_df=0.7)

tfid_x_train = tfvect.fit_transform(x_train)
tfid_x_test = tfvect.transform(x_test)

STEP 6. Passive aggressive classifier is used, if the class is correct then we keep the model and if it is incorrect we update to adjust to this misclassified information.

Classifier = Passive Aggressive Classifier (max_iter=50)
classifier.fit (tfid_x_train,y_train)

STEP 7. The model accuracy is loaded and confusion matrix is displayed.

STEP 8. In the last step we can provide the new article we want to verify the nature of the text.

RESULTS

The whole model was executed using python and the following observations were analysed. Our training dataset has more than 1000 values.

Training Sample:

Testing Sample:

Output:

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/140

In [5]:
```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)
y_train
```

Out[5]:
```
4518      offensive
4472           hate
799     non-hostile
4809           fake
1043    non-hostile
           ...
4931    non-hostile
3264           hate
1653    non-hostile
2607     defamation
2732           fake
Name: label, Length: 4582, dtype: object
```

In [6]:
```
tfvect = TfidfVectorizer(max_df=0.7)
tfid_x_train = tfvect.fit_transform(x_train)
tfid_x_test = tfvect.transform(x_test)
```

In [7]:
```
classifier = PassiveAggressiveClassifier(max_iter=100)
classifier.fit(tfid_x_train,y_train)
```

Out[7]:
```
PassiveAggressiveClassifier(max_iter=100)
```

In [8]:
```
y_pred = classifier.predict(tfid_x_test)
score = accuracy_score(y_test,y_pred)
print(f'Accuracy: {round(score*100,2)}%')
```

```
Accuracy: 67.45%
```

In [1]:
```
import pandas as pd
```

In [2]:
```
dataframe = pd.read_csv('constraint_Hindi_Train - Sheet1 (1).csv')
dataframe.head()
```

Out[2]:

	text	label
0	मेरे देश के हिन्दु बहुत निराले है। कुछ तो पक्क...	hate,offensive
1	सरकार हमेशा से किसानों की कमाई को बढ़ाने के लि...	non-hostile
2	सुशांत ने जो बिजनेस डील 9 जून को की थी, वो डील...	non-hostile
3	@prabhav218 साले जेएनयू छाप कमिने लोग हिन्दुओं...	defamation,offensive
4	#unlock4guidelines - अनलॉक-4 के लिए गाइडलाइन्स...	non-hostile

In [3]:
```
x = dataframe['text']
y = dataframe['label']
```

In [4]:
```
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.metrics import accuracy_score, confusion_matrix
```

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/140

```
In [9]: cf = confusion_matrix(y_test,y_pred, labels=['non-hostile','fake','offensive','defamation'])
        print(cf)

        [[571  16   7   2]
         [ 14 126  16  12]
         [  4  12  19   9]
         [  6  23   6  13]]

In [10]: def fake_news_det(news):
             input_data = [news]
             vectorized_input_data = tfvect.transform(input_data)
             prediction = classifier.predict(vectorized_input_data)
             print(prediction)

In [11]: fake_news_det('चीन ने UN में तर्क दिया की धरती का विश्व ही अजर, अमर को आतंकी नहीं मानता तो हम कैसे माने चुल्लू भर मूत्र में डूब मरो गद्दारो।
         ['fake']

In [12]: fake_news_det('कंगना रणावत ने मुम्बई की तुलना POK से की है, क्या अब उनके खिलाफ देशद्रोही का मुकदमा दर्ज होगा ? क्या कंगना जेल जायेगी ?या
         ['defamation']
```

The model works nicely with a good percentage of accuracy and is very useful to classify true news against the fake ones. The fake news is classified into non-hostile, defamation, offensive and fake categories and after working on the input sample provided the final output is given to the user. The final accuracy of the **proposed** model is around 67.45% that is better than of **BERT** model 52% but was quite low than LSTM model of 94%. This is because the dataset is collected from different sources and certain labelling have been done by us. The nature of the dataset is very different and this doesn't give sufficient training to the model due to which the accuracy eventually decreases.

FUTURE WORK

The model used in this project is limited only to Hindi language and makes use of only term frequency- inverse document frequency method along with passive aggressive method. The future scope can be to use more methods and create a model with better accuracy and do the same for other natural languages as well.

REFERENCES

1. Fang, Y., Gao, J., Huang, C., Peng, H., & Wu, R. (2019). Self-multi-head attention-based convolutional neural networks for fake news detection. *PloS one*, *14*(9), e0222713
2. Subramani, P., & BD, P. (2021). Prediction of muscular paralysis disease based on hybrid feature extraction with machine learning technique for COVID-19 and post-COVID-19 patients. *Personal and ubiquitous computing*, 1-14.
3. Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the association for information science and technology*, *52*(1), 1-4.
4. Gahirwal, M., Moghe, S., Kulkarni, T., Khakhar, D., & Bhatia, J. (2018). Fake news detection. *International Journal of Advance Research, Ideas and Innovations in Technology*, *4*(1), 817-819.
5. Parameshachari, B. D. (2022). Big Data Analytics on Weather Data: Predictive Analysis Using Multi Node Cluster Architecture. *International Journal of Computer Applications*,

Webology, Volume 18, Number 4, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I4/140

0975-8887.

6. Singh, T. D., Divyansha, D., Singh, A. V., Sachan, A., & Khilji, A. F. U. R. (2020, December). Debunking Fake News by Leveraging Speaker Credibility and BERT Based Model. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 960-968). IEEE.
7. Le, N. T., Wang, J. W., Le, D. H., Wang, C. C., & Nguyen, T. N. (2020). Fingerprint enhancement based on tensor of wavelet subbands for classification. *IEEE Access*, *8*, 6602-6615.
8. Rubin, V. L., Chen, Y., & Conroy, N. K. (2015). Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1-4.
9. Manzoor, S. I., & Singla, J. (2019, April). Fake news detection using machine learning approaches: A systematic review. In *2019 3rd international conference on trends in electronics and informatics (ICOEI)* (pp. 230-234). IEEE.
10. Agarwalla, K., Nandan, S., Nair, V. A., & Hema, D. D. (2019). Fake news detection using machine learning and Natural Language Processing. *International Journal of Recent Technology and Engineering (IJRTE)*, *7*(6), 844-847.
11. Guo, Z., Shen, Y., Bashir, A. K., Imran, M., Kumar, N., Zhang, D., & Yu, K. (2020). Robust spammer detection using collaborative neural network in Internet-of-Things applications. *IEEE Internet of Things Journal*, *8*(12), 9549-9558.