

PUNE RESEARCH SCHOLAR ISSN 2455-314X AN INTERNATIONAL MULTIDISCIPLINARY JOURNAL VOL 7, ISSUE 2

A STUDY ON VARIOUS TECHNIQUES APPLIED FOR DETECTION OF FAKE NEWS IN SOCIAL MEDIA

BHAGYASHRI TIJARE

Assistant Professor Prerna College of Commerce Nagpur (**MS**) **INDIA** **VIDYA MULMULE**

Assistant Professor Prerna College of Commerce Nagpur (MS) INDIA

ABSTRACT

Fake news and hoaxes have been there since before the arrival of the Internet. The widely accepted definition of Internet fake news is: "fictitious articles deliberately created to mislead readers". Social media and news outlets publish fake news to escalate readership or as part of psychological conflict. This study analyzes the generality of fake news in light of the advances in communication made possible by the emergence of social networking sites. We use Machine learning techniques to classify the datasets. The purpose of the work is to come up with a solution that can be utilized by users to detect and filter out sites containing false and deceptive information. We use simple and carefully selected features of the title and post to accurately identify information of fake posts. The experimental result shows 80.0% accuracy using logistic classifier.

Keywords—Fake News, Machine Learning

I NTRODUCTION

The increase in use of social media exposes users to misleading information, fake and satire advertisements. Fake news or wrong information is defined as fabricated information presented as the truth . It is the publication of known false information and sharing it amongst individuals . It is the intentional publishing of false information and can be verified as false through fact-checking . Social media platforms allow individuals to share fast information with only a click of one share button. In previous studies the effect of the spreading and exposure to misleading information are investigated. Some studies determined that everybody has problems with identifying fake news, not just users of a particular age, gender or

BHAGYASHRI TIJARE

VIDYA MULMULE

1P a g e



education. The literacy and education of faux news is important within the combating of the spreading of wrong information. This review identify and discuss the factors involved within the sharing and spreading of faux news. The outcome of this review should be to provide users with the abilities to detect and recognize misinformation and also to cultivate a desire to stop the spreading of false information.

The Impact of Fake News –

The internet is mainly driven by advertising . Websites with sensational headlines are extremely popular, which results in advertising companies capitalizing on the high traffic to the location. It was subsequently discovered that the creators of faux news websites and knowledge could make money through automated advertising that rewards high traffic to their websites. The question remains how misinformation would then influence the general public. The spreading of misinformation can cause confusion and unnecessary stress among the general public . Fake news that's purposely created to mislead and to cause harm to the general public is mentioned as digital disinformation. Disinformation has the potential to cause issues, within minutes, for many people. Disinformation has been known to disrupt election processes, create unease, disputes and hostility among the general public .

II. LITERATURE SURVEY

Through this section, we summarize some of the existing research works in the field of Machine learning to analyses about Fake News Detection and build a model according to the existing applications.

Conroy N. J., Rubin, V. L., & Chen, Y. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), 1-4. (2015)

- In this literature, the author mainly focused on categorizing the news based on finding the degree of accuracy or correctness in the news.
- They focused on technologies adopted for fake news detection.
- Include mainly two categories for assessment linguistic cue approach (with machine learning) and network analysis approach.
- Both approaches adopt machine learning techniques for training classifiers to suit the analysis.
- Future scope: -combination of linguistic cue and machine learning on network based behavioral data.
- Focused on structured dataset such as text

BHAGYASHRI TIJARE

VIDYA MULMULE

2P a g e



PUNE RESEARCH SCHOLAR ISSN 2455-314X AN INTERNATIONAL MULTIDISCIPLINARY JOURNAL VOL 7, ISSUE 2

• These papers show the current development of correctness assessment methods, their goals and classes with the aim to propose new hybrid system for detection.

Wu Liang and Huan Liu. "Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate." (2018).

- The author focused on classification of social media content for social media mining and text categorization problem mainly including text with hash tags and words.
- Through this literature, author focused on modeling the propagation of messages in social network.
- Here the author proposes an approach called TRACEMINER to infer the embedding of social media users in social network structure and to create a LSTM-RNN model to represent the path of messages.
- They provide optimization methods to trace miner to guarantee the correctness and evaluate the performance of the real world social network data.

III. IMPLEMENTATION & METHODOLOGY

A. Platform and Technologies

a. Jupyter Notebook

- The IPython scratch pad is otherwise called Jupyter note pad.
- It is an open source aggregate computational condition which consolidate codes, science, plots and content.
- It is a user-friendly platform.
- We upload dataset and python files in notebook and implement codes of machine

Learning /deep learning models. We get result in form of confusion matrix or

graph.(https://jupyter.org/)

b. PyCharm

- PyCharm not only supports Python programming language, but also handles code written in SQL and similar database languages along the line.
- Jupyter users will definitely agree with the packages Pycharm integrates with libraries such as

BHAGYASHRI TIJARE

VIDYA MULMULE

3Page



Numpy and Matplotlib for numerical computing and data visualization in Python.

- Apart from computation, it also supports Web Development languages.
- One interesting feature of PyCharm is the presence of a sharp debugger with a graphical interface along with offering customizations.(https://www.jetbrains.com/pycharm/)

B. Dataset and collection

- A dataset is a collection of data which mainly comprises of single statistical data matrix, database table where every row corresponds to each member in datasets and each column represents variable.
- The dataset list values for each variable such as title, id, author, label etc.
- We commonly collect various datasets.
- Exclusively fake news article datasets contain news article are forged.

C.

D. Features and Factors

- **CONFUSION MATRIX** Confusion matrix is a type of statistical classification also known as error matrix. It grants visualization and usually called as matching matrix. Each row of matrix represents immediate prediction and column represent instances of a real class.
- **HASING VECTORIZER** Other type used in text classification is Hashing vectorizer crave a less memory and faster because they use Hashes rather than tokens. It gives an improved accuracy results than the TF-IDF vectorizer using Multinomial Naïve Bayes.
- **PRECISION** Precision also called positive predicted value is the fraction of significant instances among the retrieved instances.
- **ACCURACY** The capacity of the framework to precisely characterize information depends to a vast degree on the illustrations that you give.
- **RECALL** Recall also known as sensitivity is the fraction of significant instances that have been retrieved over the total amount of relevant instances.
- **F1-SCORE** F1 score also F-score or F-measure is a measure of a test's accuracy for binary classification.

BHAGYASHRI TIJARE

VIDYA MULMULE

4Page

E. Methodologies

Naïve Bayes Model

- It uses probabilistic approaches and are based on Bayes theorem. They deal with probability distribution of variables in the dataset and predicting the response variable of value
- They are generally used for text classification and used in medical diagnosis.
- Bayes theorem P(a|b) = p(b|a)p(b)/p(a)
- There are mainly 3 types of naïve base models. Gaussian Naïve Bayes, Multinomial naïve Bayes and Bernoulli Naïve Bayes. We have used Multinomial Naïve Bayes model for our project to detect fake news.
- Naive Bayes classifier is the straightforward procedure of creating classifier models that select class names to issue example, communicated has a vector of highlight esteems, where the class names are drawn from some limited set.
- Naive Bayes classifier model have worked well in many complicated real-world situations.
- An advantage of naïve Bayes classifier is that only requires less bulk of training data to access the parameters necessary for classification.
- We use fake_or_real_news dataset for implementation.
- By implementing this model on my data set, we've got an accuracy of 89%
- Precision, Recall, f1-Score got the same value (90%).
- it is shown below: Precision recall f1-score support (https://machinelearningmastery.com/bayes-theorem-for-machine-learning/)

Type of News	Precision	recall	F1-score
FAKE	0.90964	0.90	0.90
REAL	0.90171	0.90	0.90
Total	0.901935	0.90	0.90

BHAGYASHRI TIJARE

VIDYA MULMULE



K nearest neighbor

- They store the entire data set for the implementation. It is an instant based learning which is done by approximation.
- Their data values are arranged in a feature space. They depend on the value of 'k'.
- The data value or the feature unknown to us is found out by using the value of 'k'. That is, the nearest 'k' neighbors is taken and most occurring feature is observed. Here also we use train_mixed.csv dataset for the Implementation.
- K-Neighbors classifier is used to classify the data set.
- By Implementing this model on my data set, we've got an accuracy of 53%
- (Precision=.54%), (Recall=53%), (f1- Score=50%). (https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighborsalgorithm-6a6e71d01761)

Type of News	Precision	recall	F1-score
FAKE	0.59	0.27	0.37
REAL	0.52	0.81	0.63
Total	0.55	0.54	0.50

Decision Tree

- It is used to optically serve decisions and decision making.
- It uses tree like model of both classification and regression.
- They are commonly used for data mining that support machine learning.
- We use train_ mixed.csv data set file in naïve Bayes for this model. Decision Tree Regressor is used to classify the data set into smaller subsets and at the same time a decision tree is incrementally developed.
- By implementing this model on my data set, we've got an accuracy of (73%)

BHAGYASHRI TIJARE

VIDYA MULMULE

6Page



PUNE RESEARCH SCHOLAR ISSN 2455-314X AN INTERNATIONAL MULTIDISCIPLINARY JOURNAL VOL 7, ISSUE 2

- (Precision=75%), (Recall=74%), (f1-Score=73%)
- It is shown below:

Type of	Precision	recall	F1-score
INCWS			
FAKE	0.82	0.60	0.70
REAL	0.69	0.87	0.77
Total	0.75	0.74	0.73

IV. ADVANTAGES OF THE PRESENT SYSTEM

- 1) Fake news can become extremely influential and has the ability to spread exceedingly fast. With the increase of people using social media, they are being exposed to new information and stories every day. Misinformation can be difficult to correct and may have lasting implications. For example, people can base their reasoning on what they are exposed to either intentionally or subconsciously, and if the information they are viewing is not accurate, then they are establishing their logic on lies.
- 2) In addition, since false information is able to spread so fast, not only does it have the ability to harm people, but it can also be detrimental to huge corporations and even the stock market. For instance, in October of 2008, a journalist posted a false report that Steve Jobs had a heart attack. This report was posted through CNN's iReport.com, which is an unedited and unfiltered site, and immediately people re-tweeted the fake news report. There was much confusion and uncertainty because of how widespread it became in such a short amount of time. The stock of Job's company, Apple Inc., fluctuated dramatically that day due to one false news report that had been mistaken for authentic news reporting (Rubin, 2017).
- 3) Consumers only want to hear what they believe and do not want to find any evidence against their views. For instance, someone could be a big believer of unrestricted gun control and may desire to use any information they come across in order to support and justify their beliefs further. Whether that is using random articles from incredible sites, posts from friends, re-shared tweets, or anything online that does agrees with their principles.
- 4) Consumers do not wish to find anything that contradicts what they believe because it is simply not how humans function. People cannot help but favor what they like to hear and have a predisposition for confirmation bias. It is only those who strive for certain

BHAGYASHRI TIJARE

VIDYA MULMULE

7Page



academic standards that may be able to avoid or limit any biasness, but the average person who is unaware of false information to begin with will not be able to fight these unintentional urges.

Future Scope –

In addition, not only does fake news negatively affect individuals, but it is also harmful to society in the long run. With all this false information floating around, fake news is capable of ruining the "balance of the news ecosystem" (Shu et al., 2017). For instance, in the 2016 Presidential Election, the "most popular fake news was even more widely spread on Facebook" instead of the "most popular authentic mainstream news" (Shu et al., 2017). This demonstrates how users may pay more attention to manipulated information than authentic facts. This is a problem not only because fake news "persuades consumers to accept biased or false beliefs" in order to communicate a manipulator's agenda and gain influence, but also fake news changes how consumers react to real news (Shu et al., 2017). People who engage in information manipulation desire to cause confusion so that a person's ability to decipher the true from the false is further impeded. This, along with influence, political agendas, and manipulation, is one of the many motives why fake news is generated.

V. CONCLUSION

Fake News Detection is the analysis of socially relevant data to distinguish whether it is real or fake. In this project we explored different Machine learning models like Naïve Bayes, K nearest neighbors, Decision tree. We also explored the benefit of feature extraction, features like n-gram, TF-IDF features were extracted and used in our model. We also explored the effective of word embedding and word2vec features in Deep Neural networks. We also explored the use of select best and chi2 for feature extraction in Machine learning model. With the increase in the world 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. Fake news detection app is used to detect whether the news is real or fake. In this project we explored different Machine learning models like Naïve Bayes, Naive Bayes classifier model have worked well in many complicated real-world situations. An advantage of naïve Bayes classifier is that only requires less bulk of training data to access the parameters necessary for classification. Application of Naïve Bayes are Real time Prediction, Multi class Prediction, Text classification/ Spam Filtering/ Sentiment Analysis, Recommendation System. We also further discussed the datasets and evaluation metrics. At this aim, we implemented an algorithm combining several classification methods with text models. It performed well, and the accuracy results were relatively satisfying. Moreover, to achieve a higher accuracy, we will have to implement a more sophisticated algorithm because creating a big dataset

BHAGYASHRI TIJARE

VIDYA MULMULE

8P a g e



including more types of news articles with more class Variables (labels) will help raising the accuracy score.



Conroy N. J., Rubin, V. L., & Chen, Y. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), 1-4. (2015).

Wu Liang, and Huan Liu. "Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate." (2018).

Granik Mykhailo, and Volodymyr Mesyura. "Fake news detection using naive Bayes classifier."

Electrical and Computer Engineering (UKRCON), 2017 IEEE First Ukraine Conference on. IEEE, (2017).

Buntain Cody, and Jennifer Golbeck. "Automatically Identifying Fake News in Popular Twitter

Threads." Smart Cloud (Smart Cloud), 2017 IEEE International Conference on. IEEE, (2017).

Bhatt Gaurav, et al. "Combining Neural, Statistical and External Features for Fake News Stance

Identification." Companion of the Web Conference 2018 on the Web Conference (2018).

International World Wide Web Conferences Steering Committee, (2018).

Abbasi Ahmed, Fatemeh Zahedi, and Siddharth Kaza. "Detecting fake medical web sites using recursive trust labelling." ACM Transactions on Information Systems (TOIS)30.4 (2012): 22

Kaliya Rohit Kumar "Fake News Detection Using A Deep Neural Network " Computing Communication and Automation (ICCCA)2018 4th International Conference on (2018).

Shu Kai, et al. "Fake news detection on social media: A data mining perspective." ACM SIGKDD Explorations Newsletter 19.1 (2017): 22-36

BHAGYASHRI TIJARE

VIDYA MULMULE

9Page