

Source Based Fake News Classification using Machine Learning

Avinash Bharadwaj¹, Brinda Ashar², Parshva Barbhaya³, Ruchi Bhatia⁴, Prof. Zaheed Shaikh⁵

U.G. Student, Department of Computer Engineering, K. J. Somaiya College of Engineering, Mumbai,
Maharashtra, India^{1,2,3,4}

Assistant Professor, Department of Computer Engineering, K. J. Somaiya College of Engineering, Mumbai,
Maharashtra, India⁵

ABSTRACT: The spread of misinformation on social media platforms is an ever-growing problem. Post the 2016 US Presidential elections, the rise in the propagation of fake news has been quite troubling. Organizations, politicians, individuals looking for personal gain, and even certain news media outlets engage in propagating fake news to sway people's decisions as well as distorting events to fit a bias or prejudice. The degree of authenticity of the news posted online cannot be definitively measured, since the manual classification of news is tedious and time-consuming, and is also subject to bias. To tackle the growing problem, detection, classification, and mitigation tools are a need of the hour. We present in this paper detailed and comprehensive steps in using Machine Learning models in tandem with Feature Engineering and Extraction processes on a fake news dataset to classify a given news article. The dataset contains attributes like the source of the articles, authors, title, text, etc We will finally show the difference in results gathered, as well as a practical application of the project undertaken.

KEYWORDS: Feature engineering, Feature Extraction, Fake news, Fake News classification, Machine Learning

I. INTRODUCTION

Fake news is a type of yellow journalism, which consists of unethical practices to catch the attention of readers. It is often spread via traditional media (such as newspapers) or by posting online. The misinformation passed, even though marked fake, usually finds its way to thousands and lakhs of people through social media. It is commonly written to mislead the reader, in order to damage the reputation or image of a person, entity, company, or a product.

Fake news is most commonly detected by headlines in large font, arrogant language, lavish use of pictures, use of edited and non-existing pictures, and so on. The frequency of fake news has been observed to increase in the political hemisphere, especially during the time of elections. Since it is easy to access an online platform with more than one lakh people, users take advantage to induce public polarization and create hostile situations by publishing misinformation to gain sympathy and, in the case of politics, votes.

Spreading of misinformation has drastic repercussions, and hence, maximum efforts should be taken to minimise its influence on civilians, as well as to eradicate the sources of these fake news. Our primary goal is to break down the different facets of Machine Learning used for the classification of fake articles as well as arrive at a practical approach to classify real time data. This might not decrease the scale at which falsities spread, but can act as a medium to double-check the authenticity of any suspicious news posted online.

The project and the paper cover the implementation for predicting the accuracy of a news article, and also the scope for future work. Thus, the paper consists of the research, implementation, the result of each approach and potential future advancements in the following sections:

- Section II: Related work - This section presents the technical papers reviewed as a part of a literature survey, which have played an important role in deepening our understanding.
- Section III: Methodology - In this section, the architecture of the system designed is explained, along with the requirement specifications and related diagrams.
- Section IV: Experimental results - This section includes the information and pictures of the implemented tools, the data processing and extracting techniques, and the explanation for the Machine Learning models used.
- Section V: Conclusion - This section concludes the idea behind the project along with indication toward future scope of the same.

II. RELATED WORK

H. Allcott and M. Gentzkow [1] primarily focused on highlighting what fake news is and how the content can be relayed from one user to another without any third-party filtering or fact checking. Fake news spreads at the speed of wildfire and manages to fool credible news sources at times. The birth of fake news stems from the fact that it is cheaper to provide than precise facts and consumers may enjoy partisan news. The definition and usage of fake news has evolved over time and got a surge in the term's popularity during the 2016 US Presidential elections. [2] mentions that while it was previously defined as sensational or false information that was disseminated under the guise of news reporting, it became synonymous with the spread of false information. This involves fabricated content, imposter content, manipulated as well as false context that could manipulate and mislead readers. Our thorough research also brought our attention towards satire, hate speech as well as bias. K. Sharma et al [2] in their paper have provided a proper breakdown on what exactly constitutes fake news, excluding the likes of satire and rumor, also categorizing methods of identification, viz, 1) Content based identification 2) Feedback based identification 3) Intervention based solutions.

In [9], Saad et al went about with the detection of fake news using N-gram analysis and Deep Learning models. Prior to feature engineering, they brought about the importance of preprocessing techniques for the data, which needed to be subjected to certain refinements like stop-word removal, lowercasing, tokenization, sentence segmentation and punctuation removal. Thus, for the machine learning models to be employed, stop word removal, sentence breakdown structures and stemming [8] with recurrent neural networks were observed as well. Saad et al [9] also sheds light onto the TF-IDF (Term Frequency - Inverse Document Frequency) : a weighting metric often used in information retrieval and natural language processing. It is a statistical metric used to measure how important a term is to a document in a dataset. Hence, these text preprocessing techniques, along with various classification models like Naive Bayes, SVM, Random Forest classifier and so on are applied to classify a given article into eight categories, i.e bs, bias, conspiracy, hate, state, satire, junksci and fake.

III. METHODOLOGY

The first step involved procuring datasets pertaining to the problem statement. The prominent datasets researched and shortlisted include: LIAR, Politifact, KaggleFN and NewsFN. Among the datasets shortlisted, KaggleFN seemed the most promising for preprocessing, feature extraction, and model classification. The reason is due to the fact that all the other datasets lacked the sources from where the article/statement text was produced and published from. Citing the sources for article text is crucial to check the trustworthiness of the news and further helps in labeling the data as fake or untrustworthy. Features to be considered:

- Author
- Published date
- Language
- Site_URL
- Image_URL
- Type
- Label
- Title (generated attribute without stop words after preprocessing)
- Text (generated attribute without stop words after preprocessing)
- hasImage (generated attribute to check if the source contains images)

The dataset has 8 labels by which the articles in the corpus are classified with. These are: bias, bs, hate, state, conspiracy, satire, junksci, fake.



```
In [3]: print(dataset3['type'].value_counts())
bs      600
conspiracy  430
bias     389
hate     245
satire   146
state    121
junksci  102
fake     15
Name: type, dtype: int64
```

Fig. 1: Distribution of type values after eliminating considerable skew in kaggleFN

The categories, bs (i.e. bullshit), junksci(i.e. junk science), hate, fake, conspiracy, bias, satire and state declare the category under which untrustworthy or false news fall under. The first step, which was text preprocessing was performed using the following:

1) Taking care of null/missing values 2) Transforming categorical data with the help of label encoders 3) Uppercase to lowercase 4) Number removal 5) Tokenization 6) Stop Word Removal, Stemming and Lemmatization (with POS tagging) using the Natural Language Toolkit Library. Prior to preprocessing, a considerable amount of skew towards the label, ‘bs’, was removed from the dataset. For feature engineering, the TF-IDF technique was used, while for feature extraction, the techniques Word2Vec and GloVe were used. This processed and embedded text is provided as an input to Machine learning models, where the data is made to fit the model, to get a prediction as an output. The models implemented include: 1) Support Vector Machine 2) Naive Bayes 3) Random Forest 4) Logistic Regression 5) AdaBoost 6) Decision Tree 7) Neural Networks.

IV. EXPERIMENTAL RESULTS

Using the dataset discussed above, we trained our model by preprocessing the data and applying the feature extraction and engineering techniques discussed above and finally, used the transformed data as input to the machine learning models . The accuracy results of classification are as follows:

TF-IDF

As explained by [9] Saad et al, right after preprocessing, the output is a corpus of raw texts that are stripped of stop words, stemmed and lemmatized. In order to get a sparse matrix of TF/IDF values, the following steps are taken:

- Tokenization of texts
- Counting of the tokens and
- Transforming the raw tokens into TF/IDF values.

The above steps are done with the help of the TfidfVectorizer, which transforms text to feature vectors that can be used as input to estimators/classifiers.

Train-test split ML Models used	60:40	70:30	80:20
AdaBoost	95.6 %	96.91 %	95.36 %
Random Forest	84.6 %	81.0 %	82.0 %
SVM (linear)	67.4 %	67.9 %	68.0 %
Logistic Regression	72.9 %	75.0 %	77.0 %
Naive Bayes	43.3 %	41.0 %	46.0 %
Neural Network	55.0 %	54.3 %	30.0 %

Table 1: Results of using TF-IDF with Machine learning models



For understanding the concepts of the feature extraction techniques Word2Vec and GloVe, the paper by M. Naili et al [10] was referred.

Word2Vec

The word embedding according to the document or text is created by checking the word from the input document with an embedding of all words of the English language already available on GoogleNews. This provides a matrix of vectors as an output that can be provided to Machine Learning models as an input.

Train-test split ML model used	80:20	70:30	60:40
Adaboost	24%	28%	25%
Random Forest	47%	51%	46%
SVM (linear)	40%	38%	39%
Logistic Regression	45%	43%	45%

Table 2: Results of using Word2Vec with Machine learning models

GloVe

The first step carried out was importing the pre-trained word vectors publicly available on Stanford NLP. After iterating through the corpus, word indices of all words were created. Then an embeddings matrix was created for words present in the training set of our dataset using the pre-trained word vectors. Words were accessed with the help of the word indices.

Neural Network

Train-test split	80:20	70:30	60:40
Neural Network	32.44%	32.08%	28.82%

Table 3: Results of using GloVe with a Neural network

ML models

Train-test split ML Models	80:20	70:30	60:40
AdaBoost	29.5%	21.14%	23.1%
Random Forest	48.4%	47.1%	45.9%
SVM (linear)	42.5%	42.6%	41.8%
Logistic Regression	45.1%	43.1%	41.6%
Decision Tree	39.6%	37.8%	36.5%

Table 4: Results of using GloVe with Machine learning models

As seen from the results above, TF-IDF performs the best with the ML models compared to GloVe and Word2Vec.

As a part of implementing the above methodology, a Web portal was created by pickling the models.

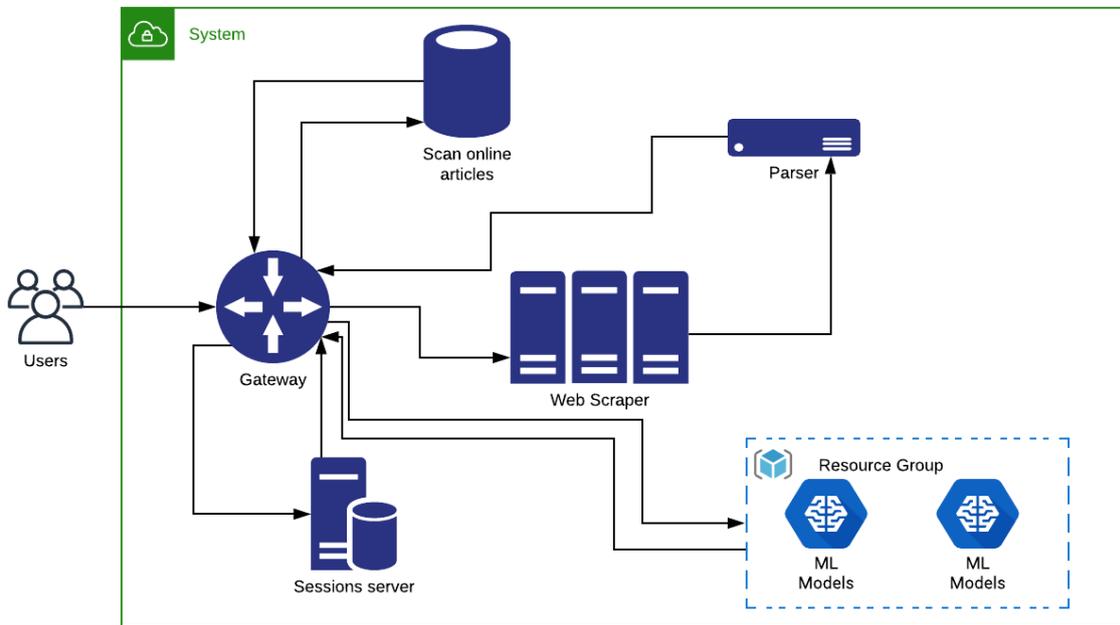


Fig. 2: System Architecture of the tool for Fake News Classification

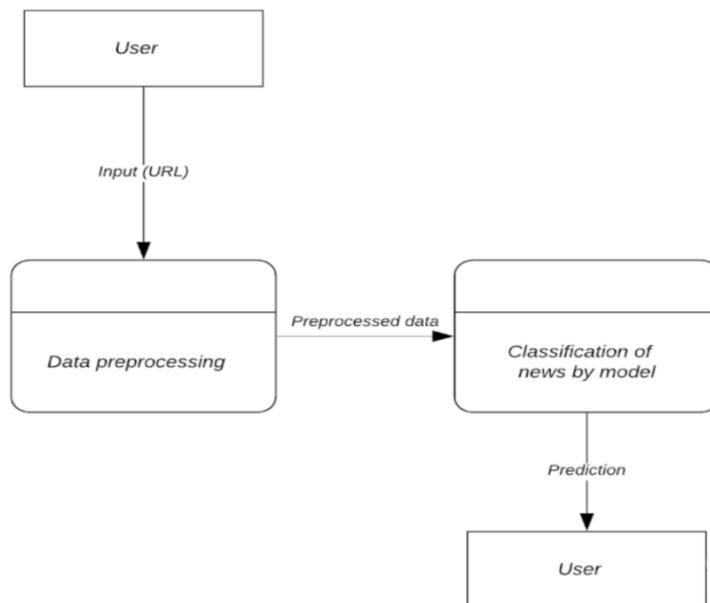


Fig. 3: Data Flow Diagram



☰
Fake News Tool Prototype
LOGOUT

Enter URL

<https://www.frontpagemag.com/point/2020/06/disney-purge-splash-mountain-daniel-greenfield/>

SUBMIT
VIEW SOURCES

Article Text:
 After revamping the Pirates of the Caribbean, Splash Mountain was bound to get purged by the most politically correct entertainment monopoly on the block The Walt Disney Company has announced plans to remodel the Splash Mountain attraction at Disneyland and Walt Disney World. The reimagined ride would replace scenes from the 1946 film The Song of the South with characters from the 2009 film The Princess and the Frog. Sometime soon, riders will see Princess Tiana, Louis the Alligator, and Dr. Facilier instead of Br'er Rabbit and Br'er Bear before their floating log falls down the flume. Hilarious Disney is going to swap in characters from an animated movie that nobody saw and nobody cares about, and won't be able to recognize. They could have just stuck Marvel characters over it. Or tied it in to Frozen, which would be the smart money. But instead Disney is trying to racially pander with The Princess and the Frog, while setting itself up for a fall. The otherwise obscure animated movie still has at least as many offended social justice critics as Song of the South for its black stereotypes and visual images. It won't be that long before there will be protests and Splash Mountain will have to be purged yet again.

Model Name	Result
adaboost	bs
logistic	satire
naive_bayes	bs
neural_network	bias
random_forest	bs
svm	bs

Keywords:
 purged characters disney purge wont walt princess song south mountain splash

Fig. 4: Working of prototype on real-time data. As seen above, the different models have classified the above article and it's extracted contents.

The tool built can be used to carry out the following functions:

1. Scrape the contents of article for a given url
2. Extract useful information about the article
3. Using the machine learning models developed, try and classify the article. The user can see the model used as well as the result after classification.

We have developed a Chrome extension that works on similar lines which makes it easier to classify and categorize news. It also saves the current verdict of the model which is easily accessible to other users.

Fig. 5: Chrome extension - For articles that have been previously classified

Some of the major challenges faced in this endeavour include:

- Lack of comprehensive datasets with news sources for people to work on and understand the subject better
- Preventing skew in the datasets that would lead to overfitting/underfitting
- Understanding how to grab relevant data as well as how to clean and pre-process data for classification

V. CONCLUSION

The project undertaken provides a comprehensive and detailed approach to the machine learning paradigm towards the topic chosen. Not only the different facets of fake news but also how to identify its sources, how it propagates and what are the possible techniques one can adopt for mitigation was understood. This project also provides a broad understanding of how Machine Learning can be used as a tool to combat fake news. For the uninitiated, this project covers everything, from dataset accumulation, to preprocessing, feature extraction and engineering and finally building accurate classifiers. With the aforementioned implementation, it is possible for people to make practical and rudimentary tools, like the web prototype as well as the Chrome extension to inform users about the articles they read and thus, help them make better decisions.

VI. FUTURE SCOPE.

Like KaggleFN (Getting Real With Fake News), there is a dire need for such comprehensive datasets that require collaboration, curation and crowdsourcing. In an era where fake WhatsApp forwards and Tweets are capable of influencing naive minds, such tools and knowledge have to be put to practical use in not only mitigating the spread of misinformation but also to inform people about the type of news they consume. Apart from this notion, it is also concluded that the next steps require work pertaining to semantic and syntactic analysis of the text as well as Natural Language Processing methods that work with Deep Learning models. The research related to this project even concluded that promising work has been done on Propagation Trees and Neural Networks, and as such, seems to be the path forward. Development of practical applications for users to gain insight from the articles they consume, fact-checking websites, built-in plugins and article parsers can further be refined, made easier to access, and more importantly, should create more awareness.

REFERENCES

- [1] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, vol. 31, no. 2, pg. 211–236, Spring 2017.
- [2] K. Sharma, F. Qian, H. Jiang and N. Ruchansky, "Combating Fake News: A Survey on Identification and Mitigation Techniques," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 3, Apr. 2019, Art. no. 21.
- [3] Y. Chen, N. J. Conroy and V. L. Rubin, "Misleading Online Content: Recognizing Clickbait as "False News"," presented at the ACM WMDD'15, Seattle, Washington, USA, Nov. 9, 2015.
- [4] Y. Yang, L. Zheng, J. Zhang and Q. Cui, "TI-CNN: Convolutional Neural Networks for Fake News Detection," Jun 29, 2019. Available: arXiv:1907.00181v1 [cs.CL].
- [5] L. Dyson and A. Golab, "Exploring the Application of NLP Methods to Machine Identification of Misleading News Sources," Lauren Dyson & Alden Golab, CAPP 30255: Advanced Machine Learning for Public Policy, University of Chicago, Winter 2017
- [6] M. Granik and V. Mesyura, "Fake News Detection using Naïve Bayes Classifier," presented at the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2017.
- [7] S. Girgis, E. Amer and M. Gadallah, "Deep Learning Algorithms for Detecting Fake News in Online Text," presented at the 2018 13th international conference on computer engineering and systems (ICCES), 2018.
- [8] Yang-Liu and Yi-Fang Brook Wu, "Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks," presented at the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), 2018.
- [9] H. Ahmed, I. Traore and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques," presented at –the International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments., Springer, 2017, pp. 127–138.
- [10] M. Naili, A. Habacha Chaibi and H. Hajjami Ben Ghezala, "Comparative study of word embedding methods in topic segmentation," presented at the International Conference on Knowledge Based and Intelligent Information and Engineering Systems, Marseille, France, Sep. 6-8, 2017.