

Fake News Detection System Using Machine Learning

Abstract

By

PRONOY BOSH

ENROLMENT NO : 12018002002231

&

MD IMRAN HOSSAIN

ENROLMENT NO : 12018002002216

Under the Supervision of

PROF. NILANJANA DUTTA ROY

Department of Computer Science and Engineering

Institute of Engineering and Management

West Bengal, India

December, 2020

Abstract

The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in human history before. With the current usage of social media platforms, consumers are creating and sharing more information than ever before, some of which are misleading with no relevance to reality. Automated classification of a text article as misinformation or disinformation is a challenging task. Even an expert in a particular domain has to explore multiple aspects before giving a verdict on the truthfulness of an article.

In this work, we propose to use a machine learning ensemble approach for the automated classification of news articles. Our study explores different textual properties that can be used to distinguish fake content from real. By using those properties, we train a combination of different machine learning algorithms using various ensemble methods and evaluate their performance on real-world datasets. The experimental evaluation confirms the superior performance of our proposed ensemble learner approach in comparison to individual learners.

Introduction:

The Internet has become compulsory in our life. It is now very easy to access the internet than it was before. There is no doubt that many young people prefer the internet to get their news rather than the newspaper, radio, etc. The internet provides many opportunities for us, we can search for anything on the internet to clear our doubts and for research purposes also. Simply saying, we can't even imagine our life without the internet. As more people are connecting to the internet, they get most of their information content through it. In a country like India where internet access has become cheap recently, a lot of people are accessing news through their digital devices. But when it comes to news publishing it creates so many issues. If it is about the news, the internet plays a very important role because through the internet, the news widespread very fast. There are so many consequences of fake news, it can cause harm to innocent people. Fake news may be made intentionally or accidentally to give harm to an individual or a group for any purposes, such as for political issues, religious purposes, and so on.

The BBC news broadcaster has done research on Indian general election during 2014. The researchers viewed about 16000 and 3000 accounts and pages from Twitter and Facebook respectively to learn how fake news gets polarized in India. Another research by the BBC resulted that nearly 72% of Indian citizens are not able to differentiate between real facts from made-up

ones. Altogether, these conclude that this project can help people in India to identify and expose the scammers to digital literacy to overcome the consequences of fake news in the country.

Materials and Methodology:

Tools needed for this project:

- Python Language
- News.csv dataset, Kaggle dataset etc.
- TfidfVectorizer
- Passive Aggressive Classifier
- pandas
- NumPy
- scikit-learn

By using those tools and the machine learning algorithms such as Logistic Regression, Support Vector Machine, K-Nearest Neighbors (KNN) etc., we can do the pre-processing on the datasets such as all the NULL valued rows are dropped. The textually based columns from the datasets are first n-gram converted and then into TFIDF vectors. These vectors are then used in finding out the Cosine similarity between each sample. And we also use the Ensemble Technique, Ensemble machine learning refers to a technique that integrates output from multiple learners and is applied to a dataset to make a prediction.

Results & Discussion:

From some tests and example, we can mention that overall, the best performing algorithm is bagging classifier (decision trees) (accuracy 94%), whereas the worst performing algorithm is Wang-Bi-LSTM (accuracy 64.25%). Individual learners' accuracy is 77.6% whereas the accuracy of ensemble learners is 92.25% in precision, recall and F-1 score. The ensemble learner XGBoost performed better in comparison to other learning models on all performance metrics. Logistic regression is a relatively simpler model but achieved an average accuracy of over 90%.

However, accuracy score alone is not a good measure to evaluate the performance of a model; therefore, we can also evaluate performance of learning models on the basis of recall, precision, and F1-score.

Conclusion:

In this research, we present the task of automatic detection of fake news using machine learning models and ensemble techniques. We have used some new publicly available fake news datasets. The primary aim of the research is to identify patterns in text that differentiate fake articles from true news. The classification of fake news from the real news is very crucial task nowadays. It is becoming an imminent threat in some situation to be not able to discern real and fake news. Our best performing models achieved accuracies that are comparable to the human ability to spot fake content.

References

1. A. Douglas, "News consumption and the new electronic media," *The International Journal of Press/Politics*, vol. 11, no. 1, pp. 29–52, 2006.
2. D. M. J. Lazer, M. A. Baum, Y. Benkler et al., "The science of fake news," *Science*, vol. 359, no. 6380, pp. 1094–1096, 2018.
3. A. D. Holan, 2016 Lie of the Year: Fake News, PolitiFact, Washington, DC, USA, 2016.
4. <https://www.engadget.com/2018/12/10/more-people-get-news-from-social-media-than-newspapers/>
5. A. Robb, "Anatomy of a fake news scandal," *Rolling Stone*, vol. 1301, pp. 28–33, 2017.
6. <https://scroll.in/latest/838693/fake-news-is-a-concern-for-83-of-indian-media-consumers-reveals-bbc-study>
7. S. Kogan, T. J. Moskowitz, and M. Niessner, "Fake News: Evidence from Financial Markets," 2019, <https://ssrn.com/abstract=3237763>.
8. <https://qz.com/india/1459818/bbc-study-shows-how-indians-spread-fake-news-on-whatsapp/>
9. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.
10. V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," 2017, <https://arxiv.org/abs/1708.07104>.

SIGNATURE OF STUDENTS:

1. Pronoy Bosh

2. Md Imran Hossain

SIGNATURE OF MENTOR:

•

DATE : 31 December, 2020