Literature Review on Fake News Detection

Introduction

[Provide an overview of fake news and its impact, as well as the importance of detecting fake news.]

Previous Work

1. [Author(s), Year, Title, Journal/Conference]

[Provide a summary of the paper's main contributions, including the data sets used, the methods applied, and the results achieved. Evaluate the paper's strengths and weaknesses.]

1. [Marco L. Della Vedova et.al,2018, Automatic Online Fake News Detection Combining Content and Social Signals,PROCEEDING OF THE 22ND CONFERENCE OF FRUCT ASSOCIATION]

The paper presents a novel fake news detection approach which, by combining content-based and social-based methods, outperforms existing approaches in the literature. Combining content-based and social-based approaches for prediction and classification tasks is a solution that has been successfully applied in other fields: in the recommender systems field, for example, the so-called hybrid recommender systems are used to overcome the limitation that collaborative filtering (i.e. social-based) methods face when an item has zero ratings [25], a situation also known as cold-start: in those cases, an additional technique based on the analysis of the item's content is combined with the collaborative filtering approach to mitigate the cold-start problem. As previously highlighted, the problem we face with fake news detection is quite similar, hence the idea of combining context-based and social-based methods to provide automatic detection tools that can work without (or with limited) social signals. This can make easier the task of early detection of fake news, that, in turns, can limit the spread of fake news as a whole.

[Amy Sliva et.al, 2017, Fake News Detection on Social Media:

A Data Mining Perspective, ACM SIGKDD Explorations Newsletter]

In this survey, we present a comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets. We also discuss related research areas, open problems, and future research directions for fake news detection on social media.

1. I. Manzoor, J. Singla and Nikita, "Fake News Detection Using Machine Learning approaches: A systematic Review," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 230-234, doi: 10.1109/ICOEI.2019.8862770.]

This paper makes an analysis of the research related to fake news detection and explores the traditional machine learning models to choose the best, in order to create a model of a product with supervised

machine learning algorithm, that can classify fake news as true or false, by using tools like python scikit-learn, NLP for textual analysis. This process will result in feature extraction and vectorization; we propose using Python scikit-learn library to perform tokenization and feature

extraction of text data, because this library contains useful tools like Count Vectorizer and Tiff Vectorizer. Then, we will perform feature selection methods, to experiment and choose the best fit features to obtain the highest precision, according to confusion matrix results.

[Garima Rawat et.al, 2023,Fake News Detection Using Machine Learning, International Conference on Artificial Intelligence and Smart Communication (AISC)]

In order to assess and compile research on the use of machine learning techniques to identify fake news, this chapter will undertake a comprehensive mapping analysis. False propaganda on social media and other platforms is widespread which is a cause of great worry because it can wreak widespread social and national damage with devastating effects. On figuring things, there has already been a lot of research. In order to model a product using supervised machine learning algorithms that can categories fake news as genuine or false using the necessary tools, a survey of the literature on fake news detection is presented in the article. Classic machine learning models are also explored. Feature extraction and vectorization are the results of this technique. We advise using this package to tokenize and extract functions from text input in Python because it has useful tools like the count vectorizer and tiff vectorizer. Then, using feature selection approaches, we investigate and select the most appropriate features to obtain the best accuracy based on the Confusion Matrix results.

[Jiawei Zhang et.al, 2020, FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network, IEEE 36th International Conference on Data Engineering (ICDE)]

This paper introduces a novel gated graph neural network, namely FAKEDETECTOR. Based on a set of explicit and latent features extracted from the textual information, FAKEDETECTOR builds a deep diffusive network model to learn the representations of news articles, creators and subjects simultaneously. Extensive experiments have been done on a real-world fake news dataset to compare FAKEDETECTOR with several state-of-the-art models, and the experimental results are provided in the full-version of this paper at [13].

1. Shaikh et.al, 2020, Fake News Detection using Machine Learning,IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)]

In this model, we used classification techniques such as Support Vector Machine (SVM), Naive Bayes, Passive attack classifier, etc. Using feature extraction technique as term Frequency Inversion Document Frequency (TF-IDF) and Support Vector Machine (SVM) as the classifier model output, the accuracy is 95.05%. The main motivation behind our work is to find the best classification algorithm that detects fake news and calculates its accuracy. We have studied different classification algorithms and in our model we have used SVM, passive attack classifier, Naive Bayes. Among the three, SVM provides the highest accuracy, but the time required by SVM is higher compared to passive-active, naive Bayes.

[Iftikhar Ahmad et.al, 2020, Fake News Detection Using Machine Learning Ensemble Methods, Hindawi Complexity Volume 2020]

Novel to this research, various ensemble techniques such as bagging, boosting, and voting classifier are explored to evaluate the performance over the multiple datasets. We used two different voting classifiers composed of three learning models: the first voting classifier is an ensemble of logistic regression, random forest, and KNN, whereas the second voting classifier consists of logistic regression, linear SVM, and classification and regression trees (CART). The criteria used for training the voting classifiers is to train individual models with the best parameters and then test the model based on the selection of the output label on the basis of major votes by all three models. We have trained a bagging ensemble consisting of 100 decision trees, whereas two boosting ensemble algorithms are used, XGBoost and AdaBoost. A k-fold (k = 10) cross validation model is employed for all ensemble learners.