Automatic Fake News Detector in Social Media Using Machine Learning and Natural Language Processing Approaches

J. Srinivas, K. Venkata Subba Reddy, G. J. Sunny Deol, and P. VaraPrasada Rao

Abstract The definition of fake news is a cooked-up story with an objective to fool or to cheat people. The current research aims to detect fake news in social media like Twitter, Watsapp and Facebook by studying the responses of the proposed model on posts acquired from Reddit online news store. Automatic fake news detection is a complex activity as it involves the model to implement natural language processing concepts in-tandem with machine learning approaches. Two feature extraction algorithms, namely CountVectoriser (CV) and term frequency-inverse document frequency (TFIDF), were employed separately for extracting the most relevant features from the dataset. These features were fed to multinomial naive Bayes (MNB), random forest (RF), support vector classifier (SVC) and logistic regression (LR) classifiers for classifying fake news creating a total of eight classification models. A solitary CV-based model was considered as the baseline model for predicting fake news in r/theonion and r/nottheonion datasets. GridsearchCV was also implemented for finding the testing and training scores for the selected parameters. Out of these models, TFIDF with MNB achieved an accuracy of 79.05% and is considered as the best.

1 Introduction

According to a survey conducted by Gartner [1], by 2022 the volume of fake news circulated in the society will be more than genuine news. This can be attributed to the fact of the exponential growth of social media users. Fake news is generally defined as synthesised news containing misinformation, rumour and falsified facts

J. Srinivas (✉)
SR University, Warangal, India

K. Venkata Subba Reddy · G. J. Sunny Deol
Kallam Haranadha Reddy Institute of Technology, Guntur, India

P. VaraPrasada Rao
Gokaraju Rangaraju Institute of Technology, Hyderabad, India

S. C. Satapathy et al. (eds.), *Smart Computing Techniques and Applications*,
Smart Innovation, Systems and Technologies 224,
https://doi.org/10.1007/978-981-16-1502-3_30