



**INFORMATICS
INSTITUTE OF
TECHNOLOGY**

**UNIVERSITY OF
WESTMINSTER**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER (UOW)

Faculty of Science and Technology

MSc Advanced Software Engineering

Fake News Detection Based on Cross-Verification and Publisher Reputation

A dissertation by

Mr. Asanka Herath

Supervised by

Mr. Saman Hettiarachchi

Student Name	Student ID	UOW ID
Herath Mudiyansele Asanka Sanjaya	2018492	17426729/1

Abstract

With the growth of social media usage over the last few years, it is known that the manipulated news content spreading has increased and it badly affects public opinion and it has become a social problem as well (Wang et al., 2018). Most of such false news is first getting initialized on social media platforms such as Twitter and Facebook (Traylor et al., 2019). Social media allows a user to post anything without having any accountability. False news has been published by different individuals and parties to gain different advantages such as political advantage.

Identifying and filtering or blocking such news content is a challenging task even for a human operator since it is hard to know which is true and which is false. Different algorithms and methods have been proposed to identify false news content. Most of the algorithms are based on classifying the news based on a past data set. This research aims to find a method that can identify fake news by cross-checking with other news sources and the news source reputation.

The solution proposed and developed in this project is an approach based on news cross-comparing with credible sources to figure out whether the news is true or fake. A domain reputation-based approach was used in order to filter out the credible sources among a list of news publishers and a model trained using Doc2Vec vector is used to compare news sentences with each other. An opposite sentence detection algorithm also proposed. The main limitation in this approach is that it doesn't validate the news article body content. Validation of the body content would lead to more accurate results while accurate summarization of body content is still a challenge.

Keywords: NLP, Fake news, Sentence similarity detection, Domain reputation, Doc2Vec