



Covid-19 Fake News Detection on Twitter Based on Author Credibility Using Information Gain and KNN Methods

Nanda Ihwani Saputri¹, Yuliant Sibaroni², Sri Suryani Prasetyowati³

^{1,2,3}Informatika, Fakultas Informatika, Universitas Telkom

¹nandaihwani@student.telkomuniversity.ac.id, ²yuliant@telkomuniversity.ac.id, ³srisuryani@telkomuniversity.ac.id

Abstract

Twitter is one of the social media that is used as a tool to share various kinds of information about various kinds of things that are of concern to social media users. One of the information shared is information about COVID-19, which is known that the COVID-19 pandemic is currently spreading throughout the world at a very alarming rate. COVID-19 is an infectious disease caused by SARS-COV-2. The World Health Organization (WHO) claims that the spread of COVID-19 is supported by the spread of false/fake news. So to find out the truth of the news, a COVID-19 fake news detector is needed so that users don't fall for the hoaxes circulating. This study aims to classify COVID-19 news on Twitter based on author credibility. Credibility in question is a person's perception of the validity of information and is a multidimensional concept that is used as a means of receiving information to assess the source of communication. The method used in this research is Information Gain and KNN. KNN (K-Nearest Neighbor) is a supervised learning algorithm that works by classifying a set of data based on classified training data. Information Gain is used to ranking the most influential attributes, and KNN is used to classify data based on learning data taken from the nearest neighbors. The research consists of 6 main stages, namely data collection (crawling data), data preprocessing, feature extraction, feature selection, data split into training data and testing data, KNN stage, and data evaluation stage. The research carried out succeeded in obtaining an accuracy value of 91%, a correlation value between credibility and hoax of 0.115, and a p-value <0.005.

Keywords: twitter; fake news, COVID-19; credibility; KNN; information gain

1. Introduction

The development of communication technology helps humans in sending and receiving information. Virtual communities are starting to form and shift traditional communities. The virtual community that we can meet is social media. Distribution of information in the form of online conversations can be done through social media. The most obvious participation and use of social media can be seen in social media such as Facebook and Twitter [1]. Based on a survey conducted by the Association of Indonesian Internet Service Providers, shows that internet penetration in Indonesia in 2020 will reach 171.17 million people, equivalent to 64.8% of Indonesia's total population [2]. Meanwhile, when a similar survey was conducted in 2017, 143 million internet users were found, equivalent to 54.7% of Indonesia's total population. This shows that there was an increase in the number of internet users from the previous year of 10.1%. This increase in internet access is in line with access to social media, both Facebook and Twitter.

Since its launch in 2006, Twitter has become one of the most popular social media platforms for sharing information, both personal information and a means of interaction in various parts of the world [3]. Dissemination of information on Twitter is done through making tweets. One of the information being disseminated is information about COVID-19, which is known that the COVID-19 pandemic is currently spreading throughout the world at a very alarming rate [4]. The World Health Organization (WHO) even claims that the spread of COVID-19 is supported by the spread of false/fake news [5].

Fake news about COVID-19 seems to spread very quickly on social media [6]. Similar trends have been seen during other epidemics, such as the Ebola, yellow fever, and Zika outbreaks [4]. This is a very worrying development because even a little exposure to fake news can cause public anxiety and distrust [5]. In addition, it is necessary to identify the creator or subject of fake news which will help eradicate a large number of fake news from its origin [7]. Generally, for news spreaders, in addition to the tweets that are made,