

## ABSTRAK

Nama : Yeppy Mangun Puspitajudin  
NIM : 41518010056  
Pembimbing TA : Ida Nurhaida, Dr, MT.  
Judul : Tweets Analysis of Hoax Detection Using  
Levenshtein Distance and Text Classification  
Methods

Hoaks memanipulasi persepsi manusia dengan memberikan informasi yang salah, tetapi dianggap sebagai kebenaran yang akan menimbulkan keresahan atau perselisihan di antara masyarakat. Apalagi jika ada sebuah tren, peristiwa ataupun permasalahan, seperti pemilihan presiden, pandemi, isu politik dan lain-lain. Kementerian Komunikasi dan Informatika (Kominfo) di Indonesia mencatat hingga 8 Agustus 2020, 1.028 hoaks tersebar di berbagai platform media sosial dan menurut riset DailySocial.id, 44% masyarakat Indonesia belum bisa mendeteksi berita hoaks. Pada penelitian ini penulis mengangkat topik Machine Learning yang berfokus pada pendeteksian penyebaran informasi palsu (hoaks) pada Twitter dengan mengimplementasikan metode Levenshtein Distance (LD) dengan kombinasi Text Classification seperti Support Vector Machine (SVM), dan Stochastic Gradient Descent (SGD). Perhitungan TF-IDF akan digunakan untuk mengukur bobot sebuah kata dari tweet yang digunakan sebagai kosakata dan menggunakan metode LD untuk mengukur jumlah perbedaan yang dimiliki pada setiap teks yang diproses, serta SVM dan SGD untuk melakukan klasifikasi teks sehingga hasil akhir berupa persentase apakah teks tersebut dianggap hoax atau tidak. Penelitian ini mengangkat 3 topik khusus untuk mendeteksi hoaks yaitu, pandemi covid19, politik, dan ekonomi di Indonesia.

**Kata Kunci:** Hoaks, Tweet Analysis, Hoax Detection, Levenshtein Distance, Text Classification.

## ABSTRACT

Name : Yeppy Mangun Puspitajudin  
Student Number : 41518010056  
Supervisor : Ida Nurhaida, Dr, MT.  
Title : Tweets Analysis of Hoax Detection Using  
Levenshtein Distance and Text Classification  
Methods

Hoax misled human perceptions with wrong information but considered truth which will cause anxiety or discord in society. Especially if there are trends or events, such as presidential elections, pandemics, political issues, and others. The Ministry of Communication and Informatics (Kominfo) in Indonesia noted that up to August 8, 2020, 1,028 hoaxes were spread on various social media platforms and according to DailySocial.id research, 44% of Indonesians have not been able to detect hoax news. In this research, the author raises the topic of Machine Learning which focuses on detecting the spread of false information (hoaxes) on Tweets by implementing the Levenshtein Distance (LD) method with the combination of Text Classification such as Support Vector Machine (SVM), and Stochastic Gradient Descent (SGD). TF-IDF calculations will be used to measure the weight of a word from the tweets which is used as a vocabulary and use the LD method to measure the number of differences that are owned in each text being processed, also SVM and SGD to perform text classification so that the final result is a percentage whether the text considered as a hoax or not. These raise 3 topics to focuses on detecting hoaxes related to the covid19 pandemic, politics, and economics in Indonesia.

### **Keywords:**

Hoax, Tweet Analysis, Hoax Detection, Levenshtein Distance, Text Classification.