



A Comparative Analysis of Text Vectorization and Machine Learning Classifiers for Fake News Detection

¹ Dr. Ashutosh Dhamija, ²Dr. Mukesh Kumar

¹Associate Professor, Department of Electronics & Communication Engineering, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana-133207, India.

ashutosh.dhamija@mmumullana.org

²Associate Professor, Department of Electronics & Communication Engineering, The Technological Institute of Textile and Sciences, Bhiwani, Haryana-127021, India.

drmukeshkumar@titsbhiwani.ac.in

Abstract: In today's digital era, the media landscape has seamlessly transitioned from print to online platforms, leading to an unprecedented increase in information accessibility and exchange. However, this transformation has also intensified a major challenge, the rapid proliferation of fake news, which refers to fabricated or misleading information that can be easily produced and disseminated. This paper addresses the growing global concern of misinformation and explores potential solutions through machine learning techniques. The proposed study develops a model designed to assess the authenticity of news articles by evaluating multiple text vectorization methods, specifically the Bag-of-Words approach using both Count Vectorizer and TF-IDF Vectorizer. Two classification algorithms, namely the Multinomial Naive Bayes Classifier and the Passive Aggressive Classifier, are employed to detect fake news. The study further investigates how text pre-processing influences overall model performance. The dataset chosen for training is comprised of 67.7% curated information, while the remaining 33.3% remains untrained raw data. Notably, the model demonstrates a noteworthy efficiency rate of 93.78% under optimum conditions. This strong result demonstrates how well the suggested methodology works to differentiate between real and fake news.

Keywords: Fake News Detection, Text Vectorization, TF-IDF, Machine learning and Feature Extraction

1. Introduction

In the ever-evolving landscape of today's digital era, the media sector has seamlessly transitioned from traditional print formats to digital platforms, fundamentally reshaping the dynamics of information production and consumption. However, amidst this transformative shift, a formidable challenge has emerged—the widespread proliferation of "fake news." This