I. INTRODUCTION

The public has come to rely on the internet as a quick source of information, with YouTube being one of the platforms utilized. As of January 2024, YouTube ranks as the second most popular social media platform globally, with 2.5 billion monthly active users, following Facebook [1]. Indonesia itself is the fourth-largest contributor to YouTube's user base, with a total of 139 million users as of October 2023 [2].

This massive user base brings challenges for YouTube, one of which is the spread of clickbait content. Clickbait is defined as a term referring to headlines designed to entice (bait) users to click on content or links by using exaggerated titles [3]. According to Loewenstein, clickbait titles exploit the "curiosity gap," which refers to the gap between what someone knows and what they want to know [4]. An example of such a title is: “ANDA TIDAK AKAN PERCAYA APA YANG TERJADI SELANJUTNYA!”—a headline designed to take advantage of the curiosity gap. This strategy is often used by creators to boost virality and generate ad revenue (AdSense).

Clickbait headlines have negative impacts, both for viewers and content creators. Viewers may feel disappointed or deceived, leading to a loss of trust in the creators. Additionally, clickbait can serve as a loophole for spreading propaganda.

Given the large number of YouTube users in Indonesia, identifying clickbait content is a crucial step to reduce the appearance of such content on users' timelines [5]. Early identification of clickbait content could provide an effective solution, especially since YouTube currently relies on manual reporting from users to address clickbait content [6].

This paper contributes to developing a machine capable of detecting clickbait in Indonesian YouTube titles, a topic that remains underexplored. We employed and compared three machine learning models—Support Vector Machine (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM)—on this topic. In order to enable the machine to recognize clickbait titles, Term Frequency-Inverse Document Frequency (TF-IDF) for SVM and RF, and FastText embeddings for LSTM were used. The purpose of this study is to provide insightful information about machine learning-based clickbait content moderation on the YouTube platform.