

ABSTRAK

REINHARD HUTAGAOL, NIM 4183550004 (2023). Klasifikasi Berita Mengenai Vaksin Booster Covid-19 di Sosial Media Twitter Menggunakan *Naïve Bayes Classifier*

Perkembangan pesat teknologi internet telah mengakibatkan keterkaitannya yang erat dengan media sosial. Media massa saat ini sering menggunakan platform media sosial sebagai sarana untuk menyebarkan berita, termasuk berita palsu. Hoax adalah informasi yang sebenarnya tidak benar, namun disajikan seolah-olah benar. Hoax dibedakan dari rumor, ilmu semu, berita palsu, dan candaan. Tujuan dari penelitian ini adalah mengembangkan model machine learning dengan menggunakan metode Naive Bayes Classifier untuk mengklasifikasikan berita di platform Twitter. Model ini dibangun menggunakan dataset berita yang diperoleh dari laman resmi kominfo. Setiap berita telah dikategorikan sebagai hoax atau fakta. Totalnya, terdapat 2472 berita dalam dataset ini, dengan 1415 di antaranya merupakan berita fakta, dan 1057 berita lainnya merupakan berita hoax. Hasil evaluasi model menunjukkan akurasi sebesar 71%, presisi sebesar 51%, recall sebesar 51%, dan f1-score sebesar 51%. Setelah itu, model yang telah dikembangkan digunakan untuk mengklasifikasikan 97 berita Twitter yang belum diberi label mengenai vaksin booster. Hasilnya menunjukkan bahwa dari 97 berita tersebut, terdapat 62 berita yang mengandung fakta dan 35 berita yang merupakan hoaks.

Kata Kunci : Hoax, Klasifikasi, *Naïve Bayes Classifier*



ABSTRACT

REINHARD HUTAGAOL, NIM 4183550004 (2023). Classifying News About Covid-19 Booster Vaccine on Twitter Social Media Using *Naïve Bayes Classifier*

The Internet technology has developed so rapidly that it is inseparable from social media. Many mass media outlets utilize social media to spread news, including hoax news. A hoax refers to information that is not true but is made to appear true. Hoaxes are different from rumors, pseudoscience, fake news, or jokes. This research aims to build a machine learning model for classifying news on Twitter using the Naïve Bayes Classifier method. The dataset used to build this model consists of news from the official Kominfo website, which has been labeled as either hoax or fact. The dataset comprises a total of 2,472 news articles, consisting of 1,415 factual news and 1,057 hoax news, resulting in an accuracy score of 71%, a precision score of 51%, a recall score of 51%, and an f1-score of 51%. Then, the built model classifies 97 unlabeled Twitter news related to booster vaccines. The results show that out of 97 news articles, 62 were classified as factual and 35 as hoaxes.

Keywords: Hoax, Classification, *Naïve Bayes Classifier*

