

## ABSTRAK

Dengan banyaknya persebaran berita saat ini, maka akan sulit untuk menentukan suatu berita tersebut tergolong dalam berita yang benar atau berita *hoax*. Berita *hoax* akan dicerna dan jika tidak berhati-hati akan mempengaruhi cara berpikir pembacanya. Informasi yang ada di berita kesehatan pasti akan mempengaruhi pemikiran seseorang mengenai kondisi tubuhnya. Padahal berita tersebut belum tentu dapat dipertanggungjawabkan keasliannya. Namun membedakan berita yang fakta dan *hoax* tidak dapat dilakukan oleh orang awam. Maka dari itu perlu adanya klasifikasi *hoax* pada berita kesehatan untuk mengidentifikasi berita tersebut sebuah *hoax* atau fakta.

Pada penelitian ini dilakukan perbandingan terhadap metode *K-Nearest Neighbor* dan *Naïve Bayes Classifier* pada klasifikasi *hoax* berita kesehatan berbahasa Indonesia. Data yang digunakan pada penelitian ini merupakan data primer sebanyak 2446 judul berita yang bersumber dari penelitian terdahulu dan hasil *web scrapping* yang telah memiliki label. Model *K-Nearest Neighbor* menggunakan nilai ketetanggaan terdekat yaitu  $K=1$ ,  $K=3$ ,  $K=5$  dan  $K=7$ . Metode TFIDF sebagai ekstraksi fitur dari teks judul berita dengan kedua metode tersebut untuk mengklasifikasikan fitur yang diperoleh sehingga dapat melakukan klasifikasi *hoax* atau fakta pada judul berita kesehatan. Lalu pada tahap terakhir dilakukan proses pengujian terhadap model *K-Nearest Neighbor* dan *Naïve Bayes Classifier* yang telah dihasilkan menggunakan tabel *confusion matrix* serta menggunakan metode validasi *k-fold cross validation* dengan dengan rasio data latih dan uji sebesar 80:20. Hasil dari pengujian tersebut berupa nilai rata-rata dari akurasi, presisi, dan *recall*.

Berdasarkan pengujian dengan *confusion matrix* dan menggunakan pendekatan *k-fold cross validation* dengan hasil bahwa model *K-Nearest Neighbor* mendapatkan hasil rata-rata terbesar pada  $K=3$  dengan akurasi 82,91%, presisi 85,3% dan *recall* 79,38%. Sedangkan model *Naïve Bayes* mendapatkan hasil rata-rata akurasi sebesar akurasi 86,42%, presisi 88,10% dan *recall* 84,05%. Pada pengujian membuktikan bahwa model *Naïve Bayes* yang dibuat lebih baik jika dibandingkan dengan model *K-Nearest Neighbor* dengan nilai  $K=3$  yang dibuktikan dengan pengukuran efektifitas *confusion matrix*.

Kata kunci: berita, klasifikasi *hoax* berita, perbandingan, , *K-Nearest Neighbor*, *Naïve Bayes*.