

IDENTIFIKASI BERITA *HOAX* MENGGUNAKAN METODE *MODIFIED K-NEAREST NEIGHBOR*

Khoirur Rosyidi

(16.04.1.1.1.00144)

Pembimbing I : Husni, S.Kom., M.T.

Pembimbing II : Dr. Wahyudi Setiawan., S.Kom., M.Kom.

ABSTRAK

Berita merupakan suatu informasi yang terbaru ataupun baru yang didapat oleh wartawan buat dilansir dalam pesan berita sehingga bisa menarik atensi untuk pembaca. Dikala ini kabar yang disebarkan kerap kali tanpa menuliskan sumbernya secara jelas. Bersamaan dengan berkembangnya teknologi informasi yang sangat mempermudah masyarakat dalam mengakses internet. Dengan kemudahan akses internet tersebut membuka kesempatan bagi masyarakat untuk mendapatkan atau menyebarkan informasi berita di jejaring sosial ataupun media online. Perihal tersebut bisa menyebabkan salah pengertian dikarenakan berita atau informasi itu belum pasti benar ataupun salah, sehingga diperlukan sesuatu sistem untuk mengklasifikasikan berita tersebut ke dalam jenis berita *hoax* atau *valid*. Proses pengklasifikasian berita *hoax* melakukan beberapa tahapan mulai pengumpulan data yang didapat dari *Mendeley* periode 2018. Peneliti setelah itu melakukan tahapan *preprocessing*, dengan tahapan pembobotan kata menggunakan *TF-IDF*, menghitung derajat kemiripan dokumen dengan *Cosine Similarity*, serta proses klasifikasi dengan menggunakan algoritma *Modified K- Nearest Neighbor*. Serta dilakukan serangkaian pengujian dan implementasi dengan *K-fold* sebanyak *K-fold3*, *K-fold4*, *K-fold5*, *K-fold6*, *K-fold8*, *K-fold10*, *K-fold12* serta *k* tetangga terdekat sebanyak *k=5*, *k=7*, *k=9*, *k=11*, *k=13*, dan didapatkan *K-fold* terbaik sebesar *K-fold8* dengan *k* tetangga terbaik sebesar *k=13*. Adapun akurasi yang didapatkan melalui serangkaian pengujian mendapatkan nilai akurasi sebesar 73%, presisi sebesar 100% serta *recal* sebesar 71%. Dengan rata- rata akurasi sebesar 61%, presisi sebesar 83% serta *recall* sebesar 63%. Berdasarkan pada hasil yang didapatkan pada pengujian tersebut dapat dilihat kalau penentuan nilai *k* bisa pengaruhi besarnya akurasi. Hasil ini masih sangat kecil disebabkan algoritma *Modified K- Nearest Neighbor* (MKNN) juga mempunyai kelemahan dan kekurangan yang sama dengan algoritma *K- Nearest Neighbor* (KNN) yaitu ada pada optimasi nilai *k*, komputasi yang kompleks dan terdapat atribut yang tidak relevan.

Kata kunci : Klasifikasi Teks ,Berita, Hoax, Modified K-Nearest Neighbor.



ABSTRACT

News is the latest or new information obtained by journalists to be reported in news messages so that it can attract the attention of readers. At this time the news that is spread is often without clearly stating the source. Along with the development of information technology that greatly facilitates people in accessing the internet. With the ease of internet access, it opens up opportunities for the public to obtain or disseminate news information on social networks or online media. This can lead to misunderstandings because the news or information is not certain to be true or false, so a system is needed to classify the news into hoax or valid news types. The process of classifying hoax news carried out several stages starting from collecting data obtained from Mendeley for the 2018 period. The researchers then carried out the preprocessing stage, with word weighting stages using TF-IDF, calculating the degree of similarity of documents with Cosine Similarity, and the classification process using the *Modified K-Nearest Neighbor* (MKNN) algorithm. As well as a series of tests and implementations with *K-fold* as much as *K-fold3*, *K-fold4*, *K-fold5*, *K-fold6*, *K-fold8*, *K-fold10*, *K-fold12* and *k* closest neighbors as much as $k=5$, $k=7$, $k=9$, $k=11$, $k=13$, and the best *K-fold* is *K-fold8* with the best *k* neighbors of $k=13$. As for the accuracy obtained through a series of tests, the *accuracy* value is 73%, *precision* is 100% and *recall* is 71%. With an average *accuracy* of 61%, *precision* of 83% and *recall* of 63%. Based on the results obtained in the test, it can be seen that the determination of the value of *k* can affect the amount of accuracy. This result is still very small because the *Modified K-Nearest Neighbor* (MKNN) algorithm also has the same weaknesses and shortcomings as the *K-Nearest Neighbor* (KNN) algorithm, namely in the optimization of *k* values, complex computations and irrelevant attributes.

Keywords : *Text Classification, News, Hoax, Modified K-Nearest Neighbor.*

