

ABSTRAK

Penyebaran berita hoaks yang berkaitan dengan sektor kesehatan dapat memberikan dampak negatif terhadap masyarakat karena dapat memengaruhi kesehatan fisik dan mental. Oleh karena itu, perlu penanganan terhadap berita hoaks agar tidak menimbulkan kerugian besar bagi masyarakat. Pada penelitian ini, model klasifikasi *machine learning* dengan metode *Support Vector Machine* (SVM) digunakan sebagai solusi untuk mendeteksi konten berisi berita hoaks dengan memanfaatkan proses *text mining*. Klasifikasi hoaks dan valid pada berita kesehatan dengan metode SVM dapat dilakukan dengan *text preprocessing*, pembobotan kata menjadi nilai dalam vektor dengan menggunakan *Term Frequency – Inverse Document Frequency* (TF-IDF), dan pemilihan jenis kernel serta parameter *cost* yang relevan. *Dataset* penelitian adalah berita kesehatan tahun 2022–2024 yang mencakup berita hoaks dan valid dengan jumlah 500 data. Evaluasi kinerja model klasifikasi dilakukan terhadap empat jenis fungsi kernel SVM, yaitu kernel *linear*, kernel *polynomial*, kernel *Radian Basis Function* (RBF), dan kernel *sigmoid*. Hasil evaluasi menunjukkan bahwa kernel *polynomial* memiliki akurasi sebesar 0,64, kernel *linear* dan kernel RBF memiliki akurasi sebesar 0,90, sedangkan kernel *sigmoid* memiliki akurasi sebesar 0,92. Dengan demikian, kernel *sigmoid* adalah fungsi kernel terbaik karena menghasilkan akurasi tertinggi. Selain itu, model yang dihasilkan dari kernel *sigmoid* memiliki nilai evaluasi *precision*, *recall*, dan *f1-score* seimbang antara kelas hoaks dan valid, artinya model stabil dan tidak condong ke salah satu kelas tertentu.

Kata Kunci: berita hoaks, klasifikasi, *text mining*, kernel, SVM

ABSTRACT

The spread of hoax news related to the health sector can have a negative impact on society because it can affect physical and mental health. Therefore, it is necessary to deal with hoax news so that it does not cause great harm to the community. In this study, a machine learning classification model with the Support Vector Machine (SVM) method is used as a solution to detect content containing hoax news by utilizing the text mining process. Classification of hoaxes and validity in health news with SVM method can be done by text preprocessing, weighting words into values in vectors using Term Frequency – Inverse Document Frequency (TF-IDF), and selecting the type of kernel and relevant cost parameters. The research dataset is health news in 2022-2024 which includes hoax and valid news with a total of 500 data. The performance evaluation of the classification model was carried out on four types of SVM kernel functions, namely linear kernel, polynomial kernel, Radian Basis Function (RBF) kernel, and sigmoid kernel. The evaluation results show that the polynomial kernel has an accuracy of 0.64, linear kernel and RBF kernel have an accuracy of 0.90, while the sigmoid kernel has an accuracy of 0.92. Thus, the sigmoid kernel is the best kernel function because it produces the highest accuracy. In addition, the model generated from the sigmoid kernel has balanced precision, recall, and f1-score evaluation values between hoax and valid classes, meaning that the model is stable and does not lean towards one particular class.

Keywords: hoax news, classification, text mining, kernel, SVM

