

ABSTRAK

Salah satu permasalahan yang dihadapi oleh layanan fasilitas kesehatan adalah banyaknya pasien yang berobat setiap hari mengakibatkan antrian yang panjang dan waktu tunggu yang lama. Hal tersebut dapat mengakibatkan menurunnya tingkat kepuasan masyarakat terhadap layanan kesehatan. Selain itu, data rekam medis pasien yang terus bertambah setiap harinya belum dimanfaatkan dengan baik. Padahal, jika data tersebut diolah dengan baik, dapat memberikan informasi yang berharga. Penelitian ini menawarkan solusi untuk permasalahan tersebut dengan mengembangkan model klasifikasi diagnosis penyakit yang akan menerima masukan berupa gejala yang dialami pasien dan menampilkan hasil klasifikasi diagnosis penyakit. Teknik yang digunakan yaitu data mining dengan metodologi *Cross Industry Standard Process for Data Mining* (CRISP-DM). Penelitian ini membandingkan hasil klasifikasi menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* yang dikombinasikan dengan *Bag of Words* dan TF-IDF untuk pemrosesan teks. Hasilnya, menunjukkan bahwa algoritma KNN dan *Bag of Words* memberikan hasil yang terbaik dengan akurasi 75%, sedangkan algoritma *Naïve Bayes* dan *Bag of Words* menghasilkan akurasi sebesar 74%. Algoritma *Naïve Bayes* dan TF-IDF dan algoritma KNN dan TF-IDF menghasilkan skor akurasi 73%. Model klasifikasi ini diharapkan dapat memberikan gambaran kepada dokter mengenai penyakit yang dialami pasien, sehingga pemeriksaan dapat dilakukan dengan lebih cepat dan efisien.

Kata kunci : *data mining, k-nearest neighbor, naïve bayes, rekam medis*