

ABSTRAK

Klasifikasi dilakukan untuk mengelompokkan sekumpulan data ke dalam kelas-kelas yang telah ditentukan terlebih dahulu berdasarkan kesamaan karakteristik yang dimiliki. Klasifikasi soal berdasarkan topik membantu para siswa dan pengajar dalam mengambil keputusan untuk menentukan soal berdasarkan kategori topiknya. Pada penelitian ini, peneliti bermaksud untuk membuat suatu model klasifikasi soal Biologi kelas 11 SMA yang dikelompokkan menjadi sembilan kategori topik yaitu Sel, Jaringan Tumbuhan dan Hewan, Sistem Gerak Manusia, Sistem Peredaran Darah, Sistem Pencernaan, Sistem Pernapasan, Sistem Ekskresi, Sistem Koordinasi, dan Sistem Reproduksi Manusia. Soal-soal dan topik didapatkan dari buku bank soal yang berjudul “*Siap Pintar Belajar Mandiri*”.

Penelitian ini membandingkan nilai akurasi dan evaluasi performansi dari dua algoritma klasifikasi yaitu, Naive Bayes dan C4.5. Untuk evaluasi performansi peneliti menggunakan *cross validation* dan mencari nilai *precision*, *recall*, dan *f1-score* menggunakan *confusion matrix*.

Dari hasil klasifikasi, diperoleh hasil akurasi algoritma Naive Bayes sebesar 72.72%, dan nilai akurasi evaluasi performansi menggunakan *cross validation* sebesar 73.09% dan nilai *precision* sebesar 73%, *recall* sebesar 73%, dan *F1-Score* sebesar 70%. Sedangkan algoritma C4.5 mendapat nilai akurasi sebesar 54.54%, dan nilai akurasi evaluasi performansi menggunakan *cross validation* sebesar 54.09% dan nilai *precision* sebesar 58%, *recall* sebesar 56%, dan *F1-Score* sebesar 55%.

Kata kunci — **Klasifikasi Soal, Biologi, Naive Bayes, C4.5, Cross Validation**