

ABSTRAK

Pada penelitian ini, peneliti bermaksud melakukan suatu klasifikasi soal Biologi SMA kelas 11 menggunakan algoritma Naïve Bayes dan Support Vector Machine yang sudah dikelompokkan menjadi 5 kategori topik yaitu Sel, Sistem Peredaran Darah, Sistem Pertahanan Tubuh, Sistem Gerak Manusia, dan Jaringan. Penelitian ini juga membandingkan performa dari dua algoritma klasifikasi, yaitu *Naïve Bayes* dan *Support Vector Machine*. Penelitian ini melalui beberapa tahap, yang pertama dengan melalui tahapan preprocessing data dengan dilakukan proses *case folding*, *tokenizing*, *Stopword Removal*, dan *Stemming*. Selanjutnya dataset dilakukan proses TF-IDF yaitu pembobotan data atau term. Metode *oversampling* SMOTE digunakan peneliti untuk mengatasi *imbalance data* dari dataset. Kemudian penggunaan *K-Fold Cross Validation* pada dataset dengan nilai k 10. Dari hasil klasifikasi, didapat hasil performa dengan menggunakan klasifikasi *Naïve Bayes* dengan *oversampling* SMOTE memiliki akurasi 76%, lalu klasifikasi *Naïve Bayes* tanpa *oversampling* SMOTE yaitu sebesar 59%, kemudian klasifikasi *Support Vector Machine* dengan *oversampling* SMOTE memiliki akurasi 70.59% dan klasifikasi *Support Vector Machine* tanpa *oversampling* SMOTE sama dengan menggunakan SMOTE yaitu 70.59%. Berdasarkan hasil akurasi yang didapat pada penelitian ini, algoritma *Naïve Bayes* merupakan metode yang lebih baik daripada *Support Vector Machine* dalam melakukan klasifikasi soal berdasarkan kategori topik.

Kata Kunci — *Klasifikasi Soal, Naïve Bayes, Support Vector Machine, SMOTE, oversampling, Cross Validation*