

## ABSTRAK

Pendidikan pada hakikatnya merupakan suatu usaha yang terpolakan untuk menciptakan proses pembelajaran aktif bagi pelajar dalam menghasilkan potensi keterampilan dan kecerdasan dengan mendapatkan pengetahuan dengan cara melakukan latihan soal. Naskah soal biasanya berisi contoh soal dengan tingkat kesulitannya dan untuk dilakukan klasifikasi berdasarkan *Revised Bloom's Taxonomy* merupakan kasus yang tidak mudah jika dilakukan dengan cara manual. Peneliti mengungkapkan untuk melakukan klasifikasi soal otomatis dengan bantuan *machine learning* berdasarkan level kognitif *Revised Bloom's Taxonomy* yang terfokus pada soal Sejarah Indonesia tingkat SMA dan tingkat C1-C3 RBT. Dataset diperoleh dari soal USBN, ulangan harian, dan bank latihan soal yang didapatkan dari internet. Metode yang digunakan adalah algoritma *K-Nearest Neighbour* dengan pendekatan jarak *Manhattan* yang melalui tahapan *preprocessing* data dengan dilakukan proses *case folding*, *tokenizing*, *filtering* dan *stemming*. Selanjutnya dataset dilakukan proses TF-IDF yaitu pembobotan kata atau *term*. Metode *oversampling* SMOTE digunakan pada penelitian untuk mengatasi *imbalance data* dari dataset yang diperoleh. Kemudian penggunaan *K-Fold Cross Validation* pada dataset dengan nilai  $k = 10$ . Hasil akurasi penelitian yang telah dilakukan dengan algoritma KNN diperoleh nilai yang cukup tinggi dengan precision 90%, recall 87%, F1-Score 87% dan accuracy 87% serta metode *oversampling* SMOTE dapat diimplementasikan dengan baik dalam menangani *imbalance data*.

Kata Kunci: **Klasifikasi soal, KNN, Manhattan, RBT, SMOTE, oversampling**