

ABSTRAK

Dalam era digital saat ini, pemanfaatan teknologi dalam pendidikan menjadi penting untuk menunjang proses pembelajaran. Penggunaan sistem otomatis dalam klasifikasi soal membantu mengurangi waktu dan usaha yang dibutuhkan dalam pengelompokan soal berdasarkan topik mengingat akan sulit dan menghabiskan banyak waktu jika dilakukan secara manual. Penelitian ini melakukan pengembangan model klasifikasi soal Matematika tingkat SMP menggunakan algoritma *Naïve Bayes* karena kesederhanaan dan kemampuannya dalam mengolah data kemudian ditingkatkan dengan metode SMOTE untuk mengatasi ketidakseimbangan data. Tujuan dari penelitian ini adalah untuk mengevaluasi performa model *Naïve Bayes* dalam klasifikasi soal Matematika dan mengidentifikasi pengaruh penggunaan simbol Matematika terhadap hasil klasifikasi. Hasil penelitian menunjukkan bahwa model yang dihasilkan mencapai akurasi 69% dengan penerapan SMOTE. Evaluasi menggunakan *cross validation* menunjukkan bahwa klasifikasi tanpa simbol matematika mencapai akurasi 89,35%, sedangkan dengan simbol matematika mencapai 88,79% dalam menggunakan SMOTE. Meski selisih hasil *cross validation* hanya 0,56%, menunjukkan bahwa baik dengan atau tanpa simbol matematika, model *Naïve Bayes* dapat menghasilkan performa yang relatif setara dengan akurasi 89% dalam mengklasifikasikan soal Matematika. Penelitian ini menunjukkan bahwa penggunaan metode SMOTE efektif dalam mengatasi masalah ketidakseimbangan data, sementara penggunaan simbol matematika memberikan dampak yang relatif kecil terhadap kinerja model.

Kata kunci— *cross validation, klasifikasi, naïve bayes, simbol, SMOTE*