

## ABSTRAK

Penentuan kelayakan kelulusan siswa merupakan salah satu aspek penting dalam mendukung pengambilan keputusan di sektor pendidikan. Namun, dalam praktiknya, variasi data yang tidak konsisten serta pengaruh regulasi pemerintah dapat berdampak signifikan terhadap kinerja model machine learning yang dihasilkan. Penelitian ini berfokus pada analisis performa algoritma Naïve Bayes ketika diterapkan pada dataset yang memiliki ketidakkonsistenan data dan ketidakseimbangan label. Dataset yang digunakan terdiri dari data penilaian 385 siswa di SD Negeri 067053 Medan Deli, yang mencakup nilai akademik, absensi, serta penilaian guru terhadap sikap dan perilaku siswa. Hasil penelitian menunjukkan bahwa model memperoleh akurasi sebesar 65%, yang dapat dikategorikan sebagai performa sedang. Analisis lebih lanjut menunjukkan bahwa data didominasi oleh kelas “Lulus” sebesar 88%. Ketidakseimbangan kelas tersebut dipengaruhi oleh regulasi pemerintah yang mewajibkan siswa sekolah dasar untuk lulus, meskipun memiliki performa akademik maupun non-akademik yang kurang baik. Kondisi ini menyebabkan model kesulitan dalam mempelajari pola yang jelas dari data. Temuan ini menunjukkan bahwa permasalahan tidak hanya terletak pada algoritma yang digunakan, tetapi juga pada kualitas dataset. Secara khusus, ketidakseimbangan kelas dan ketidakkonsistenan label memiliki peran besar dalam menurunkan performa model. Oleh karena itu, penelitian ini menekankan pentingnya validasi data dan proses pra-pemrosesan yang tepat dalam penerapan teknik machine learning pada data mining di bidang pendidikan.

Kata kunci: Data Mining, Sekolah Dasar, Naïve Bayes, Kelulusan Siswa