

# **PERBANDINGAN MODEL PEMBELAJARAN BERBASIS SMOTE MENINGKATKAN IDENTIFIKASI SISWA BERISIKO di SEKOLAH MENENGAH PERTAMA**

## **ABSTRAK**

Penelitian ini mengevaluasi efektivitas Teknik Pengambilan Sampel Berlebih Minoritas Sintetis (Synthetic Minority Over-sampling Technique/SME) dalam mengatasi ketidakseimbangan kelas dalam kumpulan data pendidikan, dengan fokus pada peningkatan model prediktif untuk mengidentifikasi siswa yang berisiko di sekolah menengah pertama di Indonesia. Dengan menggunakan kurva Receiver Operating Characteristic (ROC) dan metrik Area Under the Curve (AUC), kinerja model Decision Tree, Random Forest, dan Support Vector Machine (SVM) dinilai sebelum dan sesudah penerapan SMOTE. Model Random Forest menunjukkan peningkatan yang paling signifikan, dengan AUC meningkat dari 0,95 menjadi 0,99 setelah SMOTE, yang disebabkan oleh generalisasi berbasis ensemble pada data yang seimbang. Decision Tree menunjukkan peningkatan marjinal (AUC: 0.94→0.95), yang mencerminkan ketahanan yang melekat pada sampel sintetis, sementara SVM mengalami sedikit trade-off (AUC: 0.93→0.94), yang mungkin disebabkan oleh sensitivitas terhadap noise sintetis selama pengoptimalan hyperplane. Semua model mengungguli tebakan acak (AUC>0.93), mengonfirmasi kegunaan SMOTE dalam meningkatkan deteksi kelas minoritas untuk aplikasi penting seperti identifikasi awal siswa yang berisiko. Penelitian ini memajukan kerangka kerja praktis untuk memanfaatkan teknik pembelajaran ketidakseimbangan dalam pendidikan, menekankan wawasan yang dapat ditindaklanjuti untuk mengatasi ketidakadilan sistemik dan meningkatkan hasil pembelajaran melalui pengambilan keputusan berbasis data.

*Kata Kunci : Data Mining, Decision Tree C4.5, academic performance, educational classification, determinant factors*

