

Abstract

Data yang tidak seimbang menghadirkan tantangan yang signifikan dalam pembelajaran mesin, yang menyebabkan hasil klasifikasi menjadi bias dan menguntungkan kelas mayoritas. Masalah ini terutama terlihat pada klasifikasi kesulitan keuangan, dimana ketidakseimbangan data sering terjadi karena kelangkaan contoh-contoh tersebut dalam kumpulan data dunia nyata. Penelitian ini bertujuan untuk memitigasi ketidakseimbangan data pada perusahaan financial distress dengan menggunakan pendekatan Kmeans-SMOTE dengan menggabungkan Kmeans clustering dan Synthetic Minority Oversampling Technique (SMOTE). Berbagai pendekatan klasifikasi, termasuk Naïve Bayes dan Support Vector Machine (SVM) diterapkan pada kumpulan data kesulitan keuangan dari Kaggle untuk mengevaluasi efektivitas Kmeans-SMOTE. Hasil eksperimen menunjukkan bahwa SVM mengungguli Naïve Bayes dengan akurasi yang mengesankan (99,1%), f1-score (99,1%), Area Under Precision-Recall (AUPRC) (99,1%), dan Geometric-mean (Gmean) (98,1%).