

Abstrak

Di era digital ini, machine learning merupakan teknologi yang banyak diminati oleh organisasi dan individu. Dalam era data dan era informasi digital, kemampuan untuk mengolah data secara efisien sangat dibutuhkan. Seiring dengan bertambahnya jumlah data, machine learning memiliki berbagai masalah. Salah satunya adalah sering ditemukannya ketidakseimbangan kelas (class imbalance) seiring dengan bertambahnya jumlah data. Ketidakseimbangan kelas adalah suatu kondisi dimana suatu kelas mendominasi kelas yang lain. Salah satu contohnya adalah ketika kelas bernilai positif memiliki jumlah yang lebih sedikit dibandingkan kelas yang bernilai negatif. Kelas yang jumlahnya lebih sedikit dikategorikan sebagai kelas minoritas, sedangkan kelas yang mendominasi kumpulan data disebut kelas mayoritas. Ketidakseimbangan kelas dapat mempengaruhi kinerja klasifikasi yang tidak tepat, sehingga penanganan kelas yang tidak seimbang diperlukan untuk memperbaiki hasil klasifikasi. Klasifikasi data yang tidak seimbang menggunakan Random Forest memiliki hasil yang memuaskan, begitu juga dengan mengimplementasikan SMOTE dan ADASYN sebagai metode pengambilan sampel karena metode ini sudah umum dan mudah diimplementasikan. Dalam penelitian ini, kami menggunakan dataset protein ecoli untuk mengevaluasi performa pengklasifikasi random forest dengan dan tanpa metode oversampling. Dalam penelitian ini Dalam penelitian ini, kami menggunakan skor f1 dan akurasi seimbang sebagai metrik evaluasi utama. Kami menghitung rata-rata dengan nilai tertinggi tertinggi yaitu 84% untuk skor f1 dan 90% untuk akurasi seimbang. Baik SMOTE dan ADASYN memiliki kinerja yang sama untuk meningkatkan dan menemukan bahwa akurasi seimbang adalah metrik yang lebih cocok untuk klasifikasi yang tidak seimbang.

Kata kunci : Imbalanced Data; Random Forest; Imbalanced Ratio; SMOTE; ADASYN