

Ringkasan Ilmiah Artikel 5

Analisi SMOTE Pada Klasifikasi Hepatitis C Berbasis Random Forest dan Naïve Bayes

1. Sitasi Artikel

Sharfina, N., & Ramadhan, N. G. (2023). *Analisis SMOTE Pada Klasifikasi Hepatitis C Berbasis Random Forest dan Naïve Bayes*. *Journal of Information Technology and Computer Science (JOINTECS)*, 8(1), 33–40.

2. Latar & Tujuan

Hepatitis C merupakan salah satu penyakit serius yang menyerang hati dan sering tidak terdeteksi sejak dini. WHO mencatat lebih dari 71 juta orang terinfeksi pada 2019, tetapi kesadaran masyarakat masih rendah. Penelitian ini bertujuan mengembangkan sistem deteksi otomatis berbasis machine learning untuk mengklasifikasikan Hepatitis C. Dua algoritma diuji, yaitu Random Forest dan Naïve Bayes, dengan tambahan teknik SMOTE guna mengatasi ketidakseimbangan data.

3. Metode

Penelitian menggunakan data pasien Hepatitis C dari Kaggle dengan total 615 data dan 13 atribut. Tahapan penelitian diawali dengan preprocessing data berupa pengisian nilai kosong, encoding variabel kategorik, serta normalisasi. Selanjutnya, dilakukan penerapan teknik SMOTE untuk menyeimbangkan distribusi data antara kelas sehat (540 data) dan terinfeksi (75 data). Setelah itu, proses klasifikasi dilakukan dengan menggunakan dua model machine learning, yaitu Random Forest dan Naïve Bayes. Performa kedua model dievaluasi menggunakan confusion matrix yang menghitung akurasi, presisi, dan recall.

4. Hasil/Temuan Kunci

Hasil penelitian menunjukkan bahwa tanpa SMOTE, model Random Forest mencapai akurasi 93% dengan presisi 93% dan recall 83%, sementara Naïve Bayes hanya memperoleh akurasi 88% dengan presisi 83% dan recall 75%. Setelah data diseimbangkan menggunakan SMOTE, kinerja Random Forest meningkat menjadi 97% akurasi, 98% presisi, dan 92% recall. Sementara itu, Naïve Bayes hanya mengalami peningkatan kecil dengan akurasi 89%, presisi 84%, dan recall 77%. Temuan ini menegaskan bahwa Random Forest lebih unggul dibandingkan Naïve Bayes, terutama pada kondisi data yang tidak seimbang.

5. Kontribusi & Keterbatasan

Kontribusi utama penelitian ini adalah membuktikan efektivitas teknik SMOTE dalam meningkatkan performa klasifikasi penyakit dengan data tidak seimbang. Random Forest terbukti lebih andal dibanding Naïve Bayes. Namun, penelitian masih terbatas pada satu dataset dari Kaggle, sehingga generalisasi hasil ke populasi nyata perlu diuji lebih lanjut.

6. Takeaway Saya

Penelitian ini menunjukkan bahwa kombinasi SMOTE + Random Forest dapat meningkatkan akurasi diagnosis Hepatitis C secara signifikan, yang berpotensi membantu tenaga medis dalam deteksi dini.