

## **Ringkasan Ilmiah Artikel 5: Analisis SMOTE Pada Klasifikasi Hepatitis C Berbasis Random Forest dan Naïve Bayes**

Review Kritis Singkat: Artikel ini menonjol banget dalam pakai SMOTE buat atasi data nggak seimbang di klasifikasi Hepatitis C, dengan Random Forest yang dapet akurasi tinggi. Tapi, dataset sekunder-nya mungkin bias, dan sayangnya kurang metrik lengkap kayak F1-score atau ROC curve buat evaluasi lebih dalam. Yang paling kurang, belum ada cross-validation ekstensif atau adaptasi buat data orang Asia, yang bisa bikin generalisasinya lemah.

### **Sitasi Artikel (IEEE)**

N. Sharfina and N. G. Ramadhan, "Analisis SMOTE Pada Klasifikasi Hepatitis C Berbasis Random Forest dan Naïve Bayes," J. Inf. Technol. Eng. Comput. Sci. (JOINTECS), vol. 8, no. 1, pp. 33-40, Jun. 2023. doi: 10.31328/jointecs.v8i1.XXXX. (Terakreditasi SINTA 3, SK No. 225/E/KPT/2022, berlaku Vol. 7 No. 1/2022–Vol. 11 No. 2/2026).

### **Latar & Tujuan**

Hepatitis C Virus (HCV) ngerusak sekitar 71 juta orang dunia menurut WHO (2019), tapi kesadaran cuma 49.7%, dan data medis sering nggak seimbang yang bikin akurasi klasifikasi jeblok. Penelitian lama pakai Decision Tree atau C4.5 cuma sampe 55-96%, jadi butuh teknik oversampling biar lebih akurat. Tujuan utamanya analisis Synthetic Minority Over-sampling Technique (SMOTE) di klasifikasi HCV pakai Random Forest (RF) buat tangani outlier dan Naïve Bayes (NB) buat yang sederhana, biar dukung diagnosis dini dan urus data imbalanced di kesehatan.

### **Metode**

Dataset HCV sekunder (dari sumber terpercaya) diolah pakai SMOTE buat seimbangkan kelas mayoritas-minoritas lewat sampel sintetis. RF dipake dengan Gini impurity buat pilih fitur, NB dengan asumsi probabilitas independen. Data dibagi train-test biasa, evaluasi lewat confusion matrix (TP/TN/FP/FN). Alurnya: preprocessing (normalisasi, isi missing values), aplikasikan SMOTE, latih model (RF 100 pohon, NB Gaussian), bandingin sama baseline kayak Decision Tree. Implementasi di Python pake scikit-learn dan imbalanced-learn.

### **Hasil/Temuan Kunci**

Tanpa SMOTE, RF dapet 93% akurasi dan NB 88%; abis SMOTE, RF naik jadi 98% dan NB 89%, ngalahin Decision Tree (55%) sama C4.5 (96%). Confusion matrix nunjukin kesalahan turun drastis, RF jago kurangi false negative di kelas minor (pasien positif HCV). Tabel banding metrik bilang SMOTE bantu tingkatkan sensitivitas, RF lebih oke tangani fitur saling kait kayak enzim hati dan gejala. Hasil stabil di data imbalanced, naik 5-10% dari tanpa oversampling.

### **Kontribusi & Keterbatasan**

Kontribusi besarnya framework SMOTE + RF/NB buat klasifikasi medis imbalanced, adaptasi baru buat HCV di Indonesia, dapet akurasi >98% buat diagnosis otomatis dan keputusan klinis. Inovasi: analisis fitur penting RF buat biomarker HCV, plus evaluasi confusion matrix yang detail. Keterbatasan: dataset sekunder bisa bias (kurang wakilin orang Asia); nggak ada k-fold cross-validation buat validasi kuat; asumsi independen NB nggak selalu pas buat data medis; belum integrasi data real-time dari RS.

## **Takeaway Saya**

SMOTE di model kayak RF ini krusial banget buat data medis yang nggak rata. Buat proyek saya, ini motivasi bikin sistem prediksi kesehatan lebih akurat, tapi harus tambahin metrik evaluasi lengkap dan validasi lebih dalam biar bisa dipake di klinik beneran.