

ABSTRAK

ALVIAN SYAH BURHANI, Pemanfaatan *Ensemble Random Forest-XGBoost* untuk Klasifikasi Penyakit ISPA :Studi Kasus RSUD Latemmamala. (di bimbing oleh Fahrir Irhamna Rachman S.T., M.T dan Ir. Muhammad Faisal, S.SI., M.T., Ph.D.,IPM).

Infeksi Saluran Pernapasan Akut (ISPA) masih menjadi masalah kesehatan utama di Indonesia dengan determinan risiko yang kompleks. Proses diagnosis manual dinilai tidak efisien dan rentan terhadap subjektivitas, khususnya di RSUD Latemmamala. Penelitian ini bertujuan membangun model klasifikasi tiga kategori risiko ISPA, Tidak ISPA, Sedang, dan Berat menggunakan pendekatan *Ensemble Learning* berbasis *Soft Voting* yang mengintegrasikan algoritma *Random Forest* dan *Extreme Gradient Boosting (XGBoost)*.

Penelitian menggunakan 458 rekam medis pasien periode 2020-2025 setelah proses data cleaning diperoleh 457 data bersih. *Dataset* dibagi dengan proporsi 80% data latih dan 20% data uji menggunakan *stratified sampling*. Pra-pemrosesan meliputi pembersihan data, transformasi, normalisasi, dan penyeimbangan kelas melalui *Random Undersampling*.

Hasil evaluasi menunjukkan model *Ensemble Soft Voting* mencatatkan performa terbaik dengan akurasi 96,74%, presisi 96,97%, *recall* 96,88%, dan *F1-Score* 96,77%. Model ini berhasil mencapai nilai *recall* sempurna pada kategori ISPA Berat. Selain itu, pengujian pada 30 data pasien baru menghasilkan akurasi 90,00%, dan hasil validasi oleh dua pakar medis mengonfirmasi bahwa klasifikasi model telah sejalan dengan logika klinis. Disimpulkan bahwa model ini merupakan alat bantu keputusan yang andal untuk triase pasien secara cepat, objektif, dan akurat.

Kata Kunci : Infeksi Saluran Pernapasan Akut, *Ensemble Learning*, *Soft Voting*, *Random Forest*, *Extreme Gradient Boosting*, Klasifikasi Risiko

ABSTRACT

ALVIAN SYAH BURHANI, *Utilization of Ensemble Random Forest-XGBoost for Acute Respiratory Infection Disease Classification: A Case Study of RSUD Latemmamala. (Supervised by Fahrin Irhamna Rachman S.T., M.T and Ir. Muhammad Faisal, S.SI., M.T., Ph.D.,IPM).*

Acute Respiratory Infection (ARI) remains a major health issue in Indonesia with complex risk determinants. Manual diagnosis processes are considered inefficient and prone to subjectivity, particularly at RSUD Latemmamala. This research aims to develop a classification model for three ARI risk categories-No ARI, Moderate, and Severe-using a Soft Voting-based Ensemble Learning approach that integrates Random Forest and Extreme Gradient Boosting (XGBoost) algorithms.

The study utilized 458 patient medical records from the 2020–2025 period, yielding 457 clean records after the data cleaning process. The dataset was partitioned into an 80% Training set and a 20% testing set using stratified sampling. Data preprocessing encompassed data cleaning, transformation, normalization, and class balancing via Random Undersampling

Evaluation results showed that the Ensemble Soft Voting model achieved the best performance with 96.74% accuracy, 96.97% precision, 96.88% recall, and a 96.77% F1-Score. The model successfully achieved perfect recall for the Severe ARI category. Furthermore, testing on 30 new patient datasets yielded 90.00% accuracy, and validation by two medical experts confirmed that the model's classification aligns with clinical logic. It is concluded that this model serves as a reliable decision-support tool for rapid, objective, and accurate patient triage.

Keywords: *Acute Respiratory Infection, Ensemble Learning, Soft Voting, Random Forest, Extreme Gradient Boosting, Risk Classification.*