

Model Ensemble Random Forest dan Support Vector Machine untuk Mendeteksi Penyakit Pneumonia Menggunakan Data Sekuens Protein = Ensemble Model of Random Forest and Support Vector Machine for Detecting Pneumonia Using Protein Sequence Data

Ryan Fathurrachman, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920567424&lokasi=lokal>

Abstrak

ISPA atau infeksi saluran pernapasan akut adalah infeksi yang menyerang saluran pernapasan, baik saluran pernapasan atas maupun bawah. Salah satu penyakit yang termasuk dalam ISPA adalah pneumonia. Pneumonia merupakan infeksi paru-paru yang dapat memengaruhi kesehatan manusia secara serius. Pneumonia memengaruhi paru-paru bagian bawah dan menjadi penyebab area tersebut dipenuhi cairan lendir atau nanah. Pneumonia dikarenakan oleh berbagai agen patogen seperti virus, bakteri, dan jamur. Bakteri yang paling sering menyebabkan pneumonia adalah *Streptococcus pneumoniae*. Selain itu, *Mycobacterium tuberculosis* juga merupakan bakteri penyebab pneumonia di beberapa negara Asia. Berdasarkan hasil radiologi, pneumonia mirip dengan pneumonia tuberkulosis. Diagnosis dini sangat berperan penting dalam pengelolaan dan pengobatan efektif untuk penyakit ini. Dengan adanya kemajuan di bidang bioinformatika, sekuens protein menjadi salah satu pendekatan yang potensial untuk mendeteksi pneumonia secara cepat dan akurat. Oleh karena itu, penelitian ini adalah pendeteksian penyakit pneumonia dengan sekuens protein. Ekstraksi fitur untuk menjadi data numerik dibutuhkan pada penelitian ini dengan metode discere. Penelitian ini menggunakan metode ensemble dari model Random Forest dan Support Vector Machine (SVM) dengan weighted majority algorithm (WMA) untuk mendeteksi penyakit pneumonia menggunakan sekuens protein *Streptococcus pneumoniae* dan *Mycobacterium tuberculosis* sebagai pembandingan yang didapatkan melalui situs UniProt. Hasil penelitian ini menunjukkan bahwa metode ensemble model Random Forest dan model SVM dengan metode WMA memiliki kinerja terbaik dengan perbandingan data training dan data testing sebesar 80:20 didapat nilai akurasi sebesar 99,17%, nilai sensitivitas sebesar 99,65%, nilai spesifisitas sebesar 97,56%, dan nilai ROC-AUC sebesar 98,61%.

.....Infection of Acute Respiratory (ARI) is an infection that attacks the respiratory tract, affecting both the upper and lower respiratory tracts. One of the diseases included in ARI is pneumonia. Pneumonia is a lung infection that can seriously impact human health. It affects the lower part of the lungs and causes the area to fill with mucus or pus. Pneumonia can be caused by various pathogens such as viruses, bacteria, and fungi. The bacterium most commonly causing pneumonia is *Streptococcus pneumoniae*. Additionally, *Mycobacterium tuberculosis* is also a bacterial cause of pneumonia in several Asian countries. Based on radiological results, pneumonia is similar to tuberculosis pneumonia. Early diagnosis is crucial in the management and effective treatment of this disease. With advancements in bioinformatics, protein sequence has become a potential approach for the rapid and accurate detection of pneumonia. Therefore, this research focuses on the detection of pneumonia using protein sequences. Feature extraction is required to convert the data into numerical form using discere method. This research uses an ensemble method combining Random Forest and Support Vector Machine (SVM) models with the weighted majority algorithm (WMA) to detect pneumonia using protein sequences of *Streptococcus pneumoniae* and *Mycobacterium tuberculosis* for comparison. This protein sequences obtained from the UniProt website. The results of this research indicate

that the ensemble method of Random Forest and SVM with WMA achieved the best performance with a training to testing data ratio of 80:20 with 99,17% accuracy, 99,65% sensitivity, 97,56% specificity, and 98,61% ROC-AUC score.