

Klasifikasi sekuens protein Coronavirus penyebab COVID-19 menggunakan metode Particle Swarm Optimization-Support Vector Machine dan Seleksi Fitur Random Forest-Recursive Feature Elimination = Classification of coronavirus protein sequences cause COVID-19 disease using Particle Swarm Optimization-Support Vector Machine Method and Feature Selection of Random Forest-Recursive Feature Elimination

Dian Puspita Sari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20524537&lokasi=lokal>

Abstrak

Coronavirus yaitu kelompok virus yang menginfeksi sistem pernapasan yang dapat menyebabkan infeksi pernapasan ringan maupun berat. Salah satu virus yang termasuk ke dalam coronavirus adalah SARS-CoV-2. Penyakit yang disebabkan oleh virus SARS-CoV-2 disebut COVID-19. COVID-19 pertama kali terdeteksi pada tahun 2019 di Wuhan, China. Penyebaran COVID-19 sangat cepat dengan tingkat kematian yang tinggi terus terjadi di berbagai negara sehingga penyakit ini berstatus pandemi. Skripsi ini menyelesaikan masalah klasifikasi virus SARS-CoV-2 dengan menggunakan data sekuens protein coronavirus. Seleksi fitur pada data sekuens protein coronavirus menggunakan metode seleksi fitur Random Forest-Recursive Feature Elimination (RF-RFE). Setelah dilakukan seleksi fitur, dilakukan klasifikasi menggunakan pendekatan machine learning dengan metode Support Vector Machine (SVM) dan Particle Swarm Optimization-Support Vector Machine (PSO-SVM). Hasil terbaik performa rata-rata akurasi, spesifisitas, dan sensitivitas untuk metode SVM berturut-turut adalah 93,43%, 98,06%, dan 88,84% pada data pelatihan sebesar 80%. Untuk metode PSO-SVM, hasil terbaik rata-rata akurasi dan spesifisitas adalah 98,48% dan 98,57% pada data pelatihan sebesar 80%, sedangkan hasil terbaik rata-rata sensitivitas adalah 98,96% pada data pelatihan sebesar 90%. Oleh karena itu, pada penelitian ini dapat disimpulkan bahwa metode PSO-SVM menghasilkan performa yang lebih baik dibandingkan dengan metode SVM.

.....Coronaviruses are a group of viruses that infect the respiratory system that can cause mild or severe respiratory infections. One of the viruses that belongs to the coronavirus is SARS-CoV-2. The disease caused by the SARS-CoV-2 virus is called COVID-19. COVID-19 was first detected in 2019 in Wuhan, China. The spread of COVID-19 is very fast with a high mortality rate that continues to occur in various countries so that this disease has a pandemic status. This thesis solves the problem of classifying the SARS-CoV-2 virus using coronavirus protein sequence data. Feature selection on coronavirus protein sequence data used the Random Forest-Recursive Feature Elimination (RF-RFE) feature selection method. After feature selection, classification is carried out using a machine learning approach with the Support Vector Machine (SVM) and Particle Swarm Optimization-Support Vector Machine (PSO-SVM) methods. The best results of the average performance of accuracy, specificity, and sensitivity for the SVM method are 93.43%, 98.06%, and 88.84%, respectively, for training data of 80%. For the PSO-SVM method, the best results on average accuracy and specificity are 98.48% and 98.57% on training data of 80%, while the best results on average sensitivity are 98.96% on training data of 90%. Therefore, in this study it can be concluded that the PSO-SVM method produces better performance than the SVM method.