

Klasifikasi sekuens protein coronavirus penyebab COVID-19 menggunakan metode Naive Bayes dengan seleksi fitur Lasso = Classification of coronavirus protein sequences cause COVID-19 using Naive Bayes method with LASSO feature selection

Ghani Deori, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920554799&lokasi=lokal>

Abstrak

SARS-COV-2 merupakan jenis virus yang menyebabkan pandemi COVID-19. Pandemi COVID-19 pertama kali terdeteksi di Wuhan, Cina. Berdasarkan data World Health Organization (WHO), jumlah orang yang telah terpapar COVID-19 adalah 123.216.178 orang dan 2.714.517 orang meninggal akibat COVID-19 berdasarkan data www.who.int pada tanggal 23 Maret 2021. Pada skripsi ini, dilakukan klasifikasi untuk SARS-COV-2 dengan menggunakan sekuens protein dari SARS-COV-2. Sekuens protein SARS-COV-2 di ekstraksi fitur dengan menggunakan package discere dari Python. Package discere akan menghasilkan 27 fitur, dimana fitur-fitur diseleksi dengan menggunakan metode LASSO (Least Absolute Shrinkage and Selection Operator). Setelah dilakukan seleksi fitur, dilakukan klasifikasi dengan menggunakan dua metode, yaitu metode Absolute Correlation Weighted Naïve Bayes dan metode Naïve Bayes. Rata-rata akurasi, sensitifitas, dan spesifisitas tertinggi untuk metode Absolute Correlation Weighted Naïve Bayes berturut-turut adalah 81,85%, 74,81%, dan 89,19%, sedangkan rata-rata akurasi, sensitifitas, dan spesifisitas tertinggi untuk metode Naïve Bayes berturut-turut adalah 81,44%, 74,58%, dan 88,24%. Terlihat bahwa metode Absolute Correlation Weighted Naïve Bayes mempunyai rata-rata akurasi, sensitifitas, dan spesifisitas yang lebih tinggi dibandingkan dengan metode Naïve Bayes.

.....SARS-COV-2 is the type of virus that causes the COVID-19 pandemic. The COVID-19 pandemic was first detected in Wuhan, China. Based on data from the World Health Organization (WHO), the number of people who have been exposed to COVID-19 is 123,216,178 people and 2,714,517 people died from COVID-19 based on data from www.who.int on March 23, 2021. In this paper, the SARS-COV-2 classification is done by using the protein sequence of SARS-COV-2. The SARS-COV-2 protein sequence will be feature extraction using the discere package from Python. The discere package will produce 27 features, where the features are selected using the LASSO (Least Absolute Shrinkage and Selection Operator) method. After feature selection, classification is carried out using two methods, namely the Absolute Correlation Weighted Naïve Bayes method and the Naïve Bayes method. The highest average accuracy, sensitivity, and specificity for the Absolute Correlation Weighted Naïve Bayes method are 81.85%, 74.81%, and 89.19%, respectively, whereas the highest average accuracy, sensitivity, and specificity for the Naïve Bayes method are 81.44%, 74.58%, and 88.24%, respectively. It can be seen that the Absolute Correlation Weighted Naïve Bayes method has a higher average accuracy, sensitivity, and specificity than the Naïve Bayes method.