

Prediksi interaksi protein berdasarkan barisan asam amino pada human immunodeficiency virus menggunakan pseudo substitution matrix representation feature dikombinasikan dengan rotation forest ensemble classifier = Sequence base prediction of protein-protein interactions using pseudo substitution matrix representation feature and ensemble rotation forest classifier in HIV (human immunodeficiency virus)

Susilo Hartomo, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20467571&lokasi=lokal>

Abstrak

ABSTRAK

Human Immunodeficiency Virus (virus HIV) adalah virus penyebab penyakit Acquired Immunodeficiency Syndrome (AIDS). Virus HIV merupakan retro virus yang menginfeksi sel-sel sistem kekebalan tubuh manusia. Virus HIV dalam proses menginfeksi tubuh manusia tidak dapat lepas dari terjadinya interaksi antara protein HIV dan protein manusia. Untuk menganalisis jaringan interaksi antar protein dapat menggunakan barisan asam amino penyusun dari suatu protein. Proses analisis dapat dilakukan dengan metode konvensional maupun komputasi. Keunggulan menggunakan metode komputasi yaitu dapat menghemat waktu maupun biaya. Metode rotation forest merupakan sebuah metode ensemble untuk pemrediksian. Dalam pemrediksian interaksi protein menggunakan barisan asam amino rotation forest akan dikombinasikan dengan fitur ekstraksi dalam penelitian ini menggunakan PseudoSMR. Data untuk membuat model prediksi berupa interaksi protein beserta barisan asam amino. Data tersebut dapat diperoleh dari NCBI (National Centre for Biotechnology Information). Fitur ekstraksi menggunakan PseudoSMR akan merubah barisan asam amino berupa barisan alphabet menjadi sebuah vektor ciri. Di dalam fitur ekstraksi PseudoSMR terdapat variabel yang membuat matriks hasil dari fitur ekstraksi PseudoSMR mempunyai ukuran yang berbeda. Matriks hasil ekstraksi ciri akan dijadikan sebagai data training dan data testing dalam pembuatan model rotation forest, sehingga total data set yang digunakan ada sebanyak 6 data set. Hasil prediksi rotation forest untuk RF(PCA) yang paling bagus pada data dan untuk RF(IPCA) yang paling bagus pada data. Nilai evaluasi performa pada kisaran 0.759436 sampai 0.793178 untuk RF(IPCA) sedangkan untuk RF(PCA) pada kisaran 0.774837 sampai 0.812225. Nilai evaluasi tertinggi didapat pada saat variabel dimana merupakan panjang kolom. Semua perhitungan dan proses membuat model prediksi menggunakan software R.

<hr>

ABSTRACT

Human Immunodeficiency Virus (HIV virus) is a virus that causes the disease Acquired Immunodeficiency Syndrome (AIDS) which is a retro virus that infects cells of the human immune system. In a process of infecting the human body, HIV virus can not be separated from the interaction between proteins HIV and human. To analyze the interaction tissue between proteins, we can use sequence of amino acids from a protein compound. The process of analysis can be done by conventional and computational methods. The advantages of using computational methods that can save time and cost efficiently. Rotation forest is an ensemble method for the classification to be combined with the PseudoSMR feature extraction. To make predictive model, data is needed in the form of protein interaction along with amino acid sequence that can

be obtained from NCBI (National Center for Biotechnology Information). Feature extraction of the amino acid sequence will be transformed into a feature vector using PseudoSMR. The result of PseudoSMR by using parameter $lg = \{2,3,5,6,8,10\}$, which will be used as training data and test data in rotation forest, so there are 6 datasets. The best result for RF(PCA) occur when $lg = 5$ and for RF(IPCA) occur when $lg = 8$. The value of performance evaluation in the range 0.759436 to 0.793178 for RF(IPCA) while for RF(PCA) in the range 0.774837 to 0.812225. The highest evaluation value is obtained when the variable $K = p / 3$ where p is the length of the column. All calculations and prediction process using software R.