



ANALISIS AKURASI DATA PROTEIN VIRUS SARS-COV-2 DENGAN MENGGUNAKAN METODE DATA MINING

NRP/NAMA MAHASISWA: 1773012 / HANS CHRISTIAN
DOSEN PEMBIMBING : TIUR GANTINI, S.T., M.T.

TUGAS AKHIR
GENAP 2020/2021

LATAR BELAKANG

Bioinformatika merupakan ilmu dari biologi, matematika, dan ilmu komputer untuk menganalisis data-data biologi. Bioinformatika mempelajari data biologi dengan menerapkan teknik komputasional untuk pengelolaan dan penganalisisan informasi biologis.

Keakuratan data dibutuhkan setinggi-tingginya untuk analisis data. Oleh karena itu dalam penelitian ini akan menggunakan metode data mining dengan model klasifikasi Naïve Bayes dan J48 untuk kumpulan data protein SARS-CoV-2 yang disediakan oleh database website NCBI dan menganalisis hasil akurasi antara model klasifikasi Naïve Bayes dan J48.

TUJUAN

Mencari model klasifikasi dengan akurasi yang tepat untuk kumpulan data & Memprediksi protein virus SARS-CoV-2 dengan metode data mining J48 dan Naïve Bayes.

DATA MINING

Data mining merupakan metode untuk menggali suatu pola dari kumpulan data yang besar dan kompleks dengan menggunakan algoritma.

Teknik untuk mengidentifikasi pola pada data telah ada sejak jaman dahulu, seperti Bayes theorem (1700-an) dan evolusi regresi (1800-an). Pada tahun 1980-an, kata "database mining", telah digunakan sebagai ungkapan data mining namun, kata tersebut di merk dagang oleh HNC software sehingga peneliti akhirnya beralih pada kata "Data mining" yang mulai dipopulerkan dengan adanya jurnal Data mining and Knowledge Discovery pada konferensi internasional KDD di tahun 1995

SIMPULAN

Berdasarkan penelitian ini dapat disimpulkan:

1. Pencarian akurasi pada kumpulan data yaitu menggunakan algoritma model klasifikasi Naïve Bayes dan J48 dengan 10-fold cross-validation untuk melakukan data mining dengan aplikasi WEKA pada kumpulan data SARS-CoV-2 yang diambil melalui website NCBI.
2. Hasil penelitian dalam memprediksi protein pada kumpulan data SARS-CoV-2 menggunakan model klasifikasi Naïve Bayes dengan akurasi prediksi sebesar 88.8% dan J48 dengan akurasi prediksi sebesar 92.6%. Mengenai kinerja antara model klasifikasi Naïve Bayes dan J48 menunjukkan kinerja model Naïve Bayes memiliki nilai akurasi yang lebih rendah tetapi dalam pembuatan model klasifikasinya lebih cepat dibandingkan model J48 yang memiliki akurasi lebih tinggi namun, membutuhkan waktu yang lebih lama dalam proses pembuatan model.

SARAN

Saran mengenai penelitian ini untuk peneliti selanjutnya berupa:

1. Menambahkan kelas untuk kumpulan data sehingga dapat menguji prediksi dengan kelas yang lebih kompleks.
2. Menggunakan metode algoritma selain Naïve Bayes dan J48 untuk model klasifikasi.
3. Menambahkan data uji pada tahap pengujian data sehingga dapat menguji kinerja model klasifikasi jika data ujinya banyak.

