

INTISARI

KLASIFIKASI JENIS PENYAKIT INFEKSI MENGGUNAKAN SUPPORT VECTOR MACHINE, DECISION TREE, DAN K-NEAREST NEIGHBORS

Oleh

Mohamad Raychan Abdul Rosyid

17/418649/PPA/05433

Penyakit menular masih menjadi penyebab utama kematian di seluruh dunia disebabkan munculnya penyakit infeksi baru (emerging disease) dan munculnya kembali penyakit menular lama (re-emerging disease). Salah satu cara untuk mendiagnosis penyakit infeksi secara cepat dan akurat dengan membuat permodelan mesin pembelajaran yang bertujuan untuk membantu dokter atau praktisi kesehatan muda ketika sedang melakukan ko-as (cooperative assistant).

Metode klasifikasi yang akan digunakan pada penelitian ini adalah support vector machine k-nearest neighbors, dan decision tree. Metode tersebut dapat mengklasifikasikan jenis penyakit infeksi berdasarkan kemiripan tanda dan gejalanya. Penelitian ini juga menggunakan TF (Term Frequency) dan TF-IDF (Term Frequency-Inverse Document Frequency) sebagai ekstraksi fitur untuk memberi pembobotan setiap kata dalam suatu dokumen dengan pembersihan data menggunakan teknik-teknik seperti case folding, tokenisasi, stopwords removal, dan stemming.

Pengujian model dilakukan dengan menggunakan sebanyak 464 data dengan 20 kelas jenis penyakit infeksi. Hasil performa klasifikasi pada penelitian ini dengan menggunakan algoritme *support vector machine* sebesar 90%, algoritme *k-nearest neighbors* sebesar 85%, dan algoritme *decision tree* sebesar 61%.

Kata kunci: Penyakit Infeksi, Klasifikasi, *Text Mining*, *Supervised Learning*

ABSTRACT

INFECTION DISEASE CLASSIFICATION USING SUPPORT VECTOR MACHINE, DECISION TREE, AND K-NEAREST NEIGHBORS

By

Mohamad Raychan Abdul Rosyid

17/418649/PPA/05433

Infectious diseases are still the main cause of death worldwide due to the emergence of new infectious diseases (emerging disease) and the re-emergence of old infectious diseases (re-emerging disease). One way to diagnose infectious diseases quickly and accurately is by making a machine learning model that aims to help young doctors or health practitioners when doing co-as (cooperative assistant).

The classification methods that will be used in this study are the support vector machine k-nearest neighbors, and the decision tree. The method can classify types of infectious diseases based on the similarity of signs and symptoms. This study also uses TF-IDF (Term Frequency-Inverse Document Frequency) as a feature extraction to weight each word in a document by cleaning data using techniques such as case folding, tokenization, stopword removal, and stemming.

Model testing was carried out using 464 data with 20 classes of infectious diseases. The results of this study are performances in the form of accuracy of 90% for support vector machine, accuracy of 85% for k-nearest neighbors, and accuracy of 61% for decision tree.

Keywords: Infection Disease, Classification, *Text Mining*, *Supervised Learning*