

ABSTRAK

Tuberculosis atau yang biasa juga dikenal dengan TBC merupakan penyakit paru-paru yang diakibatkan oleh kuman yang memiliki nama latin *Mycobacterium tuberculosis*. Kasus TBC di Dunia sampai saat ini masih terbilang tinggi termasuk untuk daerah kabupaten Bangkalan. Menurut data yang didapat dari RSUD Syarifah Ambami Rato Ebu Bangkalan terdapat 100 data pasien TBC dengan 27 fitur yang diperlukan untuk mendiagnosa pasien tersebut, hal ini yang menjadi alasan penelitian ini untuk menguji apakah bisa untuk mengklasifikasikan penyakit TBC tadi dengan mengurangi jumlah fitur atau gejala yaitu dengan Seleksi Fitur *Information Gain* dan untuk Pengklasifikasiannya menggunakan metode *WKNN*. *Information Gain* berguna untuk mengetahui tingkat kepentingan atau korelevansi suatu gejala terhadap pengklasifikasian penyakit TBC dengan metode *WKNN*. Berdasarkan pengujian yang dilakukan didapatkan keakuratan sistem sebesar 53.2625% sebelum menggunakan *Information Gain* dan 57.125% setelah menggunakan *Information Gain* dengan hanya menggunakan 25 fitur.

Kata Kunci : *Sistem, Identifikasi , Tuberculosis, Weighted K-Nearest Neighbor, Seleksi Fitur, Information Gain.*

ABSTRACT

Tuberculosis or commonly known as tuberculosis is a disease lungs caused by a germ that has the Latin name *Mycobacterium tuberculosis*. The case of tuberculosis in the world is still relatively high for the district of Bangkalan. According to data obtained from RSUD Syarifah Ambami Rato Ebu Bangkalan, there are 100 TB patient data sets with 27 features necessary to diagnose this patient, this is the reason

This study is to test whether it can classify TB disease earlier by reducing the number of features or symptoms, namely the Feature Selection *Information Gain* and for classification using the method *WKNN*. *Information Gain* is useful for knowing the level of interest or the relevance of a symptom to the classification of TB disease by method *WKNN*. Based on the tests carried out, the accuracy of the system is obtained amounted to 53.2625% before using *Information Gain* and 57.125% after using *Information Gain* using only 25 features.

Keywords: System, Identification, Tuberculosis, Weighted K-Nearest Neighbor, Feature Selection, Information Gain.