

ABSTRAK

Tuberculosis merupakan penyakit menular yang dapat menyebarkan penyakit lewat percikan droplet atau liur ketika berbicara, batuk, bersin dan lain-lain. Indonesia merupakan negara yang berada di urutan ke 3 dengan jumlah kasus TBC terbesar di dunia. Aceh merupakan salah satu provinsi dengan beban tuberkulosis yang tinggi di Sumatera. Aceh Utara mempunyai jumlah kasus tertinggi (13% dari seluruh kasus di Aceh). Analisis data menggunakan bahasa pemrograman *python* dengan metode klasifikasi dan *clustering* untuk melihat keparahan penyakit TB Paru di Aceh Utara, membandingkan hasil akurasi dari ketiga metode yaitu Algoritma C5.0, *Modified K-Nearest Neighbor* dan *Random Forest*, sedangkan *clustering* menggunakan metode *Agglomerative Hierarchical Clustering* untuk pengelompokkan dan pemetaan. Hasil penelitian menggunakan metode klasifikasi yaitu Metode Algoritma C5.0 memiliki nilai akurasinya 93%, Metode *Modified K-Nearest Neighbor* dan *Random Forest* memiliki hasil akurasi 90%. Hasil pengelompokkan menggunakan *Agglomerative Hierarchical Clustering* yaitu *Cluster 1* = 16 kecamatan, *Cluster 2* = 3 kecamatan, dan *Cluster 3* = 8 kecamatan. Wilayah Tinggi terdiri 8 kecamatan, wilayah Sedang terdiri 3 kecamatan dan wilayah Rendah terdiri 16 kecamatan.

Kata Kunci: Tuberkulosis Paru, *Classification*, *Clustering*, *Python*

ABSTRACT

Tuberculosis is an infectious disease that can spread the disease through droplets or saliva when talking, coughing, sneezing and so on. Indonesia is the country that is in 3rd place with the largest number of TB cases in the world. Aceh is one of the provinces with a high burden of tuberculosis in Sumatra. North Aceh has the highest number of cases (13% of all cases in Aceh). Data analysis uses Python programming language with classification and clustering methods to see the severity of pulmonary TB disease in North Aceh, comparing the accuracy results of the three methods, namely the C5.0 Algorithm, Modified K-Nearest Neighbor and Random Forest, while clustering uses the Agglomerative Hierarchical Clustering method to grouping and mapping. The results of the research using classification methods, namely the C5.0 Algorithm Method, have an accuracy value of 93%, the Modified K-Nearest Neighbor and Random Forest methods have an accuracy result of 90%. The results of grouping using Agglomerative Hierarchical Clustering are Cluster 1 = 16 sub-districts, Cluster 2 = 3 sub-districts, and Cluster 3 = 8 sub-districts. The High region consists of 8 sub-districts, the Medium region consists of 3 sub-districts and the Low region consists of 16 sub-districts.

Keywords: Pulmonary Tuberculosis, Classification, Clustering, Python