

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasi jumlah kasus penyakit malaria di Kabupaten Asahan Menggunakan Penerapan *Random Forest*. Metode *Random Forest* dipilih karena kemampuannya dalam menangani data dengan variabel yang banyak dan kompleks serta mengurangi *overfitting*. Data kasus malaria dikumpulkan dari Dinas Kesehatan Kabupaten Asahan. Proses penelitian meliputi berbagai tahap, yaitu pengumpulan data, *preprocessing* data, pelatihan model, dan evaluasi model. Hasil penelitian menunjukkan bahwa metode *Random Forest* dapat memberikan kinerja yang baik dalam klasifikasi jumlah kasus malaria. Peneliti menggunakan dataset sebanyak 568 data, yang dibagi menjadi 80% untuk *training* dan 20% untuk *testing*. Dari total dataset yang terdiri dari 25 Kecamatan yang ada di Kabupaten Asahan terdapat 568 data yang dibagi menjadi 109 data (19.2%) untuk kategori rendah, 334 data (58.8%) untuk kategori sedang dan 125 data (22.0%) untuk kategori tinggi. Variabel yang paling berpengaruh pada penelitian ini adalah puskesmas dengan (47.53%), (43.77%) untuk Kecamatan, (6.07%) untuk umur, (2.18%) untuk bulan dan (0.45%) untuk jenis kelamin. Model *Random Forest* yang dibangun kemudian dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-Score* untuk mengukur kinerja model dalam klasifikasi. Berdasarkan hasil penelitian, model *Random Forest* berhasil mencapai tingkat akurasi sebesar 0.97%, presisi sebesar 0.97%, *recall* sebesar 0.97%, dan *F1-Score* sebesar 0.97%.

Kata Kunci : *Klasifikasi, Malaria, Random Forest, Akurasi*

ABSTRACT

This research aims to classify the number of malaria cases in Asahan Regency using the Random Forest method. The Random Forest method was chosen for its ability to handle data with numerous and complex variables while reducing overfitting. The malaria case data was collected from the Health Office of Asahan Regency. The research process includes several stages, namely data collection, data preprocessing, model training, and model evaluation. The results of the study show that the Random Forest method can provide good performance in classifying the number of malaria cases. The researcher used a dataset of 568 records, which was divided into 80% for training and 20% for testing. Of the total dataset, consisting of 25 sub-districts in Asahan Regency, 568 records were divided into 109 data (19.2%) for the low category, 334 data (58.8%) for the medium category, and 125 data (22.0%) for the high category. The most influential variable in this study is the health center (47.53%), followed by the sub-district (43.77%), age (2.18%), month (0.45%), and gender (0.45%). The Random Forest model was then evaluated using accuracy, precision, recall, and F1-Score metrics to measure the model's performance in classification. Based on the results of the study, the Random Forest model achieved an accuracy of 0.97%, precision of 0.97%, recall of 0.97%, and F1-Score of 0.97%.

Keywords: *Classification, Malaria, Random Forest, Accuracy.*