

**PERBANDINGAN KINERJA ALGORITMA C4.5 DAN RANDOM  
FOREST UNTUK KLASIFIKASI PENYAKIT *Tuberculosis* (TBC) DI  
PUSKESMAS JAJAG**

Jessica Cahya Ningrum<sup>1</sup>, Agung Nilogiri<sup>2</sup>, Qurrota A'yun<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email : [jesicacahya8@gmail.com](mailto:jesicacahya8@gmail.com)

**ABSTRAK**

Tuberkulosis adalah penyakit menular yang disebabkan oleh *Mycobacterium Tuberculosis*, yang dapat menyerang paru dan organ lainnya. Berdasarkan Global TB Report WHO 2021, Indonesia merupakan negara dengan beban TBC tertinggi ketiga di dunia. Data rekam medik pasien Tuberkulosis yang tersimpan di puskesmas dapat digunakan untuk pengembangan data menggunakan pola dari data sehingga memudahkan petugas mengambil keputusan terkait penyakit Tuberkulosis. Tujuan sistem ini bukan untuk menggantikan peran dokter, tetapi untuk memberikan rekomendasi atau kemungkinan hasil diagnosis berdasarkan gejala yang dialami pasien. Selain itu, sistem ini dapat memprediksi atau mendiagnosis penyakit TB sejak dini, sehingga membantu mengurangi penyebaran penyakit TB di masyarakat. Penelitian ini akan menggunakan Algoritma C4.5 dimana dalam pengimplementasiannya akan didapatkan pohon keputusan sebuah informasi untuk acuan dalam melakukan prediksi terhadap penyakit Tuberkulosis, kemudian akan dibandingkan dengan Random Forest yang merupakan metode klasifikasi untuk prediksi yang menghasilkan nilai akurasi yang baik dan lebih stabil. Kemudian dicari hasil akurasi, presisi, dan recall yang terbaik dari perbandingan tersebut. Hasil penelitian ini adalah hasil pengujian tertinggi pada metode Algoritma C4.5 menggunakan Fold Cross Validation dengan nilai  $K = 2$  pada Langkah Uji 1 dengan akurasi sebesar 71,2%, presisi sebesar 75% dan recall sebesar 63%. Sedangkan hasil pengujian tertinggi pada metode Random Forest menggunakan Fold Cross Validation dengan nilai  $K = 4$  pada Langkah Uji 2 dengan akurasi sebesar 73%, presisi sebesar 75,8% dan recall sebesar 68,1%. Kesimpulan dari penelitian ini adalah metode Random Forest memiliki kinerja yang lebih baik dari metode Algoritma C4.5.

**Kata Kunci :** *Tuberculosis, Klasifikasi, Algoritma C4.5, Random Forest*

# COMPARISON OF THE PERFORMANCE OF C4.5 ALGORITHM AND RANDOM FOREST FOR THE CLASSIFICATION OF *TUBERCULOSIS* (*TBC*) DISEASE AT THE JAJAG HEALTH CENTER

Jesica Cahya Ningrum<sup>1</sup>, Agung Nilogiri<sup>2</sup>, Qurrota A'yun<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

Email : jesicacahya8@gmail.com

## ABSTRAK

Tuberculosis is an infectious disease caused by *Mycobacterium tuberculosis*, which can attack the lungs and other organs. Based on the WHO Global TB Report 2021, Indonesia is the country with the third highest TB burden in the world. Medical record data of Tuberculosis patients stored in the health center can be used for data development using patterns from the data so that it makes it easier for officers to make decisions related to Tuberculosis disease. The purpose of this system is not to replace the role of doctors, but to provide recommendations or possible diagnostic results based on the symptoms experienced by patients. In addition, this system can predict or diagnose TB disease early, thereby helping to reduce the spread of TB disease in the community. This research will use the C4.5 Algorithm where in its implementation a decision tree will be obtained as a reference in making predictions for Tuberculosis disease, then it will be compared with Random Forest which is a classification method for prediction that produces good and more stable accuracy values. Then the best accuracy, precision, and recall results from the comparison are sought. The result of this study is the highest test result in the C4.5 Algorithm method using Fold Cross Validation with a value of  $K = 2$  in Test Step 1 with an accuracy of 71.2%, precision of 75% and recall of 63%. Meanwhile, the highest test result in the Random Forest method used Fold Cross Validation with a value of  $K = 4$  in Test Step 2 with an accuracy of 73%, a precision of 75.8% and a recall of 68.1%. The conclusion of this study is that the Random Forest method has better performance than the C4.5 Algorithm method.

**Keywords :** *Tuberculosis, Classification, C4.5 Algorithm, Random Forest*