

## ABSTRAK

Tuberkulosis ialah suatu penyakit menular yang disebabkan oleh *Mycobacterium Tuberculosis* yang dapat menyerang paru dan diluar paru. Untuk mendiagnosis penyakit TB dibutuhkan data yang lengkap agar mendapat hasil akurasi yang akurat. Dataset dalam penelitian ini didapat dari data pasien Tuberkulosis pada RSUD Syarifah Ambami Rato Ebhu Bangkalan. Maka dilakukannya penelitian ini untuk mengetahui implementasi imputasi missing data dengan metode *Particle Swarm Optimization* dalam penentuan nilai  $k$  optimal pada metode *k-Nearest Neighbor* serta untuk mengukur tingkat akurasi menggunakan klasifikasi *Long Short Term Memory*. Fitur yang dilakukan dalam penelitian ini menggunakan dataset asli yang terdiri dari 6 fitur yaitu umur, jenis kelamin, foto toraks, status HIV, riwayat diabetes dan lokasi anatomi. Pada penelitian ini menghasilkan beberapa teknik pengujian. Pengujian teknik imputasi data menggunakan optimasi  $k$  dengan *PSO* pada *k-NN* menghasilkan nilai  $k$  optimal yaitu  $k = 15$  dengan pemilihan parameter pengujian  $c1 = 1$  ,  $c2 = 1$ . Luaran yang dihasilkan (*k-NN*)-*PSO* berupa data lengkap yang dibagi kedalam 10 *fold cross validation*. Selanjutnya pengujian hasil klasifikasi dengan membandingkan *PSO-(k-NN)-LSTM* dengan (*k-NN*)-*LSTM*. Pada *PSO-(k-NN)-LSTM* menghasilkan tingkat akurasi lebih kecil yaitu 97.97% pada data training dan 100% pada data testing, sedangkan hasil akurasi tertinggi yang didapat  $k$  optimal tanpa optimasi dari (*k-NN*)-*LSTM* yaitu  $k = 3$  dengan hasil akurasi 98.30% pada data training dan 100% pada data testing. Dari tingkat akurasi yang diperoleh, bahwa hasil  $k$  optimal dari metode *k-NN+PSO* terbukti tidak selalu meningkatkan kinerja akurasi yang lebih signifikan.

**Kata Kunci:** Tuberkulosis, *Missing Data*, *Particle Swarm Optimization (PSO)*, *k-Nearest Neighbor(k-NN)*, *Long Short Term Memory (LSTM)*