

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tuberculosis yang selanjutnya disingkat TB adalah penyakit menular yang disebabkan oleh *Mycobacterium Tuberculosis*, yang dapat menyerang paru dan organ lainnya. (Kemenkes, 2016). TBC masih menjadi masalah kesehatan dan menempati peringkat 10 teratas penyebab kematian di dunia. Berdasarkan Global TB Report WHO 2021, Indonesia merupakan negara dengan beban TBC tertinggi ketiga di dunia. Diestimasikan terdapat 824 ribu kasus TBC baru pada tahun 2020 dengan angka kematian mencapai 93 ribu kasus atau setara dengan 11 kematian/jam.

Permasalahan yang berkembang adalah laju pertumbuhan pasien lebih cepat dibandingkan dengan jumlah dokter yang tersedia, hal ini jelas menjadi permasalahan yang besar karena pada dasarnya setiap manusia memiliki hak untuk mendapatkan pelayanan yang baik untuk penyakit yang dideritanya. Dalam konteks penanganan penyakit TB, seorang dokter akan sangat terbantu jika ada sistem yang dapat membantu mendiagnosis penyakit tersebut. 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.

Di Puskesmas Jajag Banyuwangi dalam periode tahun 2022, mendapatkan pasien *suspect* penyakit *Tuberculosis* sebanyak 350 dengan berbagai macam gejala yang berbeda outputnya sesuai hasil pemeriksaan dan hasil Lab yang dilakukan pasien *suspect*. Karena itu Puskesmas adalah pelayanan masyarakat yang dapat diartikan bahwa sistem pelayanan pertama yang dikunjungi masyarakat sebelum masyarakat datang ke Rumah Sakit atau mendapat rujukan. Oleh karena itu data rekam medik pasien yang tersimpan di puskesmas dapat digunakan untuk pengembangan data menggunakan klasifikasi sehingga dapat menciptakan informasi dan menemukan pola dari data sehingga memudahkan petugas mengambil keputusan terkait penyakit *Tuberculosis*.

Data mining adalah serangkaian langkah untuk menggali informasi atau pola yang terkandung dalam suatu basis data dengan menggunakan metode tertentu. Informasi yang diperoleh berasal dari ekstraksi dan analisis pola-pola data yang ada dalam basis data tersebut (Cholil et al., 2020). Proses ini bersifat semi-otomatis dan memanfaatkan teknik-teknik statistik, matematika, kecerdasan buatan, serta machine learning untuk mengekstrak dan mengidentifikasi pengetahuan yang berpotensi bermanfaat yang tersimpan dalam database besar. Data mining merupakan bagian dari proses KDD (Knowledge Discovery in Database), yang meliputi beberapa tahapan seperti pemilihan data, pra-pengolahan, transformasi, data mining, dan evaluasi hasil. Secara umum, KDD juga dikenal sebagai pangkalan data (Zai, 2022).

Banyak penelitian prediksi dengan teknik klasifikasi, diantaranya dilakukan oleh Kelvin & Zakarias (2022) dalam prediksi penyakit *Cerebrovascular* menggunakan Algoritma C4.5 dan Naïve Bayes, hasil dari penelitian tersebut Algoritma C4.5 memiliki performa yang lebih baik yaitu dengan tingkat akurasi sebesar 95% serta nilai presisi, recall dan f1-score masing-masing yaitu 90%, 95% dan 93%. Selain itu, penelitian lain yang dilakukan Lazuardi et al (2020) dalam prediksi penyakit tuberkulosis menggunakan Algoritma C4.5 dengan total 665 pasien *suspect* dan 7 *atribut*. Kemudian penelitian dari S. Sundaramurthy dan P. Jayavel (2020) dalam prediksi penyakit Rheumatoid Arthritis (RA) dan mempunyai akurasi rata-rata 84%.

Random Forest adalah metode ensemble learning yang diperkenalkan oleh Breiman pada tahun 2001. Metode ini menggabungkan beberapa pohon klasifikasi, di mana setiap pohon dibangun secara independen dengan bergantung pada vektor sampel acak dan distribusi yang sama untuk setiap pohon dalam hutan tersebut (Syukron & Subekti, 2018). Random Forest sudah banyak digunakan untuk penelitian dengan menggunakan teknik klasifikasi karena kinerjanya yang unggul dan sederhana, diantaranya seperti penelitian yang dilakukan oleh Sudrajat et al (2022) dalam klasifikasi penelitiannya menggunakan metode Random Forest menghasilkan akurasi yang paling tinggi sebesar 93.92%. kemudian penelitian dari Apriliah et al (2021) dalam sistem

informasi prediksi kemungkinan diabetes menggunakan algoritma klasifikasi Random Forest menghasilkan akurasi terbaik sebesar 97.88%.

Berdasarkan penjabaran di atas, dapat disimpulkan bahwa penelitian ini akan menggunakan Algoritma C4.5 dan Random Forest dalam prediksi penyakit *Tuberkulosis*. Algoritma C4.5 merupakan salah satu algoritma modern untuk melakukan data mining dimana dalam pengimplementasian Algoritma C4.5 akan didapatkan pohon keputusan sebuah informasi untuk acuan dalam melakukan prediksi terhadap penyakit *Tuberkulosis*. Random Forest merupakan metode klasifikasi untuk prediksi yang menghasilkan nilai akurasi yang baik, cara kerjanya yaitu dengan membangun beberapa pohon keputusan dan menggabungkannya untuk mendapatkan prediksi yang lebih stabil dan akurat. Oleh karena itu hasil klasifikasi menggunakan Algoritma C4.5 akan dibandingkan dengan hasil kerja menggunakan Random Forest dan mencari hasil akurasi, presisi, dan recall yang terbaik dari perbandingan tersebut.

1.2 Rumusan Masalah

Dari latar belakang yang sudah diuraikan diatas, maka rumusan masalah yang harus diselesaikan dalam penelitian ini yaitu : Berapa tingkat akurasi, presisi, dan recall *Algoritma C4.5* dan *Random Forest* untuk klasifikasi *Tuberkulosis* (TBC) berdasarkan gejala?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan peneliti yang akan dicapai dari tugas akhir ini adalah : Mengetahui tingkat akurasi, presisi, dan recall dari hasil klasifikasi pasien *Tuberkulosis* (TBC) dilihat dari gejala pasien menggunakan *Algoritma C4.5* dan *Random Forest*.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian yang sudah disebutkan diatas, maka manfaat dari hasil penelitian ini adalah : Memberikan alternatif model penelitian klasifikasi menggunakan *Algoritma C4.5* dan *Random Forest*, diharapkan dapat memberikan kemudahan untuk mengidentifikasi penyakit TBC dengan cepat.

1.5 Batasan Penelitian

Agar penyusunan penelitian ini tidak keluar dari pokok permasalahan yang dirumuskan, maka ruang lingkup pembahasan dibatasi pada :

1. Klasifikasi pasien *Tuberculosis* (TBC) menggunakan algoritma C4.5 dan random forest dengan parameter antara lain : batuk ≥ 2 Minggu, penurunan berat badan, demam, batuk darah dan nyeri dada.
2. Data yang digunakan adalah data pasien *Tuberculosis* secara acak yang diperoleh langsung dari Puskesmas Jajag Banyuwangi pada bulan Januari tahun 2023, dataset yang dipakai merupakan data tahun 2022 dari bulan Januari sampai Desember sebanyak 350 data suspect TBC.
3. Pemrograman yang dipakai dalam penelitian ini adalah Python.
4. Metode yang diimplementasikan pada penelitian ini yaitu Algoritma C4.5 dan Random Forest.

