

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Tuberkulosis merupakan penyakit menular yang disebabkan oleh infeksi bakteri *Mycobacterium Tuberculosis* (Aini, Ramadiani, & Hatta, 2017). Bakteri *Mycobacterium Tuberculosis* berkembang melalui lingkungan yang padat, daerah yang lembab, dan penularannya melalui udara (Budi & Tuntun, 2016). Spesifik organ tubuh yang diserang oleh bakteri ini yaitu paru, sehingga disebut Tuberkulosis Paru (TB paru) namun dapat juga menyerang organ tubuh lainnya seperti tulang, usus, kelenjar getah bening, dan kulit sehingga dapat disebut juga tuberkulosis luar paru (TB ekstra paru). Bentuk bakteri *Mycobacteri Tuberculosis* adalah batang dan bersifat tahan asam pada pewarnaan, sehingga disebut sebagai Basil Tahan Asam (BTA) (B. Mulyadi, 2011).

Berdasarkan Global TB Report 2018 diperkirakan insiden TBC di Indonesia mencapai 842.000 kasus dengan mortalitas 107.000 kasus. Dengan mengacu pada data tersebut maka Indonesia memiliki jumlah beban TBC tertinggi ketiga di dunia setelah India dan China. Sebagian besar kasus TBC terjadi di usia produktif yaitu antara usia 15 sampai usia 54 tahun, sehingga sangat mempengaruhi produktifitas dan berdampak pada sosial dan ekonomi (Husada, 2019).

Berdasarkan data rekam medis RSUD RAA Soewondo Pati bahwa pada tahun 2019 kasus TB Paru menduduki peringkat ke-6 dengan jumlah 5.192 pasien. Kemudian di tahun 2020 jumlah pasien TB Paru berkurang menjadi 3.409 pasien,

namun pada tahun 2020 ini jumlah kasus TB Paru menduduki peringkat pertama di RSUD RAA Soewondo Pati. Hal ini dikarenakan adanya pandemik virus Covid-19 yang menyebabkan minat pasien RSUD RAA Soewondo Pati menjadi berkurang karena RSUD RAA Soewondo Pati merupakan Rumah Sakit rujukan pertama di kabupaten Pati bagi pasien yang terkena virus Covid-19, sehingga masyarakat Kabupaten Pati takut untuk menjalani pemeriksaan.

Gejala Tuberkulosis dapat dibagi menjadi dua yaitu gejala umum dan gejala khusus yang timbul sesuai dengan organ tubuh yang terlibat. Gambaran secara klinis tidak terlalu khas terutama pada kasus baru, sehingga cukup sulit untuk mendiagnosa secara klinik. Gejala secara umum yaitu batuk-batuk selama lebih dari 3 minggu dan dapat disertai dengan darah, demam yang tidak terlalu tinggi tapi berlangsung lama dan biasanya dirasakan pada saat malam hari disertai keringat malam, demam seperti influenza yang bersifat hilang timbul, penurunan nafsu makan disertai dengan turunnya berat badan, dan mengalami perasaan tidak enak serta merasa lemah. Sedangkan gejala khusus tergantung dari organ tubuh mana yang terkena, bila terjadi sumbatan sebagian bronkus (saluran yang menuju ke paru-paru) akibat penekanan kelenjar getah bening yang membesar, akan menimbulkan suara “mengi”, suara nafas melemah yang disertai sesak, bila mengenai tulang maka akan terjadi gejala seperti infeksi dan dapat membentuk saluran yang bermuara pada kulit di atasnya dan akan mengeluarkan cairan nanah (Werdhani, 2002).

Infeksi TB terjadi karena inhalasi *droplet nuclei* yang mengandung kuman tuberkulosis. Setelah terpapar kuman TB ada empat keadaan yang bisa terjadi yaitu pertama tidak terjadi infeksi (ditandai dengan tes kulit tuberkulin yang negatif),

kedua terjadi infeksi kemudian menjadi TB yang aktif (TB primer), ketiga menjadi TB laten dimana mekanisme imun mencegah progresivitas penyakit menjadi TB aktif dan keempat menjadi TB laten tetapi kemudian terjadi reaktivasi dan berkembang menjadi TB aktif dalam beberapa bulan sampai beberapa tahun kemudian. Infeksi TB laten ini didefinisikan sebagai kondisi seseorang yang terinfeksi *M.tuberculosis* tetapi saat ini orang tersebut tidak sakit, tidak mempunyai gejala / *asymptomatic* dan gambaran foto toraks normal. Kira – kira 5% - 10% dari orang dengan infeksi laten, akan terjadi reaktivasi dan menjadi TB aktif (Martin & Hasibuan, 2010).

Salah satu upaya untuk mengendalikan TB yaitu dengan menurunkan angka kesakitan dan kematian yang diakibatkan oleh TB dengan cara memutuskan rantai penularan. Upaya yang dapat diterapkan adalah teratur dalam mengkonsumsi obat serta melaksanakan cara pengobatan yang disarankan oleh dokter sehingga dapat meningkatkan angka kesembuhan (Dewi, 2021). Maka dari itu, perlu adanya aplikasi yang dapat digunakan untuk mempermudah prediksi penyakit TBC. Salah satu teknik yang dapat digunakan untuk melakukan prediksi penyakit adalah klasifikasi (Febriyanto & Kurniawan, 2018).

Banyak penelitian prediksi penyakit dengan teknik klasifikasi, diantaranya dilakukan oleh (Darsyah, 2014) dalam klasifikasi tuberkulosis dengan pendekatan metode Supports Vector Machine (SVM). Hasil penelitian tersebut mendapatkan tingkat akurasi dalam klasifikasi sebesar 98%. Selain itu, penelitian lain yang dilakukan (Febriyanto & Kurniawan, 2018) dengan melakukan prediksi penyakit tuberkulosis menggunakan metode C4.5 dengan total 280 kasus menunjukkan bahwa

nilai rata – rata maksimal yang didapatkan mencapai 100%. Algoritma C4.5 banyak digunakan dalam klasifikasi karena mudah diinterpretasikan, tingkat akurasi yang dapat diterima, dan efisien. Sejumlah pendekatan klasifikasi telah dikembangkan selama beberapa dekade terakhir (Lu & Weng, 2007) dan termasuk juga dengan pendekatan data mining (Otukey & Blaschke, 2010).

Data mining adalah serangkaian proses untuk menggali informasi baru dengan cara mencari sebuah pola atau aturan-aturan tertentu dari data yang berjumlah besar (Huda, 2010). Data mining mempunyai lima peran utama yaitu estimasi, prediksi, klasifikasi, klaster dan asosiasi (Tan, Steinbach, & Kumar, 2006). Klasifikasi merupakan salah satu teknik yang terdapat pada data mining. Tujuan dari teknik klasifikasi data mining adalah untuk memprediksi kelas target secara akurat dengan menggunakan variabel-variabel terkait (Yusa, Utami, & Luthfi, 2107). Untuk mencapai tujuan tersebut, proses klasifikasi akan menghasilkan model yang bisa memprediksi kelas, mengkategorikan suatu objek dan mampu membedakan data ke dalam kelas-kelas yang berbeda (Aradea, Z, & A, 2011). Dimana data yang digunakan mempunyai kelas label atau target. Sehingga algoritma untuk menyelesaikan masalah klasifikasi dikategorikan kedalam Supervised Learning (Darujati & Gumelar, 2012).

Berdasarkan penelitian Maharjan tentang perbandingan algoritma C4.5 dan Naïve Bayes, dimana dapat disimpulkan bahwa algoritma C4.5 memiliki tingkat akurasi yang lebih baik dibandingkan dengan metode naïve bayes (Maharjan, 2014). Raharjo pada tahun 2017 juga melakukan penelitian komparasi algoritma C4.5, Neural Network, dan SVM dengan teknik PSO, hasil yang didapat adalah

Algoritma C4.5 memiliki tingkat akurasi tertinggi diantara ketiga algoritma tersebut (Raharjo, 2017). Selain itu pada tahun 2013 Amini melakukan penelitian komparasi antara algoritma C4.5 dengan K-Nearest Neighbor, hasil yang didapatkan setelah menerapkan algoritma dan melakukan perbandingan dan perhitungan akurasi, hasil yang didapatkan adalah algoritma C4.5 melampaui K-Nearest Neighbor dalam hal akurasi dan presisi dengan perbedaan kecil (Amini et al., 2013). Komparasi algoritma C4.5, Naïve Bayes, dan k-NN yang dilakukan oleh Jadhav dan Channe menghasilkan bahwa algoritma C4.5 memiliki hasil akurasi yang lebih akurat dan memiliki tingkat kesalahan yang lebih sedikit dan lebih mudah untuk digunakan (Jadhav & Channe, 2016). Yusa dan Utami pada tahun 2017 melakukan perbandingan algoritma klasifikasi decision tree ID3, C4.5, dan CART menggunakan dataset yang terdiri dari 55 atribut dengan 101.766 records, hasil yang didapat yaitu algoritma C4.5 memiliki nilai akurasi yang paling tinggi (Yusa et al., 2107).

Berdasarkan penelitian diatas algoritma C4.5 sering digunakan untuk masalah pengklasifikasian (Widiastuti, 2012). Karena mudah dipahami dan bisa lebih cepat untuk melakukan proses klasifikasi, salah satu kelebihanya yaitu dapat mengubah daerah pengambilan keputusan yang sebelumnya kompleks dan global menjadi lebih sederhana dan spesifik (Setiawati, Taufik, Jumadi, & Z, 2016). Algoritma C4.5 bekerja mencari setiap tahap atribut dan membaginya ke dalam bagian terbaik dari kelas tersebut (Andriani, 2013).

Algoritma C4.5 adalah algoritma yang digunakan untuk membentuk pohon keputusan (Fatoni, 2006). Algoritma C4.5 adalah pengembangan dari algoritma

ID3 dan J. Ross Quinland adalah yang pertama kali mengembangkannya (Andriani, 2012). Pengembangannya dilakukan untuk mengatasi missing data, data continue, dan pruning (Shita & Marliani, 2013). Perbaikan yang membedakan algoritma C4.5 dari ID3 adalah dapat menangani numeric attributes, missing values, dan noisy data (Raharja, 2014). Algoritma C4.5 memiliki beberapa kelebihan yaitu: dapat membuat pohon keputusan yang mudah dipahami, memiliki tingkat akurasi yang tinggi, efektif dalam menangani atribut bertipe diskret dan numerik (Kamagi & Hansun, 2014). Algoritma C4.5 akan digunakan untuk mengklasifikasikan dan menghitung tingkat akurasi data pasien penyakit tuberkulosis paru pada pasien RSUD RAA Soewondo Pati.

1.2

1.3 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah yang akan dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana gambaran umum data rekam medis pasien Tuberkulosis RSUD RAA Soewondo Pati?
2. Bagaimana hasil klasifikasi gejala penyakit tuberkulosis (TB) menggunakan Algoritma C4.5?
3. Bagaimana tingkat keakuratan C4.5 dalam klasifikasi gejala penyakit tuberculosis (TB)?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan, tujuan yang ingin dicapai melalui penelitian ini adalah:

1. Mengetahui bagaimana gambaran umum data rekam medis pasien Tuberkulosis RSUD RAA Soewondo Pati.
2. Mengetahui hasil klasifikasi gejala penyakit tuberkulosis (TB) menggunakan Algoritma C4.5.
3. Mengetahui tingkat keakuratan dari Decision Tree C4.5 dalam klasifikasi gejala penyakit tuberkulosis (TB).

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penulisan ini adalah:

1. Menambah pengetahuan tentang pengaplikasian data mining menggunakan algoritma Decision Tree C4.5 pada pengklasifikasian gejala penyakit tuberkulosis (TB).
2. Penelitian ini diharapkan dapat digunakan sebagai bahan bacaan untuk menambah ilmu dan pengetahuan bagi pihak yang memerlukan

1.6 Batasan Penelitian

Dalam suatu penelitian hendaknya diberi pembatasan agar masalah yang akan diteliti tidak terlalu menyimpang. Batasan masalah dalam penelitian ini yaitu:

1. Data yang diolah hanya data yang berhubungan dengan penyakit Tuberkulosis (TB) dengan variabel penelitian yaitu batuk, demam, badan lemas, sesak nafas, nyeri dada, batuk dahak, nafsu makan, hasil laboratorium, dan hasil rontgen, yang bersumber dari Rumah Sakit Umum Daerah (RSUD) Soewondo Pati yang merupakan data rekam medis dari pasien Tuberkulosis (TB).
2. Algoritma yang digunakan adalah Decision Tree C4.5.