

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Tuberkulosis (TBC) yang juga dikenal dengan singkatan TBC merupakan penyakit menular yang menyebabkan masalah kesehatan terbesar kedua di dunia setelah HIV. Penyakit ini disebabkan oleh basil dari bakteri *Mycobacterium tuberculosis*. Tuberkulosis sendiri dapat menyerang bagian tubuh manapun seperti sistem saraf pusat, jantung, kelenjar getah bening, tetapi yang paling umum terjadi adalah infeksi tuberkulosis pada paru-paru. TBC tetap jadi penyakit infeksi yang paling mematikan sedunia dibanding dengan penyakit infeksi lainnya. Indonesia sendiri termasuk lima besar negara dengan jumlah pengidap TBC terbanyak di Asia Tenggara dengan jumlah pengidap yang mencapai 305.000 jiwa pada 2012 [1]. Penularan penyakit ini dapat terjadi melalui percikan air liur dari batuk dan bersin orang yang mengidap TBC kepada orang yang memiliki kekebalan tubuh yang cukup rendah. Pasien penyakit TBC di RSUD Syarifah Ambami Rato Ebu Bangkalan tergolong masih banyak. Pada data asli yang didapat dari RSUD Syarifah Ambami Rato Ebu Bangkalan pasien terdiagnosa penyakit TBC di tes dengan menggunakan 27 fitur dalam pengklasifikasian penyakit TBC yang dialami pasien, dari 27 fitur ini menjadi latar belakang penelitian ini dicari berapakah fitur yang paling relevan dalam mendiagnosa penyakit TBC ini yaitu dengan menggunakan Seleksi fitur.

Seleksi fitur merupakan teknik untuk mengurangi dimensi fitur. Pengurangan dimensi tersebut dilakukan untuk mendapatkan fitur-fitur yang relevan dan tidak berlebihan sehingga dapat mempercepat proses klasifikasi dan dapat meningkatkan akurasi dari algoritma klasifikasi [2]. Metode seleksi fitur yang digunakan dalam penelitian ini adalah *Information Gain*. Metode tersebut akan melakukan proses komputasi untuk mendapatkan fitur-fitur yang paling berpengaruh terhadap dataset penyakit TBC.

Penelitian terhadap penyakit TBC sudah banyak ada seperti pada penelitian sebelumnya oleh Mikha Dayan Sinaga yang berjudul “Implementasi Theorema

Bayes Untuk Mendiagnosis Penyakit Tuberkulosis (TBC)” [3], kekurangan pada penelitian ini tidak menghitung akurasi dari system identifikasi penyakit yang telah dilakukan. Lalu seperti pada penelitian lainnya oleh Arnes Yuli Vandika dan Ahmad Cucus yang berjudul “Sistem Deteksi Awal Penyakit TBC Dengan Metode CBR” [4]. Penelitian ini masih berupa prototype dan pada penelitian ini menghasilkan ranking dari kandidat yang dapat dijadikan pasien yang dapat mendapat pola pengobatan atau resep tertentu.

Lalu pada penelitian yang sama mengenai penyakit TBC dengan metode *KNN* yang dilakukan oleh Emanuel Tes Atok, Derwin R.Sina, dan Dony M.Sihotang dengan judul “Implementasi Case Bared Reasoning Untuk Mendiagnosis Penyakit Tuberkulosis Menggunakan Algoritma K-Nearest Neighbor” tetapi pada penelitian ini hanya menggunakan 51 kasus dan mengidentifikasi 3 jenis penyakit TBC yaitu Paru-paru, Milier, dan Meningitis. Pada penelitian ini dilakukan 3 skenario pengujian dan berdasarkan ketiga skenario pengujian ini dapat diambil kesimpulan bahwa makin banyak data latih dalam basis kasus, maka semakin baik nilai akurasi yang diperoleh [5].

Berdasarkan uraian Latar Belakang ini menjadi alasan penulis melakukan penelitian ini dengan judul “Seleksi Fitur *Information Gain* Untuk Klasifikasi Penyakit Tuberculosis (TBC) Menggunakan Metode *Weighted K-Nearest Neighbor (WKNN)*”. Metode ini dipilih karena pada penelitian sebelumnya yang dilakukan oleh Sigit Bagus Setiawan, Adiwijaya, dan Mohamad Syahrul Mubarak dengan judul “Klasifikasi Topik Berita Berbahasa Indonesia menggunakan *Weighted K-Nearest Neighbor*” [6]. Kesimpulan dari penelitian ini adalah bahwa pemilihan pemberian bobot pada *weighted k nearest neighbor* cukup berpengaruh atas performansi sistem.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Tingginya jumlah pasien yang terjangkit penyakit TBC di Indonesia dan banyaknya variable atau fitur pada pasien TBC.

1.2.2 Pertanyaan Penelitian

1. Berapa Akurasi terbaik yang dihasilkan metode *WKNN* tanpa Seleksi Fitur *Information Gain*?
2. Berapa Akurasi terbaik yang dihasilkan metode *WKNN* dengan Seleksi Fitur *Information Gain*?
3. Lebih baik manakah menggunakan seleksi fitur atau tanpa seleksi fitur dalam pengklasifikasian ini?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah mencari akurasi dan jumlah fitur yang terbaik dengan menerapkan metode *WKNN* untuk pengklasifikasiannya dan *Information Gain* untuk Seleksi Fiturnya.

1.3.2 Manfaat Penelitian

Penelitian ini bermanfaat bagi peneliti berikutnya akan mengetahui tingkat akurasi yang dihasilkan dengan metode *WKNN* dan peneliti juga dapat mengetahui berapa jumlah fitur yang terbaik digunakan dalam pengklasifikasian penyakit TBC ini .

1.4 Batasan Masalah

Batasan dari system yang akan dibuat yaitu :

1. Studi kasus pada penelitian ini dilakukan di RSUD Syarifah Ambami Rato Ebu Bangkalan dengan menggunakan 100 data pasien.
2. Parameter yang digunakan untuk pengklasifikasian ini adalah hasil hitung jarak terdekat/terkecil menggunakan metode *WKNN* dengan 27 fitur pada setiap pasien
3. Seleksi Fitur yang digunakan dalam metode ini adalah *Information Gain*.

4. Pengklasifikasian pada penelitian ini adalah berdasarkan 5 diagnosa yaitu TBC Paru-paru, TBC Kelenjar, TBC Usus, TBC Otak, dan bukan penyakit TBC

1.5 Sistematika Proposal

BAB 1 PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini dijelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut diantaranya adalah penelitian-penelitian terkait Klasifikasi, TBC, metode *WKNN*, dan *Information Gain*.

BAB III METODE USULAN

Pada bab ini dijelaskan mengenai rancangan dari sistem yang akan dibuat dengan mengumpulkan data dan analisa masalah mulai dari kebutuhan software dan hardware, serta alur dari sistem yang akan dikembangkan.

REFERENSI

Berisi tentang daftar jurnal-jurnal atau referensi yang berhubungan dengan Seleksi Fitur *Information Gain* Untuk Klasifikasi Penyakit Tuberculosis(TBC) Menggunakan Metode *Weighted K-Nearest Neighbor*.