

BAB I

PENDAHULUAN

1.1 Latar Belakang

Tuberkulosis merupakan penyakit menular akut yang disebabkan oleh bakteri *Mycobacterium tuberculosis* (Buntuan, 2014). Penyakit ini ditularkan melalui udara, yaitu berasal dari butiran ludah seseorang yang terinfeksi bakteri tersebut yang keluar ketika orang tersebut batuk, bersin, atau meludah (Konstantinos, 2010). Tuberkulosis sendiri adalah salah satu dari sepuluh penyebab kematian di dunia dan menjadi penyakit paling mematikan yang disebabkan oleh infeksi. Pada tahun 2018 saja, Tuberkulosis telah mengakibatkan kematian kurang lebih 1,4 juta kematian, jauh lebih tinggi dibandingkan HIV yang hanya bertanggungjawab atas kurang lebih 730.000 kematian pada tahun yang sama (Bhargavaa dan Bhargavac, 2020). Menurut data yang dihimpun WHO tahun 2018, Indonesia menempati urutan ketiga jumlah kasus Tuberkulosis setelah India dan Cina di Asia, yaitu 824 ribu kasus dalam tahun tersebut saja. Jika dilihat dari proposinrya, Indonesia berada pada posisi kedua se-Asia, yaitu 316 kasus per 100 ribu orang, cukup tipis dibandingkan Myanmar dengan proporsi 338 kasus per 100 ribu orang. Jika dilihat dari angka kematian pada tahun yang sama, Indonesia menjadi salah satu negara dengan kematian terbesar, yaitu 93 ribu kematian dengan 35 kematian per 100 ribu orang, hanya berbeda sedikit dengan Myanmar yang mencapai 39 kematian per 100 ribu orang (WHO, 2019).

Salah satu cara untuk menanggulangi penyakit Tuberkulosis adalah dengan pendeteksian kasus sedini mungkin. Hal ini dimaksudkan untuk mencegah penularan dari penderita terhadap orang lain dan meningkatkan keefektifan pengobatan bagi sang penderita (Suharyo, 2013). Metode utama untuk mendeteksi Tuberkulosis adalah dengan mendeteksi Bakteri Tahan Asam di dalam dahak. Sayangnya, metode ini memiliki sensitivitas yang rendah dan hasilnya tergantung pada teknisi. WHO sendiri mencatat bahwa secara global, hanya 45,6% dari kasus baru Tuberkulosis yang terdeteksi positif melalui metode ini (Nour-Neamatollahi

et al, 2018). Selain itu, waktu pengujian yang berselang 2 sampai 6 pekan membuat seringkali pengobatan dilakukan sebelum hasil pengujian didapatkan.

Akibat hal tersebut, diperlukan bantuan komputasi untuk dapat mengidentifikasi *Mycobacterium tuberculosis* pada citra hapusan dahak yang terduga Tuberkulosis. Salah satu metode yang mampu membantu mendeteksi *Mycobacterium tuberculosis* pada citra hapusan dahak adalah Pengolahan Citra Digital (Digital Image Processing). Pengolahan Citra Digital adalah istilah yang digunakan untuk penggunaan komputer dalam mengolah citra digital dengan mengikuti sebuah algoritma tertentu. Citra digital sendiri adalah gambar yang terdiri dari elemen-elemen yang disebut piksel, masing-masing direpresentasikan dengan suatu nilai diskrit tertentu sebagai nilai intensitas warna atau tingkat keabuan gambar tersebut (Gonzales, 2018).

Sebelumnya, sudah ada beberapa penelitian yang membahas mengenai pendeteksian *Mycobacterium tuberculosis* pada citra hapusan dahak penderita Tuberkulosis. Makkapati (2009) meneliti tentang pendeteksian *Mycobacterium tuberculosis* menggunakan metode segmentasi dan klasifikasi berdasarkan warna dan bentuk bakteri tersebut pada citra hapusan dahak yang diberi pewarnaan ZN. Meski tidak mencantumkan perhitungan hasilnya, namun penulis mencantumkan hasil pengolahan citranya dan berhasil mengidentifikasi *Mycobacterium tuberculosis* pada citra yang digunakan. Arisgraha (2012) kemudian meneliti mengenai sistem deteksi digital *Mycobacterium tuberculosis* pada citra hapusan dahak menggunakan metode Neural Network dan menghasilkan tingkat akurasi 77,5% dengan tingkat kesalahan kurang dari 5%. Foster (2014) sendiri mengambil pendekatan yang berbeda. Alih-alih meneliti tentang bagaimana mendeteksi *Mycobacterium tuberculosis*, Foster mencoba meneliti mengenai pendeteksian jumlah bakteri penyebab Tuberkulosis tersebut menggunakan segmentasi pada citra Positron Emission Tomography (PET). Hanya saja, subjek pada penelitian mereka adalah hewan. Akurasi yang dihasilkan cukup tinggi, yaitu sebesar $91,25 \pm 8,01\%$. Mithra (2018) sendiri menggunakan metode Gaussian Fuzzy Neural Network (GFNN) dalam meneliti mengenai keberadaan *Mycobacterium*

tuberculosis di dalam citra mikroskopis dahak dan akurasi yang didapatkan menggunakan metode tersebut adalah 91,38%. Berdasarkan penelitian sebelumnya, masih sedikit penelitian yang mengambil banyaknya bakteri pada citra hapusan dahak. Padahal, banyaknya bakteri pada citra hapusan dahak menunjukkan tingkat keparahan Tuberkulosis yang diderita. Oleh karena itu, peneliti pada penelitian ini akan menghitung banyaknya *Mycobacterium tuberculosis* untuk bisa dijadikan acuan mengenai tingkat keparahan penderitanya berdasarkan citra hapusan dahak miliknya.

Pendeteksian dan perhitungan *Mycobacterium tuberculosis* selama ini masih dilakukan secara manual, sehingga memerlukan waktu yang cukup lama dan tenaga yang cukup banyak. Atas alasan ini, peneliti mengusulkan untuk membuat sebuah pendekatan pemodelan statistika yang dapat mempersingkat dan memperingan perhitungan bakteri tersebut. Salah satu pemodelan yang dapat dilakukan adalah memodelkan banyaknya bakteri yang terdeteksi pada penderita Tuberkulosis dengan hasil dari pengolahan citra hapusan dahaknya. Metode statistika yang tepat dalam menganalisis hubungan antara respon dan prediktor adalah analisis regresi. Analisis regresi adalah merupakan suatu teknik analisis yang digunakan untuk memperkirakan hubungan fungsional antara variabel terikat dengan satu atau lebih variabel bebas. Analisis regresi dapat dilakukan dengan pendekatan parametrik dan nonparametrik. Pada pemodelan parametrik, fungsi regresi diasumsikan memiliki bentuk tertentu, seperti linier, kuadratik, dan eksponensial, sehingga yang diestimasi adalah sejumlah parameternya. Sebaliknya, fungsi regresi pada pemodelan nonparametrik hanya diasumsikan mulus (*smooth*) atau termuat dalam suatu fungsi ruang tertentu sehingga memiliki fleksibilitas yang lebih tinggi (Eubank, 1998).

Dalam mendapatkan hasil pendeteksian *Mycobacterium tuberculosis* melalui citra hapusan dahak, perlu dilakukan transformasi dari gambar menjadi angka. Hal ini diperlukan karena variabel yang digunakan pada analisis regresi haruslah berupa numerik, sehingga perlu dilakukannya proses pengubahan gambar menjadi angka terlebih dahulu. Namun, hasil pengolahan citra memiliki dimensi yang

sangat besar sehingga akan menyulitkan dalam melakukan analisis regresi nantinya. Oleh karena itu diperlukan suatu metode untuk bisa mereduksi dimensi. Dalam penelitian ini, metode untuk mereduksi dimensi yang akan digunakan adalah *Genetic Algorithms* (GA). Metode ini mampu diterapkan dalam pemilihan variabel prediktor terutama saat berhadapan dengan data yang terdiri atas ratusan atau bahkan ribuan variabel prediktor (Sartono, 2010).

Citra hapusan dahak yang telah diubah menjadi angka dan direduksi menggunakan GA kemudian dijadikan variabel prediktor dalam penelitian ini. Variabel responnya sendiri adalah banyaknya *Mycobacterium tuberculosis* yang akan dianalisis menggunakan pendekatan regresi nonparametrik. Model yang umum digunakan dalam meregresikan variabel respon berupa *count data* adalah regresi Poisson, namun seringkali pada data yang digunakan terjadi pelanggaran asumsi regresi Poisson, yaitu overdispersi. Overdispersi merupakan sebutan bagi keadaan data yang memiliki nilai variansi lebih besar dibandingkan rata-ratanya. Keadaan ini merupakan pelanggaran bagi asumsi regresi Poisson karena pada regresi Poisson, variabel respon yang digunakan haruslah berdistribusi Poisson, yang dicirikan dengan nilai rata-rata dan variansi yang sama. Pelanggaran ini akan mengakibatkan kurang baiknya hasil yang didapatkan sehingga mengurangi tingkat keakuratannya. Beberapa penelitian telah dilakukan untuk mengatasi masalah overdispersi, salah satunya adalah penelitian oleh Lawless (1987) yang menyarankan untuk menggunakan regresi Binomial Negatif agar mengurangi tingkat kesalahan yang didapatkan. Temuan ini kemudian dikonfirmasi oleh Hilbe (2007) dan Chamidah *et al* (2021) bahwasanya penggunaan regresi binomial negatif benar-benar mampu mengatasi masalah overdispersi pada *count data*. Khusus dalam penelitian Chamidah *et al* (2021), metode yang digunakan adalah pendekatan regresi nonparameterik, berdasarkan estimator linier lokal, pada regresi binomial negatif sehingga hasil yang didapatkan memiliki fleksibilitas yang lebih tinggi.

Untuk estimatornya sendiri, terdapat beberapa estimator yang dapat digunakan untuk mengestimasi kurva regresi non parametrik, salah satunya adalah

polinomial lokal. Polinomial lokal adalah perluasan dari estimator linier lokal, yaitu dengan mengubah orde 1 pada linier lokal menjadi lebih dari 1. Penggunaan orde berderajat tinggi biasanya menghasilkan estimasi dengan bias yang lebih kecil, namun menghasilkan koefisien yang lebih banyak. Oleh karena itu, orde dari estimator yang digunakan dapat disesuaikan berdasarkan pola data yang digunakan. Jika datanya menunjukkan pola yang tidak monoton naik atau turun, maka orde yang digunakan sebaiknya bernilai 2 atau lebih (Liang & Cheng, 2005).

Estimator ini dipilih karena proses pembuatan modelnya yang tidak memerlukan spesifikasi fungsi tertentu. Untuk pengestimasiannya, dibutuhkan nilai *bandwidth* dan orde polinomial lokal tertentu. Selain itu, metode polinomial lokal sangat fleksibel, sehingga ideal untuk pemodelan variabel yang kompleks dan sulit diketahui model teoritisnya. Kedua kelebihan ini, dikombinasikan dengan kesederhanaannya, membuat polinomial lokal cocok digunakan sebagai estimator pada penelitian ini.

Berdasarkan uraian di atas, dalam skripsi ini akan diteliti pemodelan banyaknya *Mycobacterium tuberculosis* pada citra hapusan dahak penderita Tuberkulosis menggunakan model regresi binomial negatif dengan pendekatan regresi non parametrik berdasarkan estimator polinomial lokal dengan reduksi dimensi yang digunakan adalah *Genetic Algorithm*.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang sebelumnya, permasalahan yang akan dibahas pada penelitian ini adalah sebagai berikut:

1. Bagaimana mengolah data citra hapusan dahak penderita Tuberkulosis menjadi data numerik?
2. Bagaimana mereduksi dimensi data hasil pengolahan citra penderita Tuberkulosis menggunakan *Genetic Algorithm*?
3. Bagaimana mengestimasi dan menganalisis model data hasil reduksi dimensi pengolahan citra penderita Tuberkulosis menggunakan pendekatan

regresi binomial negatif non parametrik berdasarkan estimator polinomial lokal?

1.3 Tujuan Penelitian

Berdasarkan permasalahan yang telah diuraikan, didapatkan tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengolah data citra hapusan dahak penderita Tuberkulosis menjadi data numerik.
2. Mereduksi dimensi data hasil pengolahan citra penderita Tuberkulosis menggunakan *Genetic Algorithm*.
3. Mengestimasi dan menganalisis model data hasil reduksi dimensi pengolahan citra penderita Tuberkulosis menggunakan pendekatan regresi binomial negatif non parametrik berdasarkan estimator polinomial lokal.

1.4 Manfaat Penelitian

Hasil penelitian ini diharapkan membawa manfaat berupa:

1. Mahasiswa dapat menambah wawasan mengenai pengolahan citra, reduksi dimensi menggunakan *Genetic Algorithm*, dan estimasi model regresi binomial negatif non parametrik berdasarkan estimator polinomial lokal.
2. Program Studi menghasilkan lulusan yang dapat menggunakan dan mengembangkan ilmu statistika untuk menyelesaikan masalah di berbagai bidang kehidupan.
3. Instansi kesehatan mendapatkan bahan pertimbangan untuk pemeriksaan lebih lanjut mengenai Tuberkulosis dan metode yang lebih cepat, akurat, dan efisien.