

1. Pendahuluan

Acquired Immune Deficiency Syndrome (AIDS) adalah sekumpulan gejala yang timbul karena turunnya kekebalan tubuh yang disebabkan infeksi oleh HIV (*Human Immunodeficiency Virus*). Meskipun cenderung fluktuatif, data kasus HIV AIDS di Indonesia terus meningkat dari tahun ke tahun. Selama sebelas tahun terakhir, jumlah kasus HIV di Indonesia mencapai puncaknya pada tahun 2019, yaitu sebanyak 50.282 kasus. Berdasarkan data WHO tahun 2019, terdapat 78% infeksi HIV baru di regional Asia Pasifik sedangkan untuk kasus AIDS tertinggi terjadi pada tahun 2013, yaitu sebanyak 12.214 kasus. Berdasarkan data dari Ditjen P2P tahun 2019, ada 5 provinsi dengan jumlah kasus HIV dan AIDS terbanyak yaitu Jawa Timur, DKI Jakarta, Jawa Barat, Jawa Tengah, dan Papua.

Sedangkan presentase HIV dan AIDS berdasarkan jenis kelamin pada tahun 2019 dilaporkan pada HIV laki-laki sebesar 64,5% dan perempuan 35,5%, sedangkan pada AIDS laki-laki 68,6% dan perempuan 31,4%. Case Fatality Rate (CFR) atau jumlah kematian AIDS di Indonesia sejak tahun 2005 sampai tahun 2019 terus mengalami penurunan yang disebabkan oleh upaya pengobatan AIDS di Indonesia secara terus menerus[1].

Deteksi TBC pada penderita HIV sangatlah penting. ini berguna untuk membantu penderita HIV mendapat pengobatan jika dapat dideteksi. Sejauh ini biasanya untuk mendeteksi HIV adalah dengan cara tes darah yang meliputi tes antibodi, tes antigen, hitung CD4 dan pemeriksaan *viral load* (HIV RNA). Kemudian untuk mendeteksi TBC bisa dengan cara tes rontgen, yaitu tes dengan menggunakan sinar X-ray. Setiap pendeteksian mempunyai kekurangan dan kelebihan. Kekurangan deteksi TBC secara konvensional adalah citra foto yang masih kabur, kurang kuat, kontras, dan sebagainya. Untuk itu, salah satu alternatif yang dapat digunakan untuk mendeteksi TBC pada penderita HIV adalah dengan implementasi metode *machine learning* pada data microarray[2].

Microarray adalah *lab-on-a-chip* multipleks. Ini adalah susunan dua dimensi pada substrat padat biasanya slide kaca atau sel film tipis silikon yang menguji sejumlah besar bahan biologis menggunakan metode pemrosesan dan deteksi miniatur, multipleks dan paralel. *Microarray* merupakan sebuah microchip yang berisi data pemetaan genom manusia. Untuk membuat *microarray* yang dapat berjalan dengan baik, dibutuhkan sebuah metode agar dapat membangun program yang dapat mendeteksi HIV[3].

Tahun 2020, Sajib Kabiraj melakukan pendeteksian kanker payudara menggunakan metode algoritma Random Forest dengan data microarray yang menghasilkan akurasi 74,73% dan 73,63%[4]. Di tahun yang sama Prihandoko juga membuat model prediksi kesehatan kota menggunakan Random Forest dengan metode klasifikasi untuk mengembangkan model yang tepat untuk prediksi dan analisis indeks kesehatan suatu kota. Model prediksi tersebut menghasilkan akurasi 94,6%[5].

Pada tahun 2021, Vishan Kumar Gupta melakukan prediksi kasus terkonfirmasi COVID 19 meninggal dan sembuh di India menggunakan metode Random Forest. Pada penelitian tersebut prediksi kasus yang dikonfirmasi, kematian, dan sembuh pada dataset pengujian, Random Forest menghasilkan akurasi 83,54%, 72,79%, dan 81,27%[6].

Pada tugas akhir ini peneliti mempunyai tujuan untuk mencari model prediksi dan hasil terbaik pada setiap fitur Random Forest untuk deteksi Tuberculosis pada penderita HIV dengan menggunakan microarray.

Rumusan masalah

1. bagaimana proses seleksi fitur pada Random Forest?
2. bagaimana implementasi Random Forest untuk prediksi TB pada penderita HIV dengan menggunakan microarray?
3. bagaimana performa model Random Forest?

Tujuan penelitian

1. melakukan proses seleksi fitur pada Random Forest
2. melakukan implementasi Random Forest untuk prediksi TB pada penderita HIV dengan menggunakan microarray
3. mencari performa model Random Forest?