

1. Pendahuluan

Latar Belakang

Dengue adalah penyakit menular yang sangat endemik di negara-negara tropis dan dengan cepat menjadi beban global. Penyakit ini disebabkan oleh salah satu dari 4 serotipe virus dengue yang tergolong *Arthropod-Borne Virus*, genus *Flavivirus*, dan famili *Flaviviridae* dan ditularkan melalui gigitan nyamuk genus *Aedes*: *Aedes Aegypti* atau *Aedes Albopictus*. Penyakit dengue bervariasi dari demam ringan hingga kondisi parah demam berdarah dengue dan sindrom syok. Kasus DBD di Indonesia pada tahun 2020 menurut Kementerian Kesehatan Republik Indonesia terdapat banyaknya 108.303 kasus dan memiliki jumlah kematian sebanyak 747 orang [1,2].

Salah satu upaya yang dapat dilakukan adalah dengan melakukan prediksi penyebaran penyakit DBD sehingga dapat meminimalisir terjadinya peningkatan kasus. Salah satu metode untuk melakukan prediksi demam berdarah adalah dengan menggunakan pendekatan *Machine Learning*. Di dalam *Machine Learning* terdapat salah satu teknik yaitu *Boosting*. *Boosting* adalah pendekatan pembelajaran mesin untuk membuat *rule* prediksi yang sangat akurat dengan menggabungkan banyak *rule* lemah dan tidak akurat [3]. *AdaBoost* merupakan salah satu algoritma *boosting* yang sering digunakan untuk membuat model klasifikasi. *AdaBoost* menggunakan *loss function* dari *exponential function* untuk meningkatkan akurasi prediksi [4, 5]. Pada tahun 2020, Fahmi dkk. melakukan penelitian evaluasi kinerja pengklasifikasi untuk memprediksi kasus infeksi virus dengue. Salah satu algoritma yang digunakan adalah *AdaBoost* [5]. Pada tahun yang sama, A. Puengpreeda dkk. melakukan penelitian penyebaran penyakit DBD yang dilakukan di negara Thailand menggunakan Algoritma *AdaBoost* dengan menggunakan data cuaca harian di Thailand tahun 2014-2018 dengan hasil performansi *Mean Squared Error* (MSE) terbaik sebesar 5.52, *Mean Absolute Error* (MAE) terbaik sebesar 1,61 dan untuk *R-Squared* (R^2) terbaik yaitu sebesar 0,47 [6].

Sedangkan untuk penelitian prediksi penyebaran penyakit DBD di Kabupaten Bandung berdasarkan data cuaca di Kabupaten Bandung dilakukan oleh Muzakki pada tahun 2018 menggunakan *K-Means Clustering* dan *Support Vector Machine* (SVM) [7] namun perbedaannya dengan penelitian ini adalah pada metode, sumber dataset dan rentang waktu yang digunakan.

Pada Tugas Akhir ini diimplementasikan algoritma *AdaBoost* untuk memprediksi penyebaran penyakit DBD di Kabupaten Bandung. Data yang digunakan adalah data jumlah kejadian DBD dan data cuaca dari tahun 2009-2021. Diharapkan *AdaBoost* dapat memberikan performansi yang baik dalam melakukan prediksi dan hasil dari penelitian ini dapat memberikan informasi bagi Dinas Kesehatan Kabupaten Bandung untuk melakukan pencegahan peningkatan kasus penyebaran penyakit DBD di Kabupaten Bandung.

Topik dan Batasannya

Rumusan masalah dari penelitian ini adalah bagaimana memprediksi penyebaran penyakit DBD di Kabupaten Bandung, bagaimana korelasi antara cuaca dengan *incident rate* (IR) dan bagaimana mengukur performansi algoritma *AdaBoost* dalam memprediksi penyebaran penyakit DBD. Adapun Batasan masalah dalam penelitian ini, yaitu dataset yang digunakan merupakan data bulanan kejadian DBD dan data cuaca dari tahun 2009-2021. Data cuaca meliputi *2 meters dew point temperature*, *2 meters temperature*, *mean sea level pressure*, *surface net thermal radiation*, *surface net thermal radiation (clear sky)*, *surface pressure*, dan *relative humidity* yang diambil dari ERA 5. Sedangkan data bulanan kejadian DBD didapatkan dari Dinas Kesehatan Kabupaten Bandung.

Tujuan

Tujuan dari penelitian ini adalah memprediksi penyebaran penyakit DBD di Kabupaten Bandung, menganalisa korelasi antara cuaca dengan angka kejadian atau *incident rate* (IR), dan mengukur performansi algoritma *AdaBoost* dalam memprediksi penyebaran penyakit DBD.

Organisasi Tulisan

Bagian selanjutnya yaitu bab 2, membahas mengenai studi yang berkaitan dengan penelitian yang telah dilakukan. Pada bab 3 membahas model dan sistem yang digunakan untuk penelitian. Pada bab 4 berisi mengenai hasil dan pembahasan setiap hasil yang didapat, lalu kesimpulan akan ditampilkan pada bab 5.

1. Studi Terkait

1.1 Demam Berdarah Dengue (DBD)

Dengue adalah penyakit menular yang sangat endemik di negara-negara tropis dan dengan cepat menjadi beban global. Penyakit ini disebabkan oleh salah satu dari 4 serotipe virus dengue dan ditularkan melalui manusia melalui nyamuk *Aedes betina*. Penyakit dengue bervariasi dari demam ringan hingga kondisi parah demam berdarah dengue dan sindrom syok. Penyakit ini banyak ditemukan di daerah tropis dan subtropis, menempatkan hampir sepertiga populasi manusia, di seluruh dunia, pada risiko infeksi [1]. Menurut Kementerian Kesehatan Republik Indonesia, jumlah penderita Demam Berdarah Dengue (DBD) di Indonesia pada tahun 2020 mencapai 108.303 kasus dengan jumlah kematian sebanyak 747 orang [2]. Faktor cuaca juga sangat mempengaruhi akan

peningkatan penyebaran penyakit menular ini, seperti suhu, curah hujan dan kelembapan. Faktor-faktor ini dapat menyebabkan lebih cepatnya perkembangbiakan nyamuk sehingga jumlah penderita penyakit menular ini pada musim hujan lebih tinggi daripada musim lainnya [7].

1.2 Adaptive Boosting (AdaBoost)

Adaptive Boosting (AdaBoost) merupakan pendekatan ke *machine learning* berbasis *ensemble classifier* untuk mengambil keputusan berdasarkan ide pembuatan *prediction rule* berakurasi tinggi dengan menggabungkan banyaknya *relative lemah* dan *rule* yang tidak akurat, menyesuaikan bobot secara adaptif untuk setiap siklus grup *classifier* yang lemah. AdaBoost mencoba menghasilkan *base classifier* baru dan dapat memberikan hasil yang lebih baik karena perbedaan-perbedaan di antara *classifier* yang lemah berdasarkan kinerjanya di setiap *classifier* [8, 9]. Gambar 1 menampilkan algoritma AdaBoost:

<p>Algoritma AdaBoost:</p> <p>Input: Sampel penelitian dengan label $\{(x_i, y_i), \dots, (x_N, y_N)\}$ suatu <i>component learn</i> algoritma, jumlah perputaran T.</p> <p>Output: $f(x) = \text{sign}(\sum_{t=1}^T \alpha_t h_t(x))$</p> <p>1: <i>Initialize:</i> Bobot suatu sampel pelatihan $w_i^1 = 1/N$, untuk semua $i = 1, \dots, N$</p> <p>2: Do for $t = 1, \dots, T$</p> <p>3: Melatih komponen klasifikasi dan h_t dengan menggunakan <i>component learn</i> algoritma.</p> <p>4: Hitung kesalahan pelatihan pada h_t: $\epsilon_t = \sum_{i=1}^N w_i^t, y_i \neq h_t(x_i)$</p> <p>5: Tetapkan bobot untuk <i>component classifier</i> $h_t = \alpha_t = \frac{1}{2} \ln(\frac{1-\epsilon_t}{\epsilon_t})$</p> <p>6: <i>Update</i> sampel pelatihan $w_i^{t+1} = \frac{w_i^t \exp(-\alpha_t y_i h_t(x_i))}{c_t}, i = 1, \dots, N$ c_t adalah suatu konstanta normalisasi.</p>
--

Gambar 1. Algoritma dari AdaBoost

1.3 Ukuran Performansi

Ukuran performansi yang digunakan pada penelitian ini berupa:

1.2.1 Root Mean Square Error (RMSE)

Root Mean Square Error (RMSE) merupakan metode pengukuran nilai dari suatu model dalam penelitian. Penggunaan nilai abosolut sangat tidak diinginkan dalam banyak perhitungan sistematis sehingga membuat RMSE memiliki keuntungan dikarenakan metode ini menghindari penggunaan nilai absolut [10]. RMSE memiliki nilai terbaik yaitu 0 dan tidak terbatas untuk nilai terburuknya. Rumus RMSE ditunjukkan oleh persamaan (1):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2} \tag{1}$$

Pada persamaan (1), n merupakan banyaknya sampel yang digunakan, X_i merupakan nilai dari model prediksi dan Y_i merupakan data dari target [11].

1.2.2 Correlation Coefficient (CC)

Correlation Coefficient (CC) merupakan pengukuran kuat atau lemahnya arah hubungan linier antara dua variabel. Ketika nilai salah satu variabel meningkat, maka variabel lainnya mengalami peningkatan yang sama. *Range* dari nilai CC terletak antara -1 hingga 1. Ketika nilai $CC = 0$, maka hubungan antara variabel x dan y dapat dikatakan tidak ada [12]. Rumus CC ditunjukkan oleh persamaan (2) [13]:

$$CC = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{2}$$

Pada persamaan tersebut, x_i dan y_i merupakan nilai dari variabel data, dengan \bar{x} dan \bar{y} merupakan nilai rata-rata data tersebut [14].

1.2.3 Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error (MAPE) merupakan metode pengukuran nilai kesalahan absolut untuk mengevaluasi keakuratan prediksi. MAPE menunjukkan seberapa banyak error dalam prediksi. MAPE mengukur akurasi nilai prediksi model dalam bentuk rata-rata persentase kesalahan absolut. Semakin rendah nilai MAPE maka semakin tinggi akurasi prediksi model yang dipakai. Rumus MAPE ditunjukkan oleh persamaan (3):

$$\text{MAPE} = \frac{\sum_{t=1}^n \left| \frac{a-b}{a} \right|}{n} \times 100\% \quad (3)$$

Pada persamaan tersebut, a merupakan nilai dari data aktual, b merupakan hasil data prediksi dan n merupakan banyaknya data [15, 16].

1.4 Incident Rate (IR)

Incident Rate (IR) adalah angka insiden terjadinya Demam Berdarah Dengue (DBD) per 100.000 jumlah penduduk [17]. Rumus IR ditunjukkan oleh persamaan (4).

$$\text{IR} = \frac{\text{jumlah kasus} \times 100.000}{\text{populasi}} \quad (4)$$