

1. Pendahuluan

Latar Belakang

Penyakit Demam Berdarah Dengue (DBD) merupakan penyakit menular yang memiliki angka kematian yang cenderung tinggi. Penyakit DBD ini telah dikategorikan sebagai penyakit yang mematikan sejak tahun 2013. Kejadian penyakit DBD yang telah dikonfirmasi pada tahun 2019 di Indonesia tercatat sebanyak 138.127 orang. Kejadian yang terkonfirmasi pada tahun 2019 menunjukkan bahwa adanya peningkatan dari sekitar 65.602 orang yang terkonfirmasi pada tahun 2018. Kejadian ini mengakibatkan Provinsi Jawa Barat mengalami peningkatan *incidence rate*. *Incidence rate* di Jawa Barat pada tahun 2018 sekitar 24.75% meningkat pada tahun 2019 menjadi 51.48% [1]. Peningkatan *incidence rate* yang sangat signifikan ini dipengaruhi dari kejadian DBD pada kota/kabupaten di Jawa Barat, salah satunya Kabupaten Bandung. Berdasarkan buku Profil Kesehatan Kabupaten Bandung Tahun 2019, telah terkonfirmasi kejadian DBD sekitar 3.274 orang dengan kematian sebanyak 24 jiwa, kejadian yang terkonfirmasi pada tahun 2018 sekitar 1.778 orang dengan kematian sebanyak 11 jiwa [2].

Pada tahun 2018, M Mufli Muzakki dan Fhira Nhita melakukan riset untuk memprediksi penyebaran DBD di Kabupaten Bandung menggunakan algoritma *Support Vector Machine*. Penelitian dilakukan dengan membagi dua skenario dalam pemilihan atribut cuaca. Skenario satu dilakukan dengan menggunakan seluruh atribut cuaca meliputi *solar radiation*, *humidity*, *wind velocity*, *temperature*, *evaporation*, dan *rainfall*. Sedangkan skenario dua hanya menggunakan tiga atribut cuaca yaitu *temperature*, *evaporation*, dan *rainfall*. Akurasi yang didapat untuk skenario 1 dan skenario 2 dengan normalisasi berturut sebesar 93% dan 90%. Sedangkan akurasi untuk skenario satu dan skenario dua tanpa normalisasi berturut sebesar 93% dan 91%. Diketahui bahwa dengan pemilihan atribut cuaca dapat mempengaruhi nilai akurasi yang didapat [3].

Pada tahun 2015, Kamran Shaukat dkk. Melakukan studi tentang Prediksi DBD di Pakistan menggunakan lima algoritma untuk mencari akurasi terbaik dari dua algoritma yaitu algoritma *Naïve bayes* dan J48, dengan hasil akurasi berurut yaitu 92% dan 88% [4]. Tahun 2019, Iswaya Maalik dkk. Melakukan riset dalam membandingkan akurasi rata-rata antara algoritma XGBoost dengan algoritma *Random Forest*, dalam melakukan klasifikasi kategori dari sambatan sekuens DNA, dengan hasil akurasi rata-rata yang diperoleh sebesar 95% untuk XGBoost dan 92% untuk *Random Forest* [5].

Pada tahun 2021, Labeba Tahsin dan Shaily Roy melakukan penelitian untuk membandingkan tingkat akurasi antara beberapa algoritma *Machine Learning* untuk memprediksi tingkat keparahan covid19. Algoritma yang dipakai berturut ialah *Decision Tree*, *Random Forest*, *Logistic Regression*, XGBoost, dan *Support Vector Machine* dengan akurasi dari tiap algoritma berturut 91.69%, 90.56%, 76.47%, 95.07%, 76,19% [6]. Dalam penelitian tersebut, XGBoost merupakan algoritma dengan tingkat akurasi tertinggi yaitu sebesar 95.07%.

Pada Tugas Akhir ini dilakukan penelitian mengenai prediksi penyebaran penyakit DBD yang ada di Kabupaten Bandung dengan menggunakan algoritma XGBoost. Data yang digunakan yaitu data jumlah kejadian DBD dan data cuaca dari tahun 2011-2021. Diharapkan hasil dari penelitian ini dapat memberikan informasi bagi Dinas Kesehatan Kabupaten Bandung untuk mengurangi kasus penyebaran penyakit DBD di Kabupaten Bandung.

Topik dan Batasannya

Topik pada penelitian ini adalah melakukan prediksi terhadap penyebaran penyakit DBD di Kabupaten Bandung menggunakan metode algoritma XGBoost. Dataset yang digunakan merupakan data kejadian DBD dan data cuaca dari tahun 2011-2021. 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*, *relative humidity* yang merupakan data cuaca di Kabupaten Bandung. Pada penelitian ini telah dilakukan percobaan sebanyak tiga skenario. Skenario I yaitu mencari panjang data training terbaik. Skenario II mencari fitur cuaca yang paling berpengaruh terhadap skor performansi dalam bentuk RMSE dan CC. Skenario III melakukan *hyperparameter tuning* terhadap data yang telah terseleksi pada skenario I dan II. Setelah mendapatkan semua hasil, dilakukan analisis performansi dari tiap skenario.

Tujuan

Tujuan dari penelitian ini melakukan analisis hubungan antara cuaca dengan *incidence rate*, melakukan penerapan algoritma XGBoost untuk memprediksi penyakit DBD, dan menganalisis performansi dari algoritma XGBoost dalam memprediksi penyebaran penyakit DBD di Kabupaten Bandung.

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.