

1. Pendahuluan

Latar Belakang

Virus Dengue adalah penyebab paling umum dari penyakit Antrhopod Borne Virus secara global. Menurut World Health Organization (WHO) diperkirakan terjadi 100 juta infeksi setiap tahunnya, sekitar 500.000 diantaranya adalah pengidap demam berdarah (DBD) dan sebagian besar adalah anak-anak. DBD dapat berkembang menjadi Dengue Shock Syndrome (DSS) dengan tingkat kematian 1-2,5%, jika tidak diberikan perawatan yang tepat dapat meningkat melebihi 20% [1].

Kabupaten Bandung yang terletak di iklim tropis yang mendukung penyebaran DBD ini karena DBD ditularkan melalui nyamuk yang berjenis *Aedes Aegypti* dan *Ae. Albopictus* yang terinfeksi oleh virus Dengue [2]. DBD sangat berbahaya karena dapat memicu kematian, ditambah lagi saat ini belum ada obat yang secara spesifik untuk menangani DBD. Oleh karena itu pencegahan adalah cara yang paling efektif untuk menekan penyebaran DBD. Banyak pihak yang mencoba mencegah penyebaran demam berdarah. Banyak peneliti juga yang melakukan prediksi terhadap penyebarannya untuk membantu pihak-pihak untuk mengetahui kemungkinan yang terjadi dimasa yang akan datang.

Autoregressive integrated Moving Average (ARIMA) adalah salah satu metode yang umum digunakan dalam melakukan prediksi, khususnya untuk data time series. Model ini pernah digunakan untuk memprediksi insiden demam berdarah di Rio de Janerio, Brazil[3].

Metode lain yang sudah umum digunakan untuk prediksi data time series adalah Recurrent Neural Networks (RNN). RNN pada tahun 2016 pernah digunakan untuk mempelajari dan memantau diagnosa menggunakan data pediatric intensive care unit untuk kepentingan medis[4]. Long-Short Term Memory (LSTM) adalah unit dari RNN yang dirancang untuk menghilangkan Long Term Dependencies karena RNN ini memiliki ketergantungan antara langkah yang terpisah jauh sehingga menyebabkan vanishing/exploding gradient [5].

Topik dan Batasannya

Dalam tugas akhir ini adalah mengetahui seberapa bagus optimasi dari prediksi ARIMA menggunakan RNN dan LSTM. Dataset yang digunakan adalah *incident rate* demam berdarah dan cuaca di kabupaten Bandung. Dataset cuaca digunakan untuk mengetahui hubungan antara cuaca tersebut dan *incident rate*. Cuaca terbagi menjadi menjadi beberapa variable, seperti penyinaran, kelembapan, angin, temperatur, uap, dan curah hujan.

Tujuan

Tujuan dari tugas akhir ini adalah mengetahui optimasi dari model ARIMA menggunakan RNN dan LSTM untuk prediksi *multi-variate* dataset. Model optimasi ini disebut dengan model *hybrid* atau model yang menggabungkan beberapa model agar memiliki model yang lebih optimal.

Model ARIMA digunakan untuk memprediksi dataset. Sebelum dilakukan prediksi perlu dilakukan menentukan model yang optimal dari ARIMA untuk dataset tersebut. Setelah mendapat model dari ARIMA maka dapat dilakukan prediksi dataset. Dari hasil prediksi dataset didapatkan residual atau error dari prediksi.

Residual atau error dari prediksi menggunakan model ARIMA akan digunakan oleh RNN dan LSTM. RNN dan LSTM melakukan *training* menggunakan data residual dari prediksi ARIMA. Setelah dilakukan *training*, RNN dan LSTM kemudian melakukan prediksi residual dari ARIMA. Hasil prediksi tersebut kemudian dijumlahkan dengan prediksi dari model ARIMA.