

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil penelitian yang dilakukan, kesimpulan yang dapat diambil dari penelitian ini yaitu:

- Dengan menggunakan model GRU dataset Imputasi menggunakan KNN dengan *scalling*, hasil yang didapatkan memberikan performa terbaik dalam hasil pelatihan model untuk prediksi *multivariate* data *time series* dengan jumlah data 14400 dan memiliki 6 atribut dengan akurasi 92% dan loss 0.001. Selain itu, rasio pembagian dataset *train* dan *test* juga mempengaruhi performa model yang diuji. Yang didapatkan bahwa model GRU dengan dengan rasio 7:3 dan dataset imputasi menggunakan KNN dengan *scalling* memiliki kombinasi nilai RMSE dan R2 terbaik dibandingkan model lainnya. Model GRU 7:3 paling baik dalam memprediksi parameter. Model GRU 7:3 paling baik dalam memprediksi parameter NO2 dengan RMSE 2.48 dan R2 0.39, sedangkan paling buruk dalam memprediksi PM2.5 dengan RMSE 22.61 dan R2 -1.89. namun untuk parameter lainnya selain PM2.5 model dapat memprediksi dengan cukup baik. Sedangkan model LSTM dengan dataset imputasi menggunakan KNN yang memiliki performa terbaik untuk *loss* diantara lainnya dengan *loss* 0.008. kemudian *univariate* LSTM 7:3 yang memiliki performa terbaik dalam memprediksi PM2.5 dengan RMSE 18.29 dan R2 47% dibandingkan *multivariate*. Namun, *multivariate* masih tetap mampu memprediksi untuk parameter lainnya karena selisih perbandingan antara *multivariate* GRU 7:3 dan *univariate* LSTM 7:3 hanya kecil.
- Model GRU dengan imputasi KNN dengan *scalling*, memiliki akurasi yang cukup baik juga dalam melakukan *forecasting* untuk 31 hari dari tanggal terakhir. Hasil *forecasting* palijugang baik adalah parameter NO2 dengan RMSE 2.51 dan R2 0.62. lalu jika dibandingkan antara *multivariate* dan

univariate, maka *univariate LSTM 7:3* juga paling baik dalam melakukan *forecasting* dengan nilai RMSE 14.74 dan R2 0.65.

5.2 Saran

Penulis menyarankan beberapa hal agar mengupayakan peningkatan performa model dalam prediksi kualitas udara di Jakarta *multivariate data time series* yaitu:

- Menggunakan dataset dengan jumlah data yang lebih banyak dan minim *missing values* sehingga hasil prediksi lebih akurat.
- Menambahkan features data seperti data cuaca karena ada banyak hal lainnya yang juga mempengaruhi kualitas udara.
- Membuat model dengan konfigurasi yang lebih sesuai dengan karakteristik data agar mampu memprediksi dengan baik untuk keseluruhan parameter.
- Menambahkan fitur tambahan agar bisa menangkap karakteristik yang ada pada tiap data.
- Mengatasi ketidakseimbangan dataset pada parameter PM2.5 menggunakan teknik untuk mengatasinya sehingga hasil akurasi loss dapat meningkat.

