

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian yang telah dilakukan pada *dataset* ISPU DKI Jakarta ini adalah menggunakan sebuah *hybrid model* dengan tahapan *clustering* kemudian hasil *clustering* tersebut diprediksi. *Model hybrid* tersebut dibuat dengan bantuan *framework data mining*, yaitu *CRISP-DM*. Penggunaan algoritma *clustering* yang dilakukan secara *2-step-clustering* menggunakan *K-Means* dan *K-Medoids* dapat mengelompokkan data berdasarkan stasiun pemantau kualitas udara (SPKU) pada *dataset* ISPU. Kemudian, hasil *clustering* dilakukan pemodelan prediksi menggunakan *Long Short-Term Memory (LSTM)*.

Penggunaan 2 *optimizer* yang berbeda pada *model* prediksi menunjukkan bahwa beberapa variabel atau kolom data cocok digunakan untuk *optimizer* tertentu. Hal ini juga dapat dilihat pada hasil visualisasi dan juga perbandingan nilai *RMSE* sebelumnya. Pada penelitian ini, hasil dari *model LSTM* yang paling baik ada pada variasi *dataset* ISPU 2021 – 2023 dimana terdapat perbedaan signifikan antara *cluster* 1 dan juga *cluster* 2 *dataset* tersebut. *Cluster* 2 pada *dataset* ISPU 2021 – 2023 tersebut juga memberikan nilai *RMSE* terkecil, berkisar antara 2.91984 hingga 9.53943. Hal tersebut juga menunjukkan bahwa jumlah *dataset* sangat berpengaruh untuk menciptakan hasil prediksi yang baik. Selain itu, penerapan *model* prediksi menggunakan *LSTM* juga sangat cocok untuk memprediksi data pendukung, yaitu cuaca yang berupa suhu, kelembapan udara, dan kecepatan angin. Hasil visualisasi akhir dengan menggunakan perbandingan antara data prediksi dengan data aktual juga menunjukkan bahwa hasil prediksi masih dapat mengikuti pola data aktual, serta hasil *forecast* selama 14 hari ke depan menghasilkan tren yang bermacam-macam juga, seperti ada yang mengalami kenaikan dan penurunan, serta ada juga tren yang konstan.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan mengenai *dataset* ISPU DKI Jakarta menggunakan *hybrid model* berupa *clustering* dan prediksi, terdapat

beberapa saran untuk pengembangan bagi penelitian berikutnya dikarenakan waktu yang terbatas ketika melakukan penelitian ini, yaitu sebagai berikut:

1. Mencari atau menggunakan *dataset* ISPU pada provinsi yang berbeda dan juga memastikan agar masing-masing *dataset* lengkap sehingga hasil prediksi dapat menjadi lebih baik.
2. Menggunakan *dataset* pendukung lainnya dari sumber yang berbeda dan mencari keterkaitan antara variabel agar dapat dibuat menjadi *model multivariate* yang lebih bervariasi dan lebih baik.
3. Melakukan konfigurasi *hyperparameter* pada bagian model prediksi menggunakan *Long Short-Term Memory (LSTM)* yang masih dapat dilakukan optimasi lebih lanjut sehingga menghasilkan prediksi yang lebih baik.
4. Menggunakan algoritma *clustering* dan juga algoritma prediksi yang berbeda untuk memperkaya literatur pengetahuan dalam ranah *data mining*.

