

BAB V

KESIMPULAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. Cara pengimplementasian algoritma *multiple linear regression* untuk memprediksi temperatur udara berdasarkan kadar zat polutan dimulai dengan dilakukan pengumpulan data, kemudian dilanjutkan dengan *data cleaning* dan praproses data meliputi *merge* data *drop* kolom, pengimputasian *missing value*, dan *outliers handling*. Kemudian data yang sudah bersih dibagi menjadi 80% data latih dan 20% data uji dengan teknik *random subsampling* menggunakan *train_test_split* dari *scikit-learn*. Setelah itu, data distandardisasi untuk memastikan bobot pada setiap fitur sama. Untuk memilih model terbaik, dilakukan beberapa uji coba model dengan menggunakan *feature selection* SelectKBest ANOVA dengan parameter *F-value* untuk memilih sejumlah *k* fitur yang paling relevan. Kemudian dilakukan pemodelan dengan memanggil fungsi `LinearRegression()` dari *scikit-learn* yang implementasinya menggunakan *ordinary least squares* (OLS). Metode OLS dilakukan dengan cara menghitung koefisien garis regresi terbaik yang meminimalisasi jumlah kuadrat selisih antara nilai observasi dan nilai prediksi menggunakan persamaan matriks yang jika dirumuskan yaitu $(X^T X)^{-1} X^T Y$. Hasil dari penghitungan tersebut didapatkan nilai *intercept* dan koefisien model prediksi *multiple linear regression*.
2. Dengan menggunakan metrik MAPE, didapatkan performa model dengan nilai error sebesar 1,99806% atau dapat dikatakan model memiliki tingkat akurasi yang baik dalam memprediksi temperatur udara rata-rata harian, karena nilai kesalahannya relatif rendah.
3. Nilai hasil evaluasi model pada percobaan keenam dengan menggunakan nilai SelectKBest ANOVA *F-value* *k*=8 merupakan capaian hasil evaluasi

terbaik dari kedelapan skema percobaan model dengan perolehan nilai R^2 sebesar 0.72749, MAE sebesar 0.5593, dan MSE sebesar 0.50078.

4. Beberapa kadar zat polutan seperti nitrogen dioksida (NO_2), ozon (O_3), dan partikulat 10 (PM_{10}) masih memiliki pengaruh pada model dan dapat digunakan sebagai fitur dalam model prediksi. Hal ini dibuktikan dengan terpilihnya variabel-variabel tersebut sebagai fitur prediksi melalui *feature selection* SelectKBest ANOVA F-value.
5. Kadar zat polutan memiliki pengaruh langsung yang tidak terlalu signifikan pada temperatur rata-rata harian. Terbukti dengan semua nilai hubungan pada *correlation matrix* kadar zat polutan terhadap temperatur rata-rata yang rendah atau dapat dikatakan tidak berkorelasi kuat. Namun, ditemukan terdapat korelasi kuat antar variabel kadar zat polutan.
6. Data iklim harian seperti kadar zat polutan yang memberikan pengaruh paling signifikan pada kinerja dan nilai hasil evaluasi model yaitu kelembapan rata-rata (RH_{avg}), curah hujan (RR), lamanya penyinaran matahari (ss), arah angin terbanyak (ddd_x), kecepatan angin rata-rata (ff_{avg}), nitrogen dioksida (NO_2), ozon (O_3), dan partikulat 10 (PM_{10}).
7. Model *multiple linear regression* pada penelitian ini tidak memiliki kompleksitas yang tinggi. Waktu *training* model ini didapatkan selama 0,0358893 detik dan waktu *testing* model ini terhitung hanya 0,0004606 detik. Penggunaan memori pada model ini juga tidak besar, yaitu 262.097 MB.

5.2 Saran

Adapun saran-saran yang dapat diterapkan pada penelitian berikutnya yaitu sebagai berikut:

1. Meskipun kadar zat polutan tidak secara langsung memiliki pengaruh yang signifikan pada temperatur udara adanya korelasi kuat antar variabel kadar zat polutan dapat diteliti lebih lanjut. Hal tersebut kemungkinan menunjukkan terdapat interaksi antara jenis zat polutan tertentu yang tidak terlihat secara langsung dalam korelasi dengan temperatur. Studi lebih lanjut

dapat memperluas pemahaman tentang bagaimana polusi dapat mempengaruhi variabilitas suhu udara.

2. Memperbanyak jumlah data pada model prediksi seperti dengan menggunakan data dengan jangka waktu 10 tahun atau lebih agar hasil pemodelan lebih akurat.
3. Perlu dilakukan eksperimen pemodelan prediksi dengan algoritma lain sebagai perbandingan.
4. Model prediksi ini dapat dikolaborasikan dengan teknologi IoT yang mencatat kadar zat polutan secara *real-time*.