

PREDIKSI RATA-RATA ZAT BERBAHAYA DI DKI JAKARTA BERDASARKAN INDEKS STANDAR PENCEMAR UDARA MENGGUNAKAN METODE LONG SHORT-TERM MEMORY

¹Anisa Oktaviani, ²Hustinawati

^{1,2}Manajemen Sistem Informasi, Program Pasca Sarjana, Universitas Gunadarma
Jl. Margonda Raya No. 100, Depok 16424, Jawa Barat

¹92318030@student.gunadarma.ac.id, ²hustina@staff.gunadarma.ac.id

Abstrak

Indonesia menempati peringkat ke-6 dari 98 negara paling berpolusi di dunia pada tahun 2019. Di tahun tersebut, rata-rata AQI (Air Quality Index) sebesar 141 dan rata-rata konsentrasi PM_{2.5} sebesar 51.71 µg/m³ yang lima kali lipat diatas rekomendasi World Health Organization (WHO). Salah satu kota penyumbang polusi udara yaitu Jakarta. Berdasarkan data ISPU (Indeks Standar Pencemar Udara) yang diambil dari SPKU (Stasiun Pemantau Kualitas Udara) Dinas Lingkungan Hidup DKI Jakarta melampirkan pada tahun 2019, Jakarta memiliki kualitas udara sangat tidak sehat. Oleh karena itu perlu adanya model Artificial Intelligence dalam memperdiksi rata-rata tingkat zat berbahaya pada udara di DKI Jakarta. Salah satu algoritma yang dapat diterapkan dalam membuat model prediksi dengan menggunakan data timeseries adalah Long Short-Term Memory (LSTM). Tujuan dari penelitian ini membangun model prediksi rata-rata ISPU di DKI Jakarta menggunakan metode LSTM yang berguna bagi para pemangku kepentingan dibidang lingkungan hidup khususnya mengenai polusi udara. Penelitian mengenai prediksi rata-rata ISPU di DKI Jakarta menggunakan metode LSTM, menghasilkan nilai evaluasi MAPE 12.28%. Berdasarkan hasil evaluasi MAPE yang diperoleh, model LSTM yang digunakan untuk prediksi rata-rata ISPU di DKI Jakarta masuk kedalam kategori akurat.

Kata Kunci: DKI Jakarta, Indeks Standar Pencemar Udara, Long Short-Term Memory, Prediksi

Abstract

Indonesia was ranked 6th out of 98 most polluted countries in the world in 2019. In that year, the average AQI (Air Quality Index) was 141 and the average PM_{2.5} concentration was 51.71 µg / m³ which was five times that above the recommendations of the World Health Organization (WHO). One city that contributes to air pollution is Jakarta. ISPU data (Air Pollutant Standard Index) taken from the SPKU (Air Quality Monitoring Station) of the DKI Jakarta Environmental Service said that in 2019, Jakarta had very unhealthy quality. Therefore it is necessary to have an Artificial Intelligence model in predicting the average level of hazardous substances in the air in DKI Jakarta. One of the algorithms that can be applied in making prediction models using time series data is Long Short-Term Memory (LSTM). The purpose of this study is to build a prediction model for the average ISPU in DKI Jakarta using the LSTM method which is useful for stakeholders in the environmental sector related to air pollution. Research on the prediction of the average ISPU in DKI Jakarta using the LSTM method, resulted in a MAPE evaluation value of 12.28%. Based on the results of the MAPE evaluation obtained, the LSTM model used to predict the average ISPU in DKI Jakarta is in the accurate category.

Keywords: DKI Jakarta, Indeks Standar Pencemar Udara, Long Short-Term Memory, Prediction

PENDAHULUAN

Dewasa ini, bidang ilmu *Artificial Intelligence* (AI) semakin banyak diminati dan terus berkembang. Kecerdasan Buatan atau Artificial Intelligence (AI) adalah teknik yang digunakan untuk meniru kecerdasan yang dimiliki oleh makhluk hidup maupun benda mati untuk menyelesaikan sebuah persoalan[1]. Salah satu fungsi *Artificial Intelligence* yang banyak dimanfaatkan adalah melakukan prediksi. Prediksi adalah proses perhitungan sistematis dalam memperkirakan suatu objek dimasa yang akan datang berdasarkan data masa lalu dan sekarang. Algoritma yang dapat diterapkan dalam membuat model prediksi dengan menggunakan data *timeseries* adalah *Long Short-Term Memory* (LSTM).

LSTM adalah jenis arsitektur pengembangan dari *Recurrent Neural Network* (RNN), sehingga terdapat kelebihan yang dimiliki yaitu dapat menyimpan informasi dalam jangka waktu yang panjang dengan menambahkan *memory cell*. Selain itu, LSTM dapat menjadi solusi saat memproses data sekuensial yang panjang sehingga mengatasi terjadinya *vanishing gradient*. Permasalahan *vanishing gradient* merupakan kekurangan dari RNN dalam menangkap *long term dependencies* yang mengurangi akurasi dari suatu prediksi.

Prediksi yang berhubungan dengan aktifitas kehidupan bermasyarakat sehari-hari adalah prediksi zat berbahaya penyebab

pencemaran udara. Menurut data yang dapat dilihat dari website IQAir, pada tahun 2019 Indonesia menempati peringkat ke-6 dari 98 negara paling berpolusi di dunia [2] dengan rata-rata AQI 141 dan rata-rata konsentrasi PM2.5 51.71 $\mu\text{g}/\text{m}^3$ yang lima kali lipat diatas rekomendasi World Health Organization (WHO). Salah satu kota yang penyumbang polusi udara di Indonesia yaitu Jakarta[2].

Pada tahun 2019, Jakarta menempati peringkat ke-3 dari 10 kota paling berpolusi di Indonesia[3]. Berdasarkan data ISPU (Indeks Standar Pencemar Udara) yang diambil dari SPKU (Stasiun Pemantau Kualitas Udara) Dinas Lingkungan Hidup Jakarta menunjukkan pada tahun 2019, Jakarta memiliki 2 hari (1%) kualitas udara baik, 172 hari (48%) kualitas udara sedang, 183 hari (50%) kualitas udara tidak sehat dan 8 hari (2%) kualitas udara sangat tidak sehat [4].

Perkembangan *Artificial Intelligence* semakin masif diaplikasikan dalam penyelesaian masalah di kehidupan sehari-hari. Salah satunya dengan membangun model prediksi rata-rata ISPU di DKI Jakarta. Prediksi tersebut salah satunya dapat berguna bagi para pemangku kepentingan dibidang lingkungan hidup khususnya mengenai polusi udara. Model prediksi yang dibangun dapat dimplementasikan menggunakan metode *Long Short-Term Memory* (LSTM) karena data *time series* kualitas udara menghasilkan tingkat akurasi yang tinggi dan hasil evaluasi pengujian yang rendah [5].

Penelitian terkait mengenai prediksi terhadap kualitas udara telah banyak dilakukan oleh peneliti terdahulu. Penelitian mengimplementasikan RNN-LSTM dalam melakukan prediksi konsentrasi polutan udara di kota New Delhi dan Agra. Hasil penelitian menunjukkan bahwa metode RNN-LSTM menghasilkan akurasi prediksi yang baik untuk 1 jam dibanding 6 – 12 jam berikutnya [6]. Penelitian menerapkan metode LSTM untuk sistem monitoring real-time memprediksi Influenza Like Illness (ILI) dan kadar zat PM_{2.5} di Taiwan. Hasil evaluasi dari model prediksi ILI dan AQI adalah nilai MAPE kurang dari 10% yang berarti nilai akurasi tinggi [5]. Penelitian mengimplementasikan LSTM untuk memprediksi dan meramalkan debit air untuk lima hari kedepan di Sungai Cimandiri dalam membantu masyarakat mengambil keputusan dalam pengendalian operasi irigasi. Hasil dari penelitian menunjukkan model memperoleh hasil RSME kurang dari 10%, sehingga model yang dibuat baik untuk prakiraan jangka pendek [7]. Penelitian mengimplementasikan metode CNN-LSTM dalam memprediksi parameter polutan PM_{2.5} di Kota Beijing. Hasil penelitian menggunakan 10 sample menunjukkan bahwa model multivariate CNN-LSTM merupakan model yang memiliki akurasi tertinggi dengan RSME 17.9306 dan MAE 13.9697 [8]. Penelitian mengimplementasikan model RNN-LSTM dalam peramalan polusi udara di Beijing. Hasil penelitian menunjukkan bahwa

algoritma RNN-LSTM menghasilkan akurasi yang setara dengan data asli, tetapi hanya sampai batas waktu tertentu yaitu 5 sampai 6 jam kedepan. Setelah batas waktunya lewat hasilnya akan menurun drastis [9]. Penelitian membandingkan beberapa model yaitu SVR-KERNEL-RBF, SVR-KERNEL-POLY, SVR-KERNEL-LINERA, dan RNN-LSTM untuk memprediksi kualitas udara di kota Visakhapatnam, India. Hasil penelitian menunjukkan metode RNN-LSTM akurasi tinggi dan kinerja lebih baik dalam matrik evaluasi [10]. Penelitian mengimplementasikan kombinasi model RNN, GRU dan LSTM untuk memprediksi polutan PM_{2.5} di Distrik Keçiören, Turki dengan kisaran waktu 1, 2, dan 3 jam kedepan. Hasil penelitian terbaik diperoleh untuk prediksi 1 jam dari model terpilih dengan R² 0.832, RMSLE 0.404, RMSE 6.272 dan MAE 4.21 [11]. Penelitian membandingkan antara LSTM–RNN dengan ARIMA dalam mempelajari jeda waktu antara paparan konsentrasi tinggi PM_{2.5} dan respon kondisi kesehatan tubuh manusia. Selain itu menguji mode lag terdistribusi non-linier (DLNM) untuk mensimulasikan hubungan eksposur–respon dan distribusi waktu yang digunakan untuk mendeteksi efek lag dari PM_{2.5} pada kunjungan darurat rumah sakit pada bagian penyakit pernafasan. Hasil menunjukkan bahwa model berbasis LSTM dapat digunakan sebagai metode baru untuk analisis lag yang bermanfaat untuk mendukung kebijakan darurat rumah sakit [12]. Penelitian

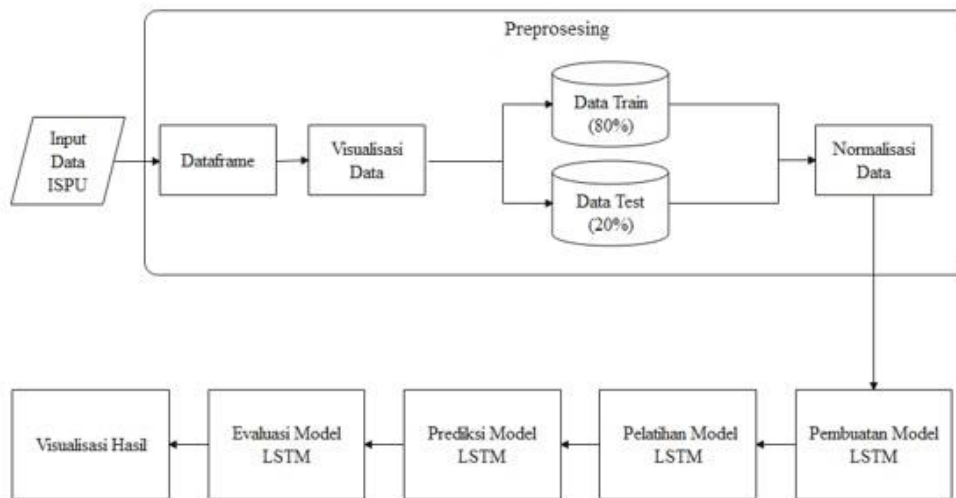
menimplementasikan metode LSTM dan DAE untuk memprediksi konsentrasi PM10 dan PM2.5. Hasil yang didapatkan adalah metode LSTM lebih akurat dibandingkan metode DAE untuk prediksi konsentrasi PM10 dan PM2.5 [13]. Penelitian mengimplementasikan metode LSTM untuk memprediksi kualitas udara di Kota Shanghai, Cina. Evaluasi pada model LSTM menghasilkan RSME 10.95 dan deviasi kuadrat 7.56 [14]. Berdasarkan telaah teoritis, hasil-hasil penelitian terdahulu, maka dapat dikembangkan hipotesis yaitu pengolahan data time series khususnya mengenai kualitas udara menghasilkan tingkat akurasi yang tinggi jika diimplementasi menggunakan metode long short-term memory.

Hipotesis lain dari penelitian ini adalah H\hasil evaluasi yang didapat dari model prediksi ISPU menggunakan metode LSTM menghasilkan nilai MAPE dengan presentasi rendah.

METODE PENELITIAN

Metode penelitian dalam melakukan prediksi rata-rata ISPU (Indeks Standar Pencemaran Udara) di DKI Jakarta menggunakan metode *Long Short-Term*

Memory terdiri dari beberapa tahapan-tahapan. Bagan umum metode penelitian ini dapat dilihat pada Gambar 1. Berdasarkan Gambar 1 tahapan awal penelitian adalah memasukkan dataset Indeks Standar Pencemar Udara (ISPU). Dataset ISPU yang didapat dikelola oleh Dinas Lingkungan Hidup Provinsi DKI Jakarta dan dapat di unduh melalui portal resmi data terbuka Indonesia yaitu website Satu Data Indonesia. Dataset ISPU kemudian diproses ke Google Colab untuk memulai *pre-processing*. Tahapan kedua adalah melakukan *pre-processing* pada dataset. *Preprocessing* adalah proses manipulasi dataset ISPU sebelum dimasukkan ke dalam model. Tujuan dari *pre-processing* adalah menyesuaikan data agar lebih kompatibel dengan *library* yang akan digunakan. Pada tahapan *pre-processing* prediksi rata-rata ISPU terdiri dari proses dataframe, visualisasi data, pembagian data, dan normalisasi. Tahapan ketiga adalah proses pembuatan model prediksi rata-rata ISPU di DKI Jakarta. Pembuatan model dengan mengimplementasikan metode LSTM, kemudian menambahkan layer. Penambahan layer disesuaikan dengan jumlah data, agar menghasilkan model dengan nilai akurasi tinggi.



Gambar 1. Metode Penelitian

Tahapan keempat adalah melakukan pelatihan model LSTM. Pelatihan model bertujuan untuk melihat nilai loss dari data dan melatih model yang dibangun agar mengenal data-data yang telah dimasukkan. Pada tahapan ini juga ditambahkan beberapa *layer* tambahan seperti *layer epoch* dan *batch size*. *Epoch* menandakan bahwa seluruh dataset telah melalui satu kali putaran proses training pada Neural Network.

Batch size adalah sample data yang didistribusikan ke Neural Network. *Layer-layer* tersebut disesuaikan nilainya hingga menghasilkan akurasi tinggi dan nilai loss yang rendah.

Tahapan kelima adalah melakukan prediksi dari model LSTM, dengan membandingkan data asli dengan data hasil prediksi model LSTM. Prediksi ISPU menggunakan data testing yang dihasilkan pada tahap preprocessing. Hasil dari prediksi akan menampilkan data aktual dengan data

hasil prediksi, sehingga dapat terlihat perbedaannya.

Tahapan keenam adalah evaluasi model dengan tujuan untuk mengetahui jumlah error yang dihasilkan dan melihat kelayakan dari model yang dibuat.

Evaluasi model LSTM menggunakan perhitungan nilai Mean Percentage Error (MAPE). Hasil dari perhitungan MAPE yang rendah menandakan bahwa model prediksi menghasilkan nilai yang mendekati nilai aktual atau data real. Tahapan ketujuh yaitu menampilkan visualisasi dari hasil prediksi model tersebut. Visualisasi data adalah teknik mengkomunikasikan data atau informasi dengan menggambarkannya ke dalam objek visual. Hasil dari prediksi rata-rata ISPU DKI Jakarta menggunakan pemodelan LSTM akan di visualisasi dalam bentuk plot diagram. Dari hasil visualisasi tersebut dapat dilihat gambaran visual dari hasil prediksi model dengan nilai aktualnya.

HASIL DAN PEMBAHASAN

Memasukkan Data Indeks Standar Pencemar Udara

Data input ISPU (Indeks Stanadar Pencemar Udara) yang digunakan berasal dari data Dinas Lingkungan Hidup DKI Jakarta yang dapat di download dari website Satu Data Indonesia. Data-data tersebut digabungkan sehingga menghasilkan *dataset* dengan jumlah observasi sebanyak 6384 data. *Dataset* tersebut merupakan data ISPU harian dari periode Januari 2017 – Juli 2020. Pada Tabel 1 terdapat rincian dari dataset ISPU.

Dalam *dataset* tersebut terdapat beberapa variabel yaitu tanggal, wilayah, PM10, SO2, CO, O3, NO2, max, critical, dan kategori. Pada Tabel 2 terdapat penjelasan dari atribut *dataset* ISPU.

Preprocessing Data ISPU

Tahap preprocessing merupakan tahap awal dalam proses penelitian ini. Tahap preprocessing dilakukan pada *dataset* dengan tujuan agar datasiap untuk dipakai dan sesuai dengan model yang dibangun. Berikut tahapan-tahapan dari *pre-processing* data ISPU:

Tabel 1. Rincian Dataset ISPU

Tahun	Jumlah Data
2017	1825
2018	1825
2019	1825
2020	909
Total	6384

Tabel 2. Atribut Dataset ISPU

Atribut	Penjelasan
Tanggal	Tanggal pengukuran kualitas udara
Wilayah	Lokasi SPKU di DKI Jakarta
PM ₁₀	Parameter Partikulat yang diukur
SO ₂	Parameter Sulfida (dalam bentuk SO ₂) yang diukur
CO	Parameter Karbon Monoksida yang diukur
O ₃	Parameter Ozon yang diukur
NO ₂	Parameter Nitrit yang diukur
Max	Nilai ukur parameter paling tinggi dari seluruh parameter yang diukur dalam waktu bersamaan
Critical	Paramater yang memiliki nilai ukur paling tinggi
Kategori	Kategori hasil perhitungan ISPU

Pembuatan Dataframe

Tahapan pertama pada *preprocessing* yaitu membuat *dataframe* baru yang berisikan atribut tanggal dan rata-rata dari ISPU DKI Jakarta. Atribut tanggal didapat dari *dataset* ISPU, sedangkan atribut rata-rata merupakan variabel baru yang didapat dari rata-rata zat pencemar udara yaitu PM₁₀, SO₂, CO, O₃, dan NO₂ pada setiap SPKU, kemudian hasil rata-rata setiap SPKU di rata-rata kembali sehingga menghasilkan nilai rata-rata dari lima SPKU di DKI Jakarta. Pada Tabel 3 terdapat rincian *df1* yang merupakan *dataframe* yang baru saja dibuat.

Visualisasi Data

Dataframe *df1* selanjutnya divisualisasi dengan tujuan untuk memberi gambaran

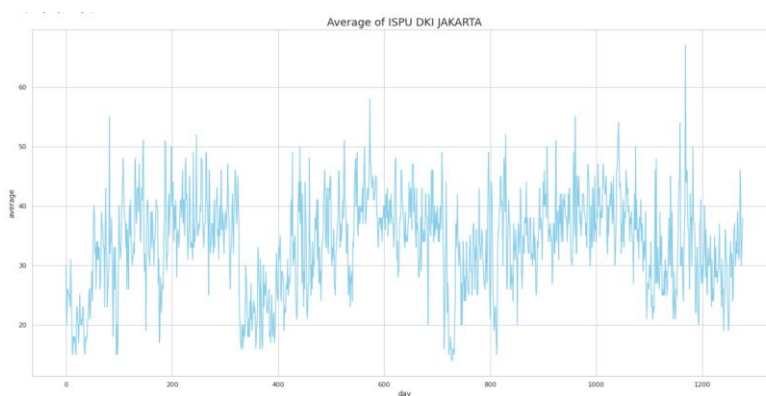
berbentuk grafik. Visualisasi yang ditampilkan yaitu grafik rata-rata ISPU DKI Jakarta dari periode Januari 2017 – Juli 2020. Keluaran yang dihasilkan berupa grafik rata-rata ISPU yang terlampir pada Gambar 2.

Pembagian Dataset

Dataframe *df1* yang telah di visualilasi selanjutnya dibagi menjadi dua bagian yaitu *training* dan *testing*. *Data training* adalah data yang digunakan untuk melatih algoritma, agar mencapai model yang diinginkan. *Data testing* adalah data yang digunakan untuk mengetahui performa dan kebenaran dari algoritma yang telah dilatih dalam model. Pembagian *dataset* diatur dengan persentase 80% untuk *training* dan 20% untuk *testing* yang tertera pada Tabel 4.

Tabel 3. Dataframe

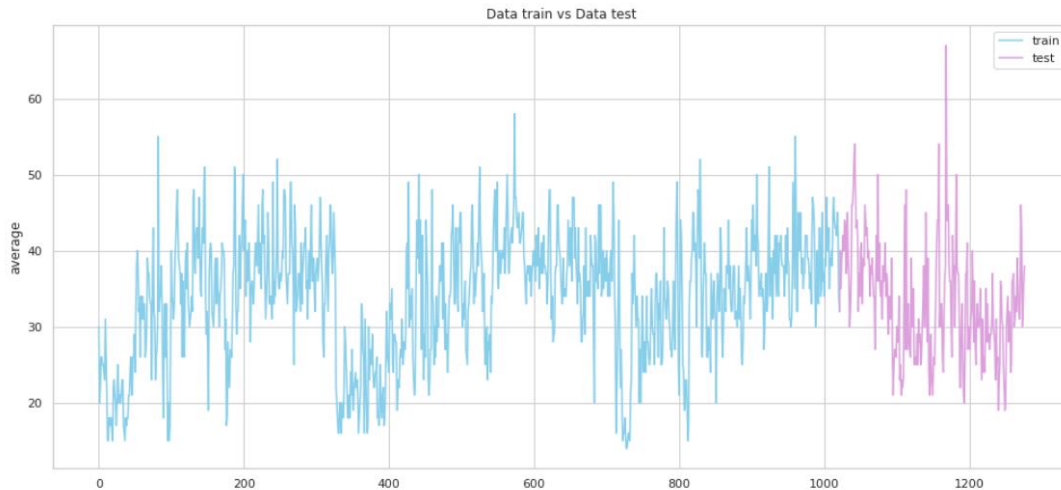
	Tanggal	Average
0	2017-1-1	30
1	2017-1-2	20
2	2017-1-3	22
....
1274	2020-6-28	32
1276	2020-6-29	36
1277	2020-6-30	38



Gambar 2. Visualisasi Nilai Rata-rata ISPU DKI Jakarta

Tabel 4. Pembagian Dataset

Dataframe	Data Training (80%)	Data Testing (20%)
Rata-rata ISPU DKI Jakarta	1022	256



Gambar 3. Visualisasi Pembagian Dataframe

Dataframe df1 yang sebelumnya telah diatur persentase pembagiannya, menghasilkan data yang terdiri dari 1022 untuk *data training* dan 256 untuk *data testing*. Pembagian data yang telah berhasil diimplementasikan, kemudian akan divisualisasi dalam bentuk grafik. Visualisasi dalam bentuk grafik bertujuan untuk mengkomunikasikan data secara jelas dan efisien. Visualisasi *dataframe* df1 terdapat pada Gambar 3.

Normalisasi Data

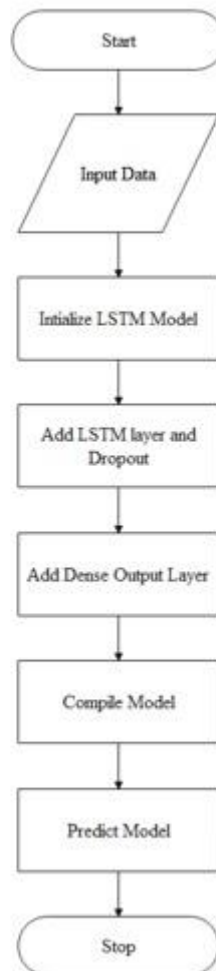
Pembagian dataframe menghasilkan data yang terdiri dari 1022 untuk data training dan 256 untuk data testing. Tahapan selanjutnya menyamaratakan perbedaan

rentang nilai pada numerical data yang dapat menimbulkan masalah dalam model machine learning.

Oleh karena itu, *feature scaling* bertujuan untuk menyamaratakan numerical data pada dataset sehingga memiliki rentang nilai (scale) yang sama. Pada Gambar 4 terdapat hasil dari normalisasi data. Setelah melakukan normalisasi data, tahap selanjutnya yaitu membuat struktur data untuk memvalidasi. Struktur data menghasilkan variabel baru yaitu *X_train* dan *y_train* dengan data yang digunakan untuk validasi sebanyak 60 data. Setelah pembuatan struktur data, selanjutnya melakukan perubahan urutan data atau reshape data dari 0 – 1.


```
array([[0.31472081],
       [0.15736041],
       [0.15736041],
       ...,
       [0.36040609],
       [0.34010152],
       [0.37563452]])
```

Gambar 4. Hasil Normalisasi Data



Gambar 5. Alur Model LSTM

Pembuatan Model LSTM

Pembuatan model merupakan representasi dari bentuk matematis algoritma LSTM. Proses pembuatan model diawali dengan memasukkan data ISPU yang didapat dari website Satu Data Indonesia, data

tersebut telah dilakukan proses pre-processing yang terdiri dari pembuatan dataframe, pembagian dataframe menjadi data testing dan data training, normalisasi data, dan membangun struktur data. Setelah melakukan pre-processing data akan masuk ke lapisan

LSTM, lapisan pada LSTM dapat terdiri dari beberapa layer dengan tujuan untuk menghasilkan model yang akurat. Tahap terakhir adalah lapisan Dense Output yang bertujuan untuk menghasilkan prediksi. Pada Gambar 5 terdapat gambaran dari alur pembuatan model LSTM. Pada tahap awal dalam membuat model LSTM adalah pembuatan class sequential sebagai container agar lapisan LSTM dapat ditambah secara berurut dan saling terhubung. Model yang dibangun terdiri dari 5 lapisan yaitu 4 lapisan LSTM dan 1 lapisan Dense Output. Lapisan LSTM menggunakan 64 unit neuron dan dropout sebesar 0.2. Jumlah lapisan LSTM tersebut dipilih karena menghasilkan nilai evaluasi paling optimal. Berikut pada tabel 4.1 terdapat rincian dari jumlah layer LSTM dan hasil evaluasi yang diperoleh. Pada Tabel 5 terdapat rincian angka yang digunakan.

Dari tabel 5 dapat diketahui bahwa layer LSTM dengan 64 neuron dan 0.1 dropout menghasilkan nilai MAPE paling optimal. Oleh karena itu, layer tersebut diterapkan pada pembuatan model LSTM prediksi rata-rata indeks pencemar udara.

Lapisan LSTM disertai dengan argumen `return_sequences` bernilai `true` yang dapat di artikan sebagai argumen yang mengeluarkan urutan agar lapisan LSTM berikutnya memiliki input 3D. Lapisan Dense Output memiliki 1 unit neuron. Complie model menggunakan fungsi optimasi RMSprop dengan learning rate 0.001 dan loss yang digunakan adalah mean squared error.

Pelatihan Model LSTM

Model LSTM yang telah dibangun, selanjutnya akan masuk ke tahap pelatihan. Pelatihan model bertujuan untuk mengetahui hasil dari nilai loss dari data. Model pelatihan menggunakan *layer epoch* dan *batch size*. Jumlah *epoch* yang digunakan sebanyak 8 dan jumlah *batch_size* sebanyak 4.

Implementasi pada tahap pelatihan model LSTM dengan *layer epoch* yang digunakan sebanyak 8 dan jumlah *batch_size* sebanyak 4 menghasilkan nilai loss sebesar 0.0109. Pada Gambar 6 terdapat tampilan hasil dari pelatihan model.

Tabel 5. Percobaan Jumlah Layer LSTM

Neuron	Dropout	MAPE
64	0.1	12.49%
164	0.2	12.28%
128	0.1	12.35%
128	0.2	12.76%

```

Epoch 6/8
241/241 [=====] - 31s 127ms/step - loss: 0.0109
Epoch 7/8
241/241 [=====] - 31s 127ms/step - loss: 0.0109
Epoch 8/8
241/241 [=====] - 31s 127ms/step - loss: 0.0109
<tensorflow.python.keras.callbacks.History at 0x7f343bf366d8>

```

Gambar 6. Hasil Pelatihan Model

	Predict	Real
0	[32.28756]	[34]
1	[34.432434]	[31]
2	[32.498215]	[33]
3	[33.75685]	[33]
4	[33.728245]	[26]

Gambar 7. Hasil Prediksi Model

Tabel 6. Tingkat Akurasi MAPE

Nilai MAPE	Akurasi
< 10%	Sangat Akurat
11% - 20%	Akurat
21% - 50%	Cukup Akurat
>50%	Tidak Akurat

(Sumber : Lewis (1982, p. 40 [16]))

Prediksi Model LSTM

Tahap yang dilakukan setelah melakukan proses training pada model yaitu prediksi model LSTM. Prediksi dilakukan pada *data testing*. Hasil dari prediksi tersebut dapat dilihat pada Gambar 7.

Evaluasi Model LSTM

Hasil prediksi yang dihasilkan, selanjutnya akan di evaluasi. Evaluasi model menggunakan LSTM dengan metode *Mean Absolute Percentage Error* (MAPE) dengan

mencari nilai kesalahan absolut disetiap periode dibagi dengan nilai observasi yang aktual pada periode tersebut dan dibuat rata-rata dari *absolute percentage error*. Pada Tabel 6 terdapat penjelasan mengenai nilai MAPE untuk data industri dan bisnis serta interpretasinya[15]. Perhitungan MAPE dapat ditulis dengan rumus seperti persamaan 1.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (1)$$

Keterangan :

Y_t = Nilai aktual pada periode t

\hat{Y}_t = Nilai *forecast* pada periode t

Evaluasi diterapkan pada perbandingan antara data prediksi dengan *data test*. Pada Gambar 8 terdapat nilai evaluasi kesalahan dari prediksi *data training*. Nilai MAPE yang dihasilkan pada prediksi *data training* yaitu 12.28%.

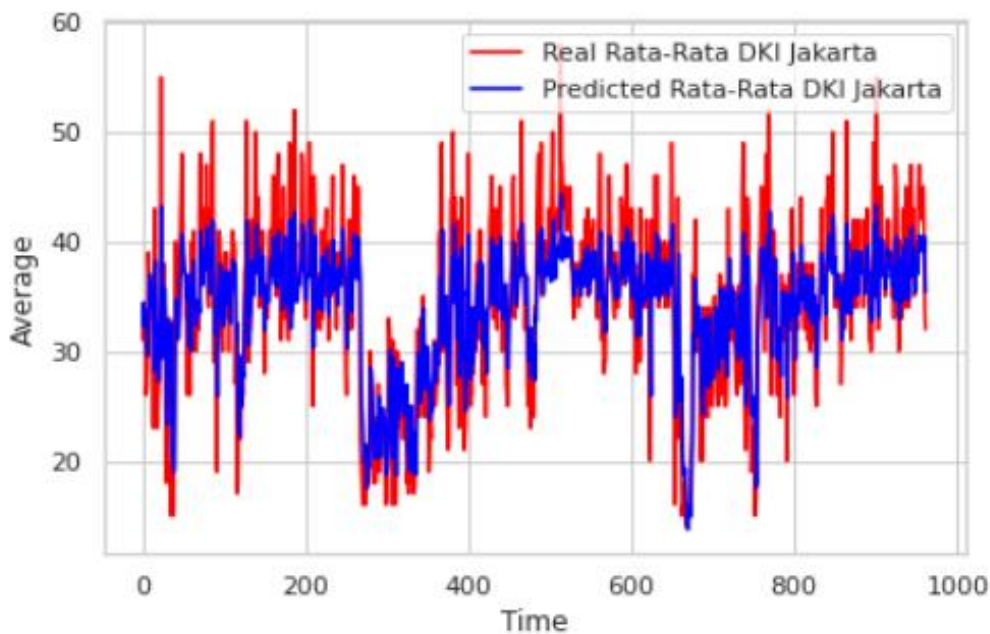
Hasil prediksi tersebut masuk ke dalam kategori MAPE tingkatan akurat. Visualisasi Model LSTM Evaluasi model LSTM yang dibangun mendapat kategori MAPE tingkatan akurat. Tahapan selanjutnya setelah

melakukan evaluasi model yaitu melakukan visualisasi.

Hasil dari prediksi rata-rata ISPU DKI Jakarta menggunakan pemodelan LSTM akan di visualisasi dalam bentuk *plot diagram*. Diagram tersebut membandingkan data rata-rata asli yang diwakilkan garis berwarna merah dengan data rata-rata hasil prediksi dari model yang diwakilkan dengan garis berwarna biru. Visualisasi tersebut dapat dilihat pada Gambar 9.

0.06183527458466975
MAPE : 12.28%

Gambar 8. Nilai MAPE pada Prediksi Data Training



Gambar 9. Hasil Visualiasi Prediksi Data

KESIMPULAN DAN SARAN

Prediksi rata-rata ISPU (Indeks Standar Pencemar Udara) dilakukan pada lima SPKU (Stasiun Pemantau Kualitas Udara) di DKI Jakarta dari periode Januari 2017 – Juni 2020. Data yang digunakan dalam membuat model prediksi adalah dari Dinas Lingkungan Hidup Provinsi DKI Jakarta yang didapat dari website data.jakarta.go.id. Prediksi rata-rata ISPU menggunakan metode *long short-term memory*. Pemodelan LSTM menggunakan *layer neuron* sebanyak 64 dan *dropout* 0.2. Hasil evaluasi model prediksi rata-rata ISPU di DKI Jakarta menggunakan metode LSTM dilakukan pada data test, menghasilkan nilai MAPE 12.28% dan masuk kedalam kategori model yang akurat. Berdasarkan hasil penelitian prediksi rata-rata ISPU di DKI Jakarta yang dilakukan maka untuk evaluasi lebih lanjut sebaiknya dilakukan penambahan data ISPU untuk dapat menghasilkan performa model yang semakin baik. Selain itu, hasil penelitian dapat dikembangkan lebih lanjut oleh pihak Dinas Lingkungan Hidup DKI Jakarta dalam meningkatkan informasi yang terdapat dalam website resmi instansi tersebut, sehingga dapat mempermudah pengguna yang ingin melihat informasi mengenai ISPU beserta rata-ratanya di setiap SPKU wilayah DKI Jakarta.

DAFTAR PUSTAKA

- [1] A. Ahmad, “Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning,” in *Academia*, 2017. [Online] Available: <https://www.academia.edu> [Accessed: Nov 26, 2020].
- [2] IQAir, “World Air Quality,” 2019 *World Air Qual. Rep.*, pp. 1–22, 2019, [Online]. Available: <https://www.iqair.com/world-most-polluted-cities/world-air-quality-report-2019-en.pdf>. [Accessed: Nov 26, 2020].
- [3] AQIAir, “AQIAir Website.” [Online] Available: <https://www.iqair.com>. [Accessed: Nov 26, 2020].
- [4] L. Arumingtyas, “Jakarta Air Quality,” *mongabay.com*, para. 8, Apr. 25, 2020. [Online] <https://www.mongabay.co.id/2020/04/25/polusi-udara-pembunuh-senyap-di-jabodetabek/>. [Accessed: Nov 26, 2020].
- [5] C. T. Yang, L. Y. Lin, Y. T. Tsan, P. Y. Liu, and W. C. Chan, “The Implementation of a Real-time Monitoring and Prediction System of PM2.5 and Influenza-Like Illness Using Deep Learning,” *J. Internet Technol.*, vol. 20, no. 7, pp. 2237–2245, 2019, doi: 10.3966/160792642019122007020. [Online Serial]. Available: <https://jit.ndhu.edu.tw/article/view/2210>. [Accessed: Dec. 1, 2020]
- [6] V. Chaudhary, A. Deshbhratar, V.

- Kumar, and D. Paul, "Time Series Based LSTM Model to Predict Air Pollutant 's Concentration for Prominent Cities in India," *Udm'18*, 2018. [Online Serial]. Available: http://philippe-fourm-nier-viger.com/utility_mining_workshop_2018/PAPER1.pdf. [Accessed: Dec. 1, 2020].
- [7] Y. Sudriani, I. Ridwansyah, and H. A Rustini, "Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 299, no. 1, 2019, doi: 10.1088/1755-1315/299/1/012037. [Online Serial]. Available: <https://iopscience.iop.org/article/10.1088/1755-1315/299/1/012037>. [Accessed Dec. 5, 2020].
- [8] T. Li, M. Hua, and X. U. Wu, "A Hybrid CNN-LSTM Model for Forecasting," *IEEE Access*, vol. 8, Feb, 2020. [Online serial]. Available: <https://ieeexplore.ieee.org/abstract/document/8979420>. [Accessed Dec. 2, 2020].
- [9] V. Reddy, P. Yedavalli, S. Mohanty, and U. Nakhat, "Deep Air: Forecasting Air Pollution in Beijing, China," 2017. [Online Serial] Available: <https://www.semanticscholar.org/paper/Deep-Air-%3A-Forecasting-Air-Pollution-in-Beijing-%2C-Reddy-Mohanty/271279939e545be172f7b754b67d6811c11c8fae>. [Accessed Dec. 7, 2020].
- [10] K. S. Rao, G. L. Devi, and N. Ramesh, "Air Quality Prediction in Visakhapatnam with LSTM based Recurrent Neural Networks," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 2, pp. 18–24, 2019, doi: 10.5815/ijisa.2019.02.03. [Online Serial] Available: <https://www.semanticscholar.org/paper/Air-Quality-Prediction-in-Visakhapatnam-with-LSTM-Rao-Devi/f34d77e3375bd512d14abe3cbacc29e31a2f83aa>. [Accessed Dec. 7, 2020].
- [11] Y. A. Ayturan *et al.*, "Short-term prediction of pm2.5 pollution with deep learning methods," *Glob. Nest J.*, vol. 22, no. 1, pp. 126–131, 2020, doi: 10.30955/gnj.003208. [Online Serial] Available: https://journal.gnest.org/publication/gnest_03208 [Accessed Dec. 10, 2020].
- [12] J. Lu, P. Bu, X. Xia, L. Yao, Z. Zhang, and Y. Tan, "A New Deep Learning Algorithm for Detecting the Lag Effect of Fine Particles on Hospital Emergency Visits for Respiratory Diseases," *IEEE Access*, vol. 8, pp. 145593–145600, 2020, doi: 10.1109/ACCESS.2020.3013543. [Online Serial] Available: <https://ieeexplore.ieee.org/document/91>

- 53761 [Accessed Dec. 10, 2020].
- [13] T. Xayasouk, H. M. Lee, and G. Lee, "Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models," *Sustain.*, vol. 12, no. 6, 2020, doi: 10.3390/su12062570. [Online Serial] Available: <https://www.mdpi.com/2071-1050/12/6/2570> [Accessed Dec. 10, 2020].
- [14] Y. Zhu and X. Zhou, "Prediction of Air Quality Index Based on Wavelet Transform Combination Model," *Proc. - 2019 11th Int. Conf. Intell. Human-Machine Syst. Cybern. IHMSC 2019*, vol. 1, no. Itaic, pp. 157–160, 2019, doi: 10.1109/IHMSC.2019.00044. [Online Serial] Available: <https://ieeexplore.ieee.org/document/8941558> [Accessed Dec. 13, 2020].
- [15] J. J. Montaña Moreno, A. Palmer Pol, A. Sesé Abad, and B. Cajal Blasco, "El índice R-MAPE como medida resistente del ajuste en la previsioón," *Psicothema*, vol. 25, no. 4, pp. 500–506, 2013, doi: 10.7334/psicothema2013.23. [Online Serial] Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=4655890> [Accessed Dec. 13, 2020].
- [16] C. Lewis, *Industrial and Business Forecasting Methods*. London: Butterworth-Heinemann, 1982.