

PREDIKSI POLUSI UDARA PERKOTAAN DI SURABAYA MENGUNAKAN *RECURRENT NEURAL NETWORK – LONG SHORT TERM MEMORY*

Muh. Anas Faishol¹⁾, Endroyono²⁾, dan Astria Nur Irfansyah³⁾

^{1, 2, 3)}Departemen Teknik Elektro, Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia 60111

e-mail: anas.18071@mhs.its.ac.id¹⁾, endroeleven@gmail.com²⁾, irfansyah@ee.its.ac.id³⁾

ABSTRAK

Udara merupakan salah satu kebutuhan primer makhluk hidup. Jika udara tercemar maka kehidupan manusia dan makhluk hidup lain akan terganggu. Karena itu perlu ada penanganan khusus agar udara tetap terjaga kualitasnya. Salah satu cara untuk mempermudah pencegahan terjadinya polusi udara adalah membuat prediksi polusi udara dengan memanfaatkan data masa lalu. Melalui Dinas Lingkungan Hidup, Pemerintah Kota Surabaya telah melakukan pemantauan kualitas udara di kota Surabaya setiap 30 menit untuk berbagai parameter kualitas udara di antaranya CO, NO, NO₂, NO_x, PM₁₀, SO₂, dan data meteorologi seperti arah angin, arah hembusan angin, kecepatan angin, kecepatan hembusan angin, radiasi global, kelembaban, dan temperatur udara. Data tersebut sangat berguna dalam membangun sebuah model prediksi untuk prakiraan polusi udara di masa yang akan datang. Dengan besarnya jumlah dan varian data yang dihasilkan dari pemantauan kualitas udara di kota Surabaya, dibutuhkan sebuah algoritma yang mumpuni untuk mengolahnya. Salah satu algoritma yang bisa digunakan adalah Recurrent Neural Network-Long Short Term Memory (RNN-LSTM). RNN-LSTM dibangun untuk pengolahan data sekuensial seperti data time series. Ada beberapa analisis yang dibuat dalam penelitian ini, yaitu analisis tren, analisis korelasi nilai polutan terhadap data meteorologi, dan prediksi polutan Karbon Monoksida menggunakan RNN-LSTM di kota Surabaya dikorelasikan dengan data meteorologi. Hasil penelitian ini menunjukkan bahwa model prediksi terbaik menggunakan RNN-LSTM dengan perhitungan RMSE mendapatkan error sebesar 1,880 dan skenario jumlah hidden layer 2 serta epoch 50. Hasil prediksi yang dibangun dapat digunakan sebagai acuan dalam menentukan kebijakan pemerintah kota untuk menangani polusi udara ke depan.

Kata kunci: Long Short Term Memory, polusi udara, Recurrent Neural Network, Surabaya.

PREDICT URBAN AIR POLLUTION IN SURABAYA USING RECURRENT NEURAL NETWORK – LONG SHORT TERM MEMORY

Muh. Anas Faishol¹⁾, Endroyono²⁾, and Astria Nur Irfansyah³⁾

^{1, 2, 3)}Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia 60111

e-mail: anas.18071@mhs.its.ac.id¹⁾, endroeleven@gmail.com²⁾, irfansyah@ee.its.ac.id³⁾

ABSTRACT

Air is one of the primary needs of living things. If the condition of air is polluted, then the lives of humans and other living things will be disrupted. So it is needed to perform special handling to maintain air quality. One way to facilitate the prevention of air pollution is to make air pollution forecasting by utilizing past data. Through the Environmental Office, the Surabaya City Government has monitored air quality in Surabaya every 30 minutes for various air quality parameters including CO, NO, NO₂, NO_x, PM₁₀, SO₂ and meteorological data such as wind direction, wind direction, wind speed, wind speed, global radiation, humidity, and air temperature. These data are very useful to build a prediction model for the forecast of air pollution in the future. With the large amount and variance of data generated from monitoring air quality in Surabaya city, a qualified algorithm is needed to process it. One algorithm that can be used is Recurrent Neural Network - Long Short Term Memory (RNN-LSTM). RNN-LSTM is built for sequential data processing such as time-series data. In this study, several analyses are performed. There are trend analysis, correlation analysis of pollutant values to meteorological data, and predictions of carbon monoxide pollutants using the Recurrent Neural Network - LSTM in the city of Surabaya correlated with meteorological data. The results of this study indicate that the best prediction model using RNN-LSTM with RMSE calculation gets an error of 1,880 with the number of hidden layer 2 and epoch 50 scenarios. The predicted results built can be used as a reference in determining the policy of the city government to deal with air pollution going forward.

Keywords: Air pollution, Long Short Term Memory, Recurrent Neural Network, Surabaya.

I. PENDAHULUAN

Manusia membutuhkan udara untuk hidup. Udara yang baik bagi makhluk hidup terutama untuk manusia adalah yang banyak mengandung oksigen sehingga terasa segar dan sejuk dalam tubuh. Kondisi udara di perkotaan di masa modern ini sudah berubah dari batas normal [1]. Hal ini disebabkan oleh polusi udara. Sisa hasil pembakaran mesin kendaraan bermotor atau emisi gas buang, kegiatan industri, sisa pembakaran pembangkit listrik, adanya bencana alam seperti meletusnya gunung berapi dan kebakaran hutan, pemanfaatan amonia, dan kegiatan memasak rumah tangga, serta penggunaan peralatan rumah tangga yang mengandung CFC mengakibatkan udara tercemar[2]. Udara yang tercemar memiliki dampak negatif bagi kesehatan manusia, di antaranya penyakit jantung, stroke, kanker paru-paru, diabetes, penyakit menular seperti pneumonia, dan kematian dini akibat penyakit pernapasan[2].

Menurut Laporan *State of Global Air* tahun 2018 yang diterbitkan oleh *Health Effects Institute* di Amerika dan *Institute for Health Metrics and Evaluation's Global Burden of Disease Project*, polusi udara adalah penyebab paling besar ke-6 yang memperpendek usia manusia atau kematian dini. Jumlah korban bisa melebihi kematian yang disebabkan faktor risiko lainnya seperti kecanduan alkohol, aktivitas fisik, konsumsi Natrium tinggi, atau kolesterol [2]. Namun tingkat kematian dini tersebut tidak sama di semua negara. Indonesia merupakan salah satu dari negara Asia yang terpengaruh oleh polusi udara. Kelima negara dengan tingkat kematian paling tinggi terdapat di Asia yaitu China, India, Pakistan, Indonesia dan Bangladesh [2]. Organisasi Kesehatan Dunia (WHO) memperkirakan bahwa pada tahun 2016 polusi udara luar, di perkotaan dan pedesaan, menyebabkan 4,2 juta kematian di seluruh dunia, dan 3,8 juta kematian lain akibat penyakit tidak menular (termasuk penyakit jantung, stroke, dan kanker) yang disebabkan oleh polusi udara rumah tangga. Dinyatakan bahwa 91% populasi dunia tidak menghirup udara bersih, dan lebih dari setengah populasi perkotaan terpapar pada tingkat polusi udara luar setidaknya 2,5 kali di atas standar keamanan yang ditetapkan oleh WHO [3].

Sektor transportasi sebagai tulang punggung aktivitas masyarakat berkontribusi besar pada pencemaran udara, hal ini dikarenakan saat ini mayoritas kendaraan bermotor masih menggunakan bahan bakar fosil. Menurut data Badan Pusat Statistik, jumlah kendaraan bermotor di Indonesia yang mencakup mobil penumpang, mobil bis, mobil barang, dan sepeda motor, dalam kurun waktu 10 tahun terakhir (2007-2017) mengalami pertumbuhan yang sangat tinggi yakni sebesar 153%. Sepeda motor mendominasi dengan 82% di tahun 2017 [4].

Seiring meningkatnya jumlah kendaraan bermotor yang beredar di masyarakat, menyebabkan meningkatnya polutan yang berupa emisi gas buang. Emisi gas buang merupakan gas yang dibuang oleh kendaraan bermotor dari sisa hasil proses pembakaran bahan bakar di dalam mesin. Emisi gas buang terdiri dari gas Karbon Monoksida (CO) yang beracun, Karbon Dioksida (CO₂), air (H₂O), senyawa Nitrogen Oksida (NO_x), Sulfur Oksida (SO₂), senyawa *Hydro Carbon* (HC), dan partikular debu termasuk Timbel (PB), sebagai akibat proses pembakaran mesin yang tidak sempurna serta partikel lepas [5]. Polusi udara karena emisi gas buang ini menjadi salah satu masalah yang muncul seiring berkembangnya sebuah kota menuju kota maju. Sebuah kota dianggap maju jika mampu memanfaatkan sumber daya yang ada secara cerdas.

Smart city menjadi label yang melekat di sebuah kota untuk menjawab tantangan dari berbagai macam masalah yang muncul di masyarakat untuk dipecahkan. Istilah *smart city* telah menjadi tren tersendiri dalam pelayanan di pemerintah daerah, baik di tingkat kota maupun kabupaten. *Smart city* berarti bahwa sebuah kota mampu memanfaatkan modal yang dimiliki, seperti sumber daya manusia, kekuatan hubungan sosial, dan infrastruktur telekomunikasi modern untuk mewujudkan pertumbuhan ekonomi berkelanjutan dan kualitas kehidupan yang tinggi, dengan manajemen sumber daya yang bijaksana melalui pemerintahan berbasis partisipasi masyarakat [6]. Kota pintar tercipta dengan tujuan untuk meningkatkan kenyamanan penghuninya melalui fasilitas mobilitas [7], efisiensi dan hemat energi [8], perbaikan kualitas air dan udara [9], identifikasi masalah dan pemberian solusi cepat, tanggap dalam pemulihan setelah bencana [10], dan integrasi data untuk pengambilan keputusan melalui pemanfaatan *cloud computing* dalam mengolah *big data* [11].

Smart environment merupakan aspek *smart city* yang bertujuan untuk melindungi dan memelihara lingkungan kota, baik keamanan maupun alam yang fokus pada pembahasan kemajuan teknologi serta pemanfaatannya. Lingkungan pintar berarti lingkungan yang mampu menghadirkan kenyamanan bagi masyarakat. Salah satu cara mengembangkan lingkungan pintar adalah penggunaan teknologi canggih untuk memproses data sehingga melahirkan informasi dan menganalisisnya [12]. Kota Surabaya merupakan salah satu kota di Indonesia yang telah menerapkan *smart city*, terutama di aspek *smart environment*. Melalui Dinas Lingkungan Hidup, Pemerintah Kota Surabaya telah melakukan pemantauan kualitas udara ambien pada beberapa parameter udara yaitu CO, PM₁₀, O₃, NO_x, NO₂, SO₂ dan beberapa parameter data meteorologi. Data hasil pemantauan tersebut disimpan dalam bentuk *file* yang volume-nya tiap hari bertambah besar. Dengan volume data dan varian yang besar ini memungkinkan

untuk diolah sehingga memberikan pengetahuan baru untuk menangani permasalahan lingkungan. Misalnya dengan mengimplementasikan teknologi *artificial intelligence*.

Saat ini, *Artificial Intelligence* (AI) makin masif diaplikasikan dalam kehidupan masyarakat untuk membantu penyelesaian masalah. Salah satu konsepsi AI yang memanfaatkan data besar untuk menyelesaikan masalah adalah *forecasting*. Peramalan (*forecasting*) adalah suatu teknik analisis perhitungan untuk memperkirakan kejadian di masa yang akan datang dengan menggunakan pengalaman di masa lampau [13]. Peramalan dapat dilakukan dengan pendekatan kualitatif maupun kuantitatif [14]. Tujuan peramalan adalah memprediksi prospek ekonomi dan kegiatan usaha serta pengaruh lingkungan terhadap prospek tersebut [15]. *Forecasting* merupakan bagian terpenting bagi setiap perusahaan, organisasi ataupun pemerintahan dalam setiap pengambilan keputusan manajemen. Peramalan bisa menjadi dasar bagi para perencana untuk membuat perencanaan jangka pendek, menengah maupun jangka panjang bagi organisasi komersial maupun tidak. Salah satu algoritma yang dipakai dalam peramalan yang memanfaatkan *Artificial Intelligence* yang berbasis data runtun waktu adalah *Recurrent Neural Network- Long Short Term Memory* (RNN-LSTM).

Jaringan saraf berulang atau *Recurrent Neural Network* (RNN) adalah jenis arsitektur jaringan saraf tiruan dengan masukan yang diproses secara berulang-ulang. Ciri khas RNN adalah data yang dijadikan masukan adalah data sekuensial. Karena data diproses melalui banyak lapis (*layer*) sehingga RNN masuk dalam kategori *deep learning*. Seiring dengan berkembangnya teknologi *hardware* untuk membantu proses *deep learning*, RNN telah mengalami kemajuan yang pesat dan telah diterapkan ke dalam berbagai bidang. Beberapa penelitian yang menggunakan RNN antara lain dalam bidang pemrosesan bahasa alami (NLP) [16][17][18], pengenalan suara [19], sistesis musik [20], manajemen keuangan [21], analisis deret DNA [22], analisis video [23], dan sebagainya. Seiring berkembangnya implementasi RNN, dikembangkan pula metode *Long Short Term Memory* (LSTM). Berbeda dengan konsep RNN, LSTM memiliki *memory cell* dan *input gates* untuk menyimpan informasi memori yang telah lama tersimpan.

Dalam penelitian ini, dibuat analisis tren, analisis korelasi antara partikular CO dengan data meteorologi, dan analisis prediksi menggunakan metode peramalan *Recurrent Neural Network-Long Short Term Memory* (RNN-LSTM) terhadap partikular Karbon Monoksida (CO) sebagai salah satu partikular utama penyebab polusi udara di bidang transportasi. Dari penelitian ini diharapkan para pemangku kepentingan khususnya di bidang pengendalian polusi lingkungan, dapat menggunakan metode ini sebagai acuan pengambilan keputusan dalam upaya mengendalikan polusi udara untuk tahun-tahun yang akan datang.

II. PENELITIAN TERKAIT DAN LITERATUR

A. Penelitian Terkait

Ada beberapa penelitian menggunakan *Recurrent Neural Network-Long Short Term Memory* (RNN-LSTM) yang menjadi rujukan dalam penelitian ini, diantaranya adalah Analisis dan Implementasi *Long Short Term Memory Neural Network* untuk Prediksi Harga Bitcoin [24], *Predicting Trends in Air Pollution in Delhi using Data Mining* [25], *Time Series Based LSTM Model to Predict Air Pollutant's Concentration for Prominent Cities in India* [26], dan *Deep Air: Forecasting Air Pollution in Beijing, China* [27].

Pada artikel [24], disampaikan bahwa penelitian ini membangun model LSTM untuk memprediksi harga Bitcoin menggunakan data jual beli indeks harga Bitcoin terhadap US Dollar yang diambil dari *blockchain.info* pada tanggal 22 Mei 2016 – 19 Mei 2018 dengan jumlah dataset sebesar 728. Pada penelitian tersebut didapatkan hasil yang terbaik yaitu dengan komposisi data latih 70% dan data uji 30%, jumlah neuron hidden 25, dan *max epoch* adalah 100 mendapatkan akurasi rata-rata pada data latih 95,36% dan data uji 93,5%.

Pada penelitian [25], membahas tren variasi polutan udara seperti *Sulphur Dioxide* (SO₂), *Nitrogen Dioxide* (NO₂), *Particulate Matter* (PM), *Carbon Monoxide* (CO), dan *Ozone* (O₃) menggunakan teknik *data mining* regresi linier dan *neural network multilayer perceptron*. Dataset berasal dari *Central Pollution Time Series Analysis Control Board* (CPCB) antara tahun 2011 sampai dengan 2015. Di penelitian ini polutan udara dikorelasikan dengan beberapa parameter kendaraan seperti kendaraan roda dua, bus, mobil dan kendaraan lainnya. Hasil penelitian yang didapat adalah ada peningkatan jumlah PM10 sebesar 45,9% di tahun-tahun mendatang. Namun jumlah CO dan NO₂ menunjukkan sedikit peningkatan karena meningkatnya jumlah kendaraan roda dua di jalan. Sedangkan polutan lain seperti SO₂ menunjukkan penurunan.

Pada penelitian [26], dibahas tentang prediksi parameter udara seperti PM2.5, PM10, CO, NO₂, SO₂, O₃ dengan penunjang fitur meteorologi (suhu, tekanan, kelembaban, kecepatan angin, arah angin, indeks UV, tutupan awan, hujan), informasi lalu lintas, dan informasi festival. Data diambil di kota Delhi dan Agra, India. Dataset polutan udara yang digunakan berasal dari *website Central pollution Control Board* (CPCB) dengan periode data tiap satu jam. Data meteorologi diambil dari *website accuweather*, sedangkan data informasi lalu lintas berasal dari API peta

“Here”. Dataset dikumpulkan oleh ITO Delhi dan Sanjay Pallace Stasiun Agra selama 3 bulan dari 12 Februari sampai dengan 6 Mei 2018. Hasil penelitian ini menyebutkan bahwa nilai RMSE untuk 2 lokasi sensor bernilai <5 untuk prediksi 12 jam berikutnya, nilai <8 untuk prediksi 6 jam berikutnya, dan nilai < 5 untuk jam berikutnya.

Pada penelitian [27], diinvestigasi penggunaan RNN-LSTM sebagai kerangka kerja prediksi berdasarkan data deret waktu polusi udara dan informasi meteorologi di Beijing. Model prakiraan yang dibangun menggunakan regresi nonlinier menggunakan *Support Vector Regression* (SVR) dan LSTM. Hasil penelitian ini adalah model LSTM adalah prediktor yang lebih baik dalam jeda waktu yang lebih lama daripada metode *nonlinear* lainnya seperti SVR.

Selain itu ada beberapa penelitian yang menunjukkan bahwa performansi LSTM lebih baik daripada algoritma lain dalam membuat peramalan, diantaranya *A Comparison of ARIMA and LSTM in Forecasting Time Series* [28], *Comparative Study of Short-Term Wind Speed Forecasting Techniques Using Artificial Neural Networks* [29], dan *Comparison of BPA-MLP and LSTM-RNN for Stocks Prediction* [30].

Pada penelitian [28], dataset merupakan data series keuangan bulanan berupa pasar saham dan data ekonomi dari situs web keuangan yahoo dengan membagi 70% sebagai data latih dan 30% sebagai data uji. Studi empiris yang dilakukan dan dilaporkan dalam artikel ini menunjukkan bahwa algoritma LSTM mengungguli algoritma berbasis tradisional seperti model ARIMA. Data yang terkait dengan seri waktu keuangan atau pasar saham menunjukkan bahwa rata-rata Rooted Mean Squared Error (RMSE) menggunakan model Rolling ARIMA dan Rolling LSTM, masing-masing adalah 511,481 dan 64,213, serta menghasilkan rata-rata 87,445 pengurangan tingkat kesalahan yang dicapai oleh LSTM. Di sisi lain, data terkait ekonomi menunjukkan pengurangan 84,394 dalam RMSE dimana nilai rata-rata RMSE untuk Rolling ARIMA dan Rolling LSTM masing-masing dihitung sebagai 5,999 dan 0,936. Nilai RMSE jelas menunjukkan bahwa model berbasis LSTM mengungguli model berbasis ARIMA dengan margin tinggi (antara 84% –87% penurunan tingkat kesalahan).

Penelitian [29] membahas tentang perbandingan performansi antara LSTM dengan RNN. Studi ini menggunakan dataset kecepatan angin direkam dari ladang angin di dekat desa Managuli di bagian selatan Karnataka dari tahun 2010 sampai dengan 2014 dengan periode perekaman 1 jam. Hasil evaluasi dua model jaringan yang dibangun menunjukkan bahwa LSTM berkinerja lebih baik dan lebih cepat jika urutan yang diberikan memiliki dependensi jangka panjang sedangkan RNN berkinerja lebih baik dengan dependensi jangka pendek. Waktu yang dibutuhkan oleh jaringan LSTM untuk prediksi deret waktu jauh lebih pendek dibandingkan dengan model RNN pada data yang sama. Dengan beberapa masukan data, LSTM mudah beradaptasi terhadap perubahan dalam pola data.

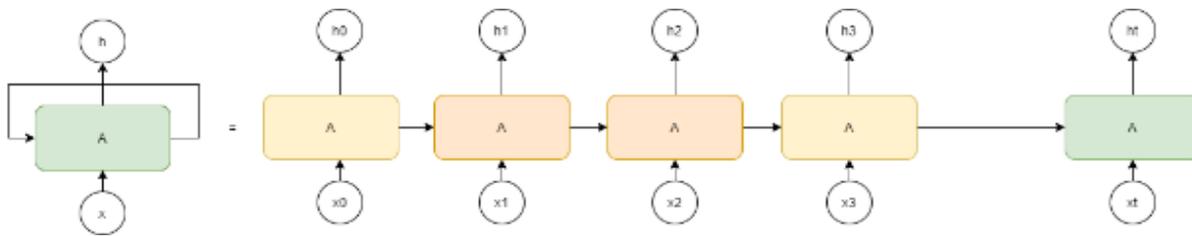
Penelitian [30] menyajikan pendekatan untuk memprediksi rasio pasar saham menggunakan jaringan saraf tiruan dengan dua teknik yang berbeda yaitu BPA-MLP (*Multilayer Perceptrons*) dan RNN-LSTM. Pengujian dilakukan pada dataset yang berbeda, seperti saham Facebook, saham Google, dan saham Bitcoin. Hasil penelitian ini menunjukkan bahwa akurasi kasus terbaik sebesar 97% untuk algoritma MLP dan 99,5% untuk algoritma LSTM.

B. Indeks Standar Pencemar Udara (ISPU)

ISPU adalah standar perhitungan indeks kualitas udara yang dipergunakan secara resmi di Indonesia. Perhitungan indeks tersebut sesuai dengan Keputusan Menteri Negara Lingkungan Hidup nomor: KEP 45/MENLH/1997 tentang Indeks Standar Pencemar Udara. Dalam [31] didefinisikan bahwa Indeks Standar Pencemar Udara adalah angka yang menggambarkan kondisi kualitas udara ambien di lokasi dan waktu tertentu yang didasarkan kepada dampak terhadap kesehatan manusia, nilai estetika dan makhluk hidup lainnya. Nilai ISPU tidak mempunyai satuan. Parameter yang termasuk dalam ISPU adalah *Particulate Matter* (PM10), Karbon Monoksida (CO), Sulfur Dioksida (SO₂), Nitrogen Dioksida (NO₂) dan Ozon (O₃). Menurut [32], penentuan ISPU dihitung dengan persamaan (1) dengan mengacu pada Batas Indeks Standar Pencemar Udara dalam Satuan SI seperti pada Tabel I.

TABEL I
BATAS INDEKS STANDAR PENCEMAR UDARA DALAM SATUAN SI.

| Indeks Standar Pencemar Udara | PM10 (µg/m ³) | SO ₂ (µg/m ³) | CO (µg/m ³) | O ₃ (µg/m ³) | NO ₂ (µg/m ³) |
|-------------------------------|---------------------------|--------------------------------------|-------------------------|-------------------------------------|--------------------------------------|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 50 | 50 | 80 | 5 | 120 | 0 |
| 100 | 150 | 365 | 10 | 235 | 0 |
| 200 | 350 | 800 | 17 | 400 | 1130 |
| 300 | 420 | 1600 | 34 | 800 | 2260 |
| 400 | 500 | 2100 | 46 | 1000 | 3000 |
| 500 | 600 | 2620 | 57,5 | 1200 | 3650 |



Gambar 1. Perulangan RNN.

$$I = \frac{Ia - Ib}{Xa - Xb}(Xx - Xb) + Ib, \tag{1}$$

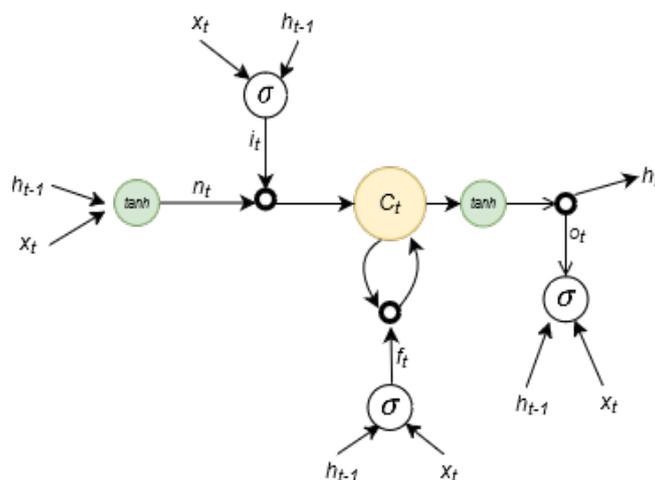
di mana:

- I = ISPU terhitung
- Ia = ISPU batas atas
- Ib = ISPU batas bawah
- Xa = Ambien batas atas
- Xb = Ambien batas bawah
- Xx = Kadar ambien nyata hasil pengukuran

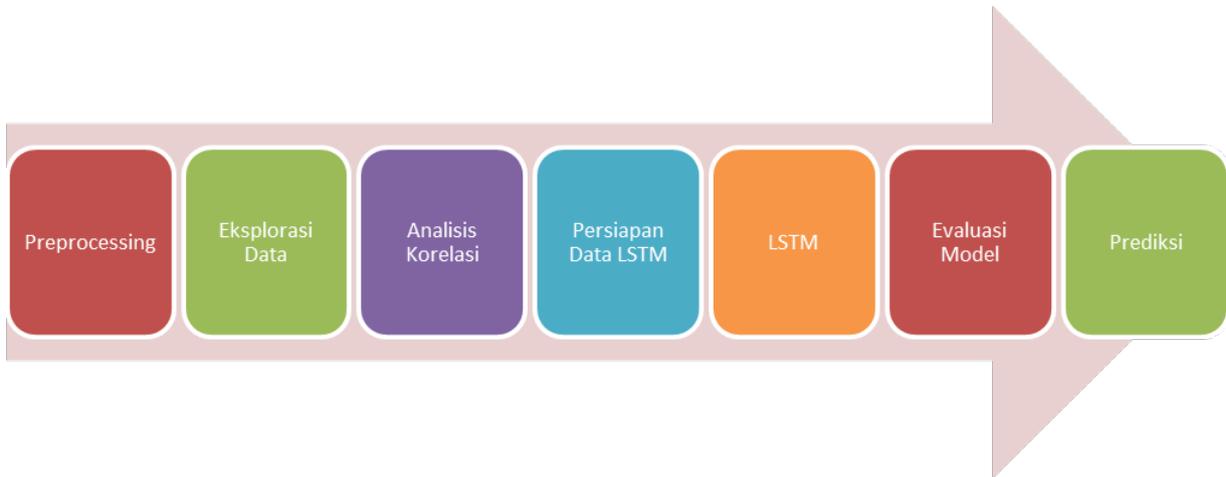
C. Recurrent Neural Network-Long Short Term Memory

Jaringan saraf berulang atau *Recurrent Neural Network* (RNN) adalah jenis *artificial intelligence* yang membutuhkan banyak lapisan (*layer*) untuk memproses data masukan, sehingga RNN dimasukkan dalam kategori *deep learning*. Proses berjalannya jaringan adalah masukan di proses dan di panggil secara berulang-ulang. Masukan yang di pakai biasanya adalah data sekuensial. Suatu sampel termasuk dalam data sekuensial apabila mempunyai suatu urutan waktu atau suatu sampel di dapat secara berurutan. Selain itu suatu sampel dalam urutan mempunyai hubungan erat satu dengan yang lain. Proses perulangan yang dilakukan dalam arsitektur RNN dapat menyimpan informasi dari masa lalu, sehingga secara otomatis informasi tersebut di simpan dalam jaringan dan digunakan untuk pembelajaran selanjutnya.

Gambar 1 adalah potongan arsitektur jaringan RNN A. RNN tersebut mendapat *input* x_t dan menghasilkan *output* h_t . Alur perulangan tersebut memproses informasi dari satu bagian menuju bagian selanjutnya. Perulangan RNN sebenarnya akan memproses *input* dari skala waktu 0 sampai t . Namun jaringan RNN tidak dapat menghubungkan pembelajaran informasi yang lama karena memori lama yang tersimpan akan tergantikan dengan memori baru seiring waktu berjalan. Kekurangan ini telah ditemukan oleh [33] sehingga dalam jaringan LSTM dikembangkan sel memori (*memory cells*) dan unit gerbang (*gate units*) untuk mengatur memori pada setiap masukannya. Cara kerja LSTM adalah membuat perubahan pada RNN dengan menambahkan *memory cell* yang dapat menyimpan informasi untuk jangka waktu yang lama. *Memory cells* digunakan untuk mengatasi terjadinya *vanishing gradient* pada RNN saat memproses data sekuensial yang panjang.



Gambar 2. Struktur Jaringan LSTM.



Gambar 3. Metodologi penelitian.

Gambar 2 merupakan struktur dari blok memori LSTM yang terdiri dari sel memori dan gerbang. x_t dan h_t merupakan *input* dan kondisi tersembunyi, pada waktu t dan f_t , i_t , dan o_t adalah gerbang yang dikenal dengan *forget gates*, *input gates*, dan *output gates*. n_t merupakan kandidat dari *input* yang disimpan, dan jumlah yang di simpan tersebut kemudian akan di kontrol oleh sebuah gerbang *input*. Kalkulasi dari setiap gerbang, kandidat *input*, kondisi sel, dan kondisi tersembunyi dilakukan oleh persamaan di bawah ini:

$$f_t = \sigma(A_f x_t + B_f h_{t-1} + b_f) \quad (2)$$

$$i_t = \sigma(A_i x_t + B_i h_{t-1} + b_i) \quad (3)$$

$$o_t = \sigma(A_o x_t + B_o h_{t-1} + b_o) \quad (4)$$

$$n_t = \tanh(A_n x_t + B_n h_{t-1} + b_n) \quad (5)$$

$$c_t = c_{t-1} * f_t + n_t * i_t \quad (6)$$

$$h_t = \tanh(c_t) * o_t \quad (7)$$

$$\sigma(z) = (1 + e^{-z})^{-1} \quad (8)$$

di mana A dan B adalah matriks bobot dan b adalah vektor-vektor bias. Simbol * merupakan perkalian *element-wise* atau perkalian satu-satu setiap elemen. Fungsi aktivasi *tanh* digunakan untuk memperbaharui kondisi sembunyi. Fungsi aktivasi *sigmoid* untuk gerbang ditandai oleh sigma pada rumus (3) dan dihitung oleh fungsi *sigmoid* pada rumus (8), dimana keluaran dari fungsi tersebut berada diantara [0,1]. Jika keluaran bernilai 0, berarti tidak ada informasi yang masuk ke dalam jaringan. Jika keluaran bernilai lebih dari 0, bermakna lebih banyak informasi yang dibolehkan. Sedangkan apabila keluaran bernilai 1, maka semua informasi dapat digunakan. Berdasarkan pada refleksi selektif seperti ini, maka LSTM memiliki kapabilitas dalam proses belajar yang lebih lama untuk pola-pola temporal.

D. Root Mean Square Error (RMSE)

Sebuah model peramalan yang dibangun perlu dilakukan evaluasi untuk mengukur tingkat akurasi atau tingkat kesalahan hasil prakiraan yang dibuat. Menurut [34], salah satu metode alternatif untuk mengevaluasi ukuran kesalahan dalam peramalan adalah nilai tengah akar kuadrat atau *Root Mean Square Error* (RMSE). Evaluasi RMSE dihitung berdasarkan nilai rata-rata dari jumlah kuadrat kesalahan antara hasil prakiraan dengan nilai aktual. Nilai digunakan untuk mengukur besarnya kesalahan yang dihasilkan oleh model prakiraan. Suatu model prakiraan mendekati variasi nilai aktualnya jika hasil perhitungan RMSE rendah. Persamaan yang digunakan untuk menghitung paramater tersebut adalah sebagai berikut:

$$RMSE = \sqrt{\frac{1}{m \times n} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} [f(i, j) - g(i, j)]^2}. \quad (9)$$

III. METODOLOGI

Penelitian ini diselesaikan melalui beberapa tahapan utama yaitu *preprocessing*, eksplorasi data, analisis korelasi, membangun model LSTM dan evaluasi model. Metodologi penelitian seperti terlihat pada Gambar 3.

Out[25]:

| day time | CO_ISPU | DD | DD_Boe | FF | FF_Boe | GRAD | HUMair | TEMPair |
|---------------------|---------|------------|------------|---------|---------|---------|------------|-----------|
| 2015-03-14 00:30:00 | 1.1400 | 296.910004 | 282.029999 | 2.30880 | 13.6850 | 1.05950 | 100.000000 | 26.833000 |
| 2015-03-14 01:00:00 | 1.1400 | 299.429993 | 292.450012 | 1.72050 | 12.8990 | 1.05950 | 100.000000 | 26.825001 |
| 2015-03-14 01:30:00 | 1.1115 | 295.049988 | 355.640015 | 3.76080 | 13.5280 | 1.05950 | 100.000000 | 26.737000 |
| 2015-03-14 02:00:00 | 1.1400 | 289.369995 | 286.549988 | 3.41000 | 14.3140 | 1.05950 | 100.000000 | 26.798000 |
| 2015-03-14 02:30:00 | 1.1400 | 261.929993 | 291.690002 | 1.86330 | 9.7526 | 1.05950 | 100.000000 | 26.639999 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2019-08-04 22:00:00 | 10.7930 | 73.483002 | 27.249001 | 1.06590 | 5.6628 | 0.62913 | 69.844002 | 25.659000 |
| 2019-08-04 22:30:00 | 7.9987 | 120.269997 | 190.240005 | 0.44746 | 5.5055 | 0.55825 | 73.008003 | 25.502001 |
| 2019-08-04 23:00:00 | 7.7557 | 98.171997 | 107.480003 | 0.48607 | 3.6179 | 0.73545 | 74.320999 | 25.198999 |
| 2019-08-04 23:30:00 | 7.7557 | 24.028000 | 185.820007 | 0.23439 | 5.8201 | 0.65445 | 75.448998 | 25.118999 |
| 2019-08-04 23:59:00 | 7.7557 | 79.831001 | 79.732002 | 0.19430 | 3.3033 | 0.71267 | 77.317001 | 25.136999 |

77040 rows x 8 columns

Gambar 4. Hasil *preprocessing*.

A. Parameter Penelitian

Data penelitian ini berasal dari stasiun pemantauan udara ambien di Dinas Lingkungan Hidup Kota Surabaya tanggal 14 Maret 2015 s.d. 4 Agustus 2019. Dinas Lingkungan Hidup Kota Surabaya memiliki 7 SUF (Stasiun Pemantauan Udara Ambien), namun dalam penelitian ini hanya fokus pada 1 stasiun yaitu di Taman Prestasi (SUF1). Pemantauan dilakukan setiap setengah jam sekali, sehingga total data adalah 77.039 sampel dalam sehari. Laporan yang dikeluarkan dari stasiun pemantauan udara menggunakan satuan mg/m^3 sehingga harus dilakukan konversi ke dalam perhitungan standar ISPU. Parameter udara yang digunakan pada penelitian ini adalah CO (*Carbon Monoxide* - Karbon Monoksida) serta dikorelasikan dengan data meteorologi, yaitu: DD (*Wind direction* - Arah angin), DD_Boe (*Wind gust direction* - Arah hembusan angin), FF (*Wind velocity* - Kecepatan angin), FF_Boe (*Wind gust velocity* - Kecepatan hembusan angin), GRAD (*Global radiation* - Radiasi global), TEMPair (*Temperature air* – Temperatur udara), dan HUMair (*Humidity air* - Kelembaban udara).

B. Preprocessing Data

Pada tahap ini, data mentah dalam bentuk *excel* digabungkan menjadi satu file dataset dan disimpan dalam format CSV (*Comma Sparared Value*). Informasi data yang ada dalam tiap file laporan adalah tanggal, parameter, satuan, waktu (pukul) saat pengambilan data, dan lokasi stasiun pemantauan (SUF-1 sampai dengan SUF-7). Dari laporan tersebut data yang di-ekstrak adalah tanggal, parameter, waktu, SUF-1. Jika ada nilai yang bukan angka (--) dan nilai 0 diganti dengan *nan values*. Hal ini dimaksudkan agar mudah dalam proses selanjutnya. Satuan CO yang dikeluarkan dari stasiun pemantauan udara menggunakan satuan mg/m^3 kemudian dikonversikan dalam perhitungan standar ISPU. Hasil data *preprocessing* dapat dilihat pada Gambar 4.

C. Eksplorasi Data

Eksplorasi data ini dimaksudkan untuk melihat karakteristik data seperti tren naik dan turun kandungan Karbon Monoksida. Analisis tren ini dilakukan untuk data harian, mingguan dan bulanan. Data digambarkan pada *chart line*. Selain itu akan dilihat jumlah data yang hilang.

D. Analisis Korelasi

Dalam penelitian ini, analisis korelasi antar parameter polutan CO dengan parameter meteorologi dilakukan menggunakan analisis Pearson. Analisis korelasi ini dimaksudkan untuk menilai seberapa besar pengaruh data meteorologi mempengaruhi nilai polutan CO. Hasil analisis korelasi parameter polutan CO dengan meteorologi diinterpretasikan untuk melihat kekuatan dan arah hubungan dua variabel. Hasil analisis korelasi tersebut berupa angka koefisien. Adapun angka koefisien memiliki kriteria sebagai berikut:

- Kedua variabel tidak mempunyai hubungan, jika angka koefisien korelasi bernilai 0
- Kedua variabel mempunyai hubungan semakin kuat, jika angka koefisien korelasi mendekati 1
- Kedua variabel mempunyai hubungan semakin lemah, jika angka koefisien korelasi mendekati 0
- Kedua variabel disebut berhubungan linier sempurna positif, jika angka koefisien korelasi sama dengan 1
- Kedua variabel disebut berhubungan linier sempurna negatif, jika angka koefisien korelasi sama dengan -1

TABEL II
DATASET POLUTAN CO (MG/M³).

| tanggal | jam | CO |
|------------|----------|---------|
| 17/07/2017 | 00.30.00 | 1,1904 |
| 17/07/2017 | 01.00.00 | 0,98445 |
| 17/07/2017 | 01.30.00 | 0,8981 |
| 17/07/2017 | 02.00.00 | 0,70745 |
| 17/07/2017 | 02.30.00 | 0,54753 |
| 17/07/2017 | 03.00.00 | 0,5611 |
| 17/07/2017 | 03.30.00 | 0,73969 |
| 17/07/2017 | 04.00.00 | 0,83613 |
| 17/07/2017 | 04.30.00 | 0,79978 |
| 17/07/2017 | 05.00.00 | 0,70891 |
| 17/07/2017 | 05.30.00 | 1,0624 |
| 17/07/2017 | 06.00.00 | 1,4158 |
| 17/07/2017 | 06.30.00 | 2,2167 |
| 17/07/2017 | 07.00.00 | 3,8966 |
| 17/07/2017 | 07.30.00 | 4,7683 |

TABEL III
MODEL SUPERVISED LEARNING PROBLEM.

| var(t-3) | var(t-2) | var(t-1) | var(t) |
|----------|----------|----------|---------|
| 0 | 1,1904 | 0,98445 | 0,8981 |
| 1,1904 | 0,98445 | 0,8981 | 0,70745 |
| 0,98445 | 0,8981 | 0,70745 | 0,54753 |
| 0,8981 | 0,70745 | 0,54753 | 0,5611 |
| 0,70745 | 0,54753 | 0,5611 | 0,73969 |
| 0,54753 | 0,5611 | 0,73969 | 0,83613 |
| 0,5611 | 0,73969 | 0,83613 | 0,79978 |
| 0,73969 | 0,83613 | 0,79978 | 0,70891 |
| 0,83613 | 0,79978 | 0,70891 | 1,0624 |
| 0,79978 | 0,70891 | 1,0624 | 1,4158 |
| 0,70891 | 1,0624 | 1,4158 | 2,2167 |
| 1,0624 | 1,4158 | 2,2167 | 3,8966 |
| 1,4158 | 2,2167 | 3,8966 | 4,7683 |

Interpretasi selanjutnya adalah melihat arah hubungan variabel. Arah hubungan variabel ada dua, yaitu searah dan tidak searah. Arah korelasi dilihat dari angka koefisien korelasi. Misalkan parameter meteorologi disimbolkan sebagai X dan polutan CO sebagai Y. Hubungan searah artinya angka koefisien bernilai positif atau > 0 . Dalam hubungan dengan variabel adalah jika variabel X nilainya tinggi, maka variabel Y juga tinggi. Sedangkan hubungan tidak searah adalah jika angka koefisien bernilai negatif. Dalam hubungan dengan variabel adalah jika variabel X nilainya tinggi, maka variabel Y bernilai rendah.

E. Persiapan Data LSTM

Dalam tahap ini data dipersiapkan masuk ke *machine learning*. Tahapannya adalah interpolasi data, normalisasi data dan mengubah dataset ke dalam model *Supervised Learning Problem*. Interpolasi data adalah cara untuk mengisi nilai yang kosong pada deret data *time series*. Tujuannya adalah agar memperoleh hasil prediksi yang baik karena data tiap waktu memiliki nilai. Dalam penelitian ini interpolasi yang digunakan adalah *spline*.

Normalisasi adalah proses perubahan nilai pada rentang tertentu dengan penskalaan data. Dalam penelitian ini, penulis menggunakan metode *Min-Max Normalization*. Normalisasi *Min-Max* bertujuan untuk merubah nilai seluruh variabel yang digunakan sebagai masukan pada rentang nilai antara 0 sampai 1. Sebelum masuk dalam proses *machines learning*, dataset harus diubah ke dalam model *Supervised Learning Problem*. *Supervised Learning Problem* adalah membingkai ulang masalah pembelajaran diawasi dari data masukan dan keluaran. Caranya adalah menggunakan langkah waktu sebelumnya sebagai variabel *input* dan menggunakan langkah waktu berikutnya sebagai variabel *output*. Tabel II merupakan nilai polutan CO pada tanggal 17/07/2017 dari pukul 00.30.00 sampai dengan 07.30.00. Jika diubah ke dalam model *Supervised Learning Problem* maka bisa dilihat pada tabel III. Di tabel III dapat diartikan bahwa $var(t)$ adalah nilai yang akan diprediksi yang masih terpengaruh nilai $var(t-1)$, $var(t-2)$, $var(t-3)$.

F. Inisialisasi Parameter LSTM

Setelah dataset siap dalam bentuk *Supervised Learning Problem*, langkah selanjutnya adalah menentukan inisialisasi parameter LSTM yang dibutuhkan. Langkah ini terdiri dari membagi dataset menjadi data latihan

(*training*) dan data uji (*testing*), mendefinisikan neuron (*hidden layer*), dan mendefinisikan jumlah *epoch* untuk iterasi dalam satu jaringan LSTM. Setelah itu proses pembelajaran jaringan LSTM bisa dilakukan. Model LSTM memerlukan masukan dengan format *batch size*, *time steps*, dan *features* [35]. *Batch size* menunjukkan banyak sampel data latih yang ingin dimasukkan sebelum memperbaharui bobot dari sebuah jaringan saraf tiruan, *time steps* mendefinisikan banyak indeks waktu yang ingin dipelajari, dan *features* adalah jumlah atribut yang digunakan untuk merepresentasikan setiap *time steps*. Dalam penelitian prediksi polusi udara di kota Surabaya ini, nilai *batch size* yang digunakan adalah 150, dengan nilai *time steps* 1, dan nilai *features* 8. Dalam pembelajaran *neural network* sering muncul masalah *overfitting*, dimana salah satu cara mengatasinya adalah dengan menggunakan teknik regularisasi *dropout* [36]. Pada penelitian ini, *dropout* diberikan nilai peluang sebesar 0,2.

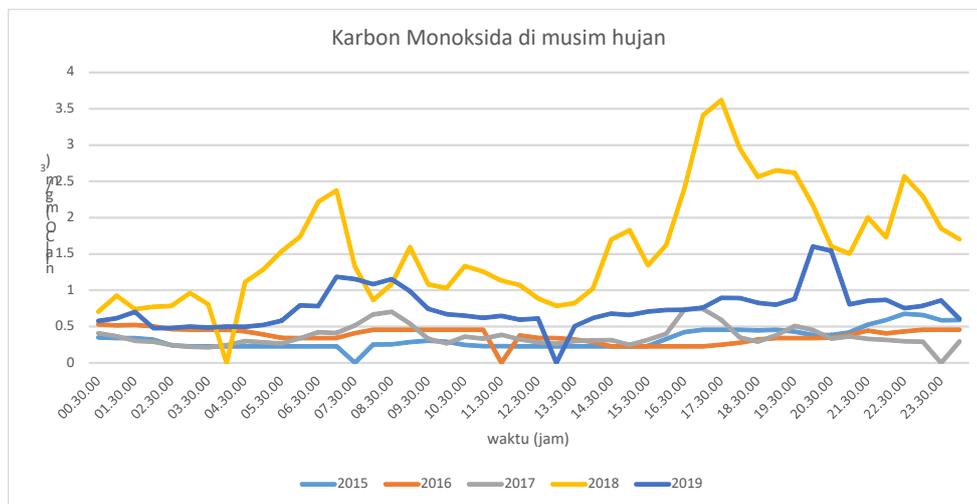
G. Evaluasi Model

Jika proses iterasi jaringan LSTM selesai dan mendapatkan model, kemudian dilakukan pengujian dengan data *testing*. Proses yang dilakukan antara lain membuat prediksi, denormalisasi dan perhitungan *error* dengan metode RMSE. Prediksi dibuat berdasarkan model pada proses algoritma LSTM. Hasil prediksi masih dalam nilai normalisasi *Min Max*, sehingga perlu dilakukan konversi ke nilai aslinya. Setelah didapat nilai asli, maka hasil prediksi dibandingkan dengan nilai aktual menggunakan perhitungan evaluasi RMSE.

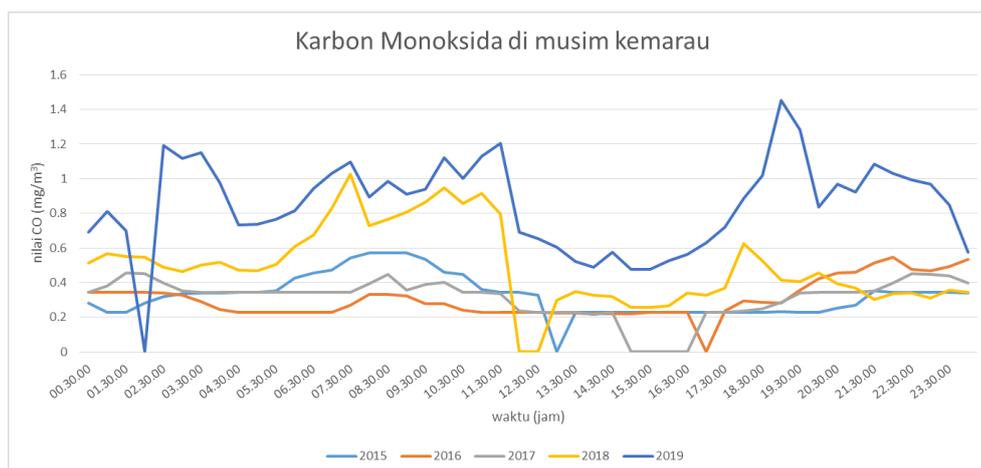
IV. HASIL DAN DISKUSI

A. Eksplorasi Data

Indonesia merupakan negara yang secara astronomis terletak di garis khatulistiwa sehingga beriklim tropis. Musim yang dimiliki Indonesia ada dua, yaitu musim hujan dan musim kemarau. Musim kemarau berlangsung antara bulan April-September dan bersamaan dengan bertiupnya angin musim timur. Sedangkan musim hujan terjadi antara bulan Oktober-Februari dan disertai bertiupnya angin musim barat. Bulan Maret merupakan bulan peralihan musim, yang biasanya disebut musim pancaroba.



Gambar 5. Grafik harian karbon monoksida pada musim hujan.



Gambar 6. Grafik harian karbon monoksida pada musim kemarau.

TABEL IV
KORELASI ANTAR-VARIABEL.

| Parameter | CO |
|-----------|-----------|
| CO | 1 |
| DD | 0.104867 |
| DD_Boe | 0.053871 |
| FF | -0.235809 |
| FF_Boe | -0.308153 |
| GRAD | -0.113533 |
| HUMair | 0.224145 |
| TEMPair | -0.206248 |

TABEL V
KOMBINASI PENGUJIAN BERDASARKAN JUMLAH *NEURON* DAN PEMBAGIAN DATA UJI DAN DATA LATIH.

| Max Epoch | Neuron Hidden | SKENARIO I | SKENARIO II | SKENARIO III |
|-----------|---------------|------------|-------------|--------------|
| | | 75% : 25% | 85% : 15% | 95% : 5% |
| 50 | 2 | 2,884 | 3,108 | 1,880 |
| 50 | 5 | 2,827 | 3,114 | 1,911 |
| 50 | 10 | 2,740 | 3,097 | 1,900 |
| 50 | 15 | 2,796 | 3,042 | 1,904 |
| 50 | 20 | 2,804 | 3,056 | 1,915 |
| 50 | 50 | 2,791 | 3,013 | 1,895 |
| 50 | 75 | 2,767 | 3,016 | 1,935 |
| 50 | 100 | 2,772 | 2,994 | 1,901 |

Gambar 5 dan Gambar 6 menunjukkan bahwa ada perbedaan tren harian pada nilai Karbon Monoksida antara musim hujan dan musim kemarau. Data diambil secara acak di dua musim, tanpa membedakan hari kerja atau hari libur. Di musim hujan atau pun musim kemarau, tren mulai naik terjadi rata-rata pada pukul 05.00 dan 16.00. Perbedaan terjadi pada tren turun. Di musim hujan nilai Karbon Monoksida mulai turun pada pukul 09.00, sedangkan pada musim kemarau tren turun terjadi pada pukul 10.00 atau 11.30.

Analisis *missing values* dari seluruh variabel terakumulasi sebesar 31,7% dengan rincian variabel CO memiliki *missing values* sebesar 5,29% setara dengan 4.077 dari 77.040 data pengamatan, variabel DD sebesar 3,60% setara dengan 2.774 data pengamatan, DD Boe sebesar 3,58% setara dengan 2.761 data pengamatan, FF sebesar 3,58% setara dengan 2.760 data pengamatan, FF Boe sebesar 3,55% setara dengan 2.734 data pengamatan, GRAD sebesar 10,38% setara dengan 8.000 data pengamatan, HUMair sebesar 0,87% setara dengan 671 data pengamatan, dan TEMPair sebesar 0,83% setara dengan 639 data pengamatan.

B. Analisis Korelasi

Hasil perhitungan korelasi antar variabel dapat dilihat pada Tabel IV. Hasil tersebut menggambarkan bahwa jika dilihat dari arahnya, maka data meteorologi yang memiliki korelasi searah dengan variabel polutan CO atau dalam hal ini bernilai positif adalah DD (*Wind direction* - Arah angin), DD_Boe (*Wind gust direction* - Arah hembusan angin) dan HUMair (*Humidity air* - Kelembaban udara). Sedangkan data meteorologi yang berkorelasi tidak searah atau bernilai negatif adalah FF (*Wind velocity* - Kecepatan angin), FF_Boe (*Wind gust velocity* - Kecepatan hembusan angin), GRAD (*Global radiation* - Radiasi global) dan TEMPair (*Temperature air* - Temperatur udara). Rata-rata nilai korelasi negatif adalah -0,22 sehingga dapat disebutkan bahwa data meteorologi cukup mempengaruhi nilai polutan CO. Artinya bahwa jika data meteorologi yang berkorelasi negatif nilainya bertambah tinggi, maka akan mengurangi nilai polutan CO.

C. Analisis Prediksi menggunakan RNN-LSTM

Skenario pengujian dalam penelitian ini adalah menganalisis dampak jumlah neuron *hidden* dan kombinasi data latih dan data uji terhadap nilai *error* yang didapatkan. Pengujian dilakukan dengan membuat beberapa skenario uji yaitu dengan membagi data latih dan data uji dengan kombinasi 75% : 25, 85% : 15%, dan 95% : 5%, mengubah jumlah neuron *hidden* dari 2 sampai dengan 100, serta jumlah *epoch* 50.

Tabel V merupakan hasil pengujian dengan menggunakan beberapa komposisi jumlah neuron *hidden* dan pembagian data latih – data uji. Dari hasil pengujian didapatkan hasil yang berbeda dilihat dari pembagian dataset menjadi data latih dan data uji. Pertama, dengan komposisi data latih 75% dan data uji 25% atau pembagian data latih sejumlah 57.779 sampel dan 19.260 sampel untuk data uji, mendapatkan hasil terbaik pada jumlah neuron *hidden* 10 dengan nilai RMSE 2,74. Komposisi kedua adalah dengan data latih 85% dan data uji 15% atau pembagian data latih sejumlah 65.483 sampel dan data uji sejumlah 11.556 sampel, mendapatkan hasil terbaik pada

jumlah neuron *hidden* 100 dengan nilai RMSE 2,994. Komposisi ketiga dengan data latih 95% dan data uji 5% atau pembagian data pelatihan sejumlah 73187 sampel dan 3852 sampel untuk data uji, mendapatkan hasil terbaik pada jumlah neuron 2 dengan nilai RMSE 1,880. *Hidden neuron* bekerja dengan mengolah nilai masukan dan menghubungkannya dengan neuron keluaran, sehingga jumlah *hidden neuron* akan menentukan nilai keluaran yang dihasilkan oleh LSTM. Tidak ada aturan pasti mengenai jumlah hidden neuron yang paling optimal dalam memprediksi *time series*. Sehingga jumlah hidden neuron yang didapatkan melalui percobaan. Setelah mendapatkan *error* terkecil dari ketiga skenario pengujian, selanjutnya data akan dilakukan validasi. Validasi dilakukan dengan cara hasil prediksi dibandingkan dengan hasil aktual dengan jumlah data sesuai dengan skenario pada Tabel V. Skenario dengan data validasi tersebut kemudian divisualisasikan dalam bentuk grafik seperti terlihat pada Gambar 7. Gambar 8 menunjukkan hasil komparasi antara nilai aktual dan prediksi untuk 48 data (dalam 1 hari). Gambar 8 menunjukkan bahwa model yang dibangun menggunakan algoritma RNN-LSTM dapat mempelajari tren harian dengan baik.

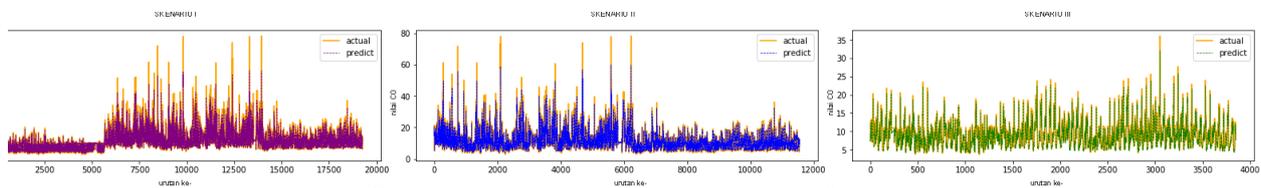
Dari hasil analisis yang telah dibuat dapat diambil *knowledge discovery* sebagai berikut:

1. Analisis tren baik harian, mingguan maupun bulanan diharapkan dapat memberikan informasi yang penting untuk pemangku kepentingan di bidang pengendalian polusi udara untuk membuat langkah-langkah nyata dalam pengendalian polusi udara pada saat tren naik, misalnya pembatasan kendaraan bermotor, pengalihan jalur transportasi pada jalan-jalan yang tidak terlalu ramai, mendorong masyarakat untuk lebih memanfaatkan transportasi umum, menggalakkan kegiatan *Car Free Day* dan lain sebagainya. Untuk pencegahan polusi udara dalam jangka waktu yang panjang misalnya dengan penambahan Ruang Terbuka Hijau sebagai paru-paru kota.
2. Analisis korelasi Karbon Monoksida dengan data meteorologi menunjukkan bahwa kandungan Karbon Monoksida pada Taman Prestasi tidak hanya dihasilkan dari tempat itu saja, melainkan ada pengaruh dari kondisi di luar tempat tersebut. Hal ini dapat dilihat bahwa arah angin dan arah hembusan jika menuju ke tempat tersebut serta kelembaban udara tinggi maka akan mempengaruhi kenaikan kandungan Karbon Monoksida di Taman Prestasi.
3. Analisis Algoritma RNN-LSTM menunjukkan bahwa algoritma ini sangat bagus untuk membuat peramalan data polusi udara di kota Surabaya yang bersifat musiman dan secara umum model yang dibuat memberikan hasil yang baik dengan dilihat dari perhitungan nilai RMSE yang relatif kecil.

V. KESIMPULAN

Penelitian ini membuat prediksi polutan Karbon Monoksida berdasarkan nilai di masa lalu dengan dikorelasikan data meteorologi seperti arah angin, arah hembusan angin, kecepatan angin, kecepatan hembusan angin, radiasi global, kelembaban udara, dan temperatur udara. Data diperoleh dari Dinas Lingkungan Hidup Kota Surabaya mulai dari tanggal 14 Maret 2015 sampai dengan 4 Agustus 2019. Dari data tersebut dibuat analisis tren, analisis korelasi nilai polutan terhadap data meteorologi, dan analisis prediksi menggunakan algoritma RNN-LSTM.

Hasil analisis tren menunjukkan bahwa tren naik terjadi pada pukul 09.00 dan 16.00, sedangkan tren turun pada musim penghujan terjadi pada pukul 09.00 dan pada musim kemarau pada pukul 10.00 atau 11.30. Hasil analisis



Gambar 7. Validasi data uji dengan data aktual pada 3 skenario.



Gambar 8. Grafik komparasi nilai aktual dan nilai prediksi dalam satu hari.

korelasi polutan Karbon Monoksida dengan data meteorologi menunjukkan bahwa nilai CO berbanding lurus

dengan DD, DD_Boe, dan HUMair, serta berbanding terbalik dengan FF, FF_Boe, GRAD, dan TEMPair. Hasil analisis prediksi menggunakan RNN-LSTM menunjukkan bahwa dari beberapa skenario pembagian data latih dan data uji, menghasilkan model terbaik pada komposisi data latih dan data uji 95% : 5%. Dengan menggunakan perhitungan *error* RMSE didapat nilai sebesar 1,880. Jadi dapat disimpulkan bahwa RNN-LSTM merupakan algoritma yang menggunakan pendekatan data, sehingga performa model akan semakin baik seiring dengan bertambahnya data. Jika data semakin besar, jaringan *neural networks* akan mengurangi nilai *error* yang didapat. Artinya akurasi prediksi bisa meningkat atau semakin baik.

UCAPAN TERIMA KASIH

Atas selesai hingga terbitnya artikel ini, ucapan terima kasih dan penghargaan yang setinggi-tingginya diberikan kepada Bapak Dr. Adhi Dharma Wibawa selaku koordinator program S2 PETIK serta atas kebersamaan dan dukungan yang begitu luar biasa dari teman-teman Mahasiswa S2 Program Pengelola TIK Pemerintahan (PETIK) hasil kerja sama ITS dengan Kementerian Komunikasi dan Informatika Angkatan 2018. Dan tidak ketinggalan ucapan terima kasih dihaturkan kepada Dinas Lingkungan Hidup Kota Surabaya yang telah mendukung dataset dalam penelitian ini.

DAFTAR PUSTAKA

- [1] M. E. Marlier, A. S. Jina, P. L. Kinney, R. S. DeFries, "Extreme Air Pollution in Global Megacities," *Current Climate Change Reports*, vol. 2, hal. 15-27, 2016.
- [2] Health Effects Institute. 2018. State of Global Air 2018. Special Report. Boston, MA:Health Effects Institute.
- [3] World Health Organization, "World health statistics 2018: monitoring health for the SDGs," *Sustainable development goals*, Geneva, 2018.
- [4] Badan Pusat Statistik.Jumlah Kendaraan Bermotor. Tersedia: <https://www.bps.go.id/linkTableDinamis/view/id/1133>
- [5] A. Syahrani, "Analisis Kinerja Mesin Bensin Berdasarkan Hasil Uji Emisi," *SMARTek*, vol. 4, no. 4, 2006.
- [6] A. Caragliu, C. Del Bo, dan P. Nijkamp, "Smart cities in Europe," *Serie Research Memoranda 0048*, VU University Amsterdam, 2009.
- [7] C. Benevolo, R. P. Dameri, B. D'Auria, "Smart Mobility in Smart City," dalam *Lecture Notes in Information Systems and Organisation*, vol. 11, Springer, Cham, 2016.
- [8] A. Bhati, M. Hansen, dan C. M. Chan, "Energy conservation through smart homes in a smart city: A lesson for Singapore households," *Energy Policy*. vol. 104, hal. 230-239, 2017.
- [9] S. Suakanto, S. H. Supangkat, Suhardi, dan R. Saragih, "Smart city dashboard for integrating various data of sensor networks," dalam *Proc. International Conference on ICT for Smart Society*, 2013.
- [10] N. Suri, Z. Zielinski, M. Tortonesi, C. Fuchs, M. Pradhan, K. Wrona, J. Furtak, D. B. Vasilache, M. Street, V. Pellegrini, G. Benincasa, A. Morelli, C. Stefanelli, E. Casini, dan M. Dyk, "Exploiting smart city IoT for disaster recovery operations," dalam *Proc. IEEE World Forum on Internet of Things*, hal. 458-463, 2018.
- [11] E. Al Nuaimi, H. Al Neyadi, N. Mohamed, dan J. Al Jaroodi "Applications of big data to smart cities," *J. Internet Serv Appl.*, vol 6, no. 25, 2015.
- [12] H. Kumar, M. K. Singh, M. P. Gupta, dan J. Madaan. "Moving towards smart cities: Solutions that lead to the Smart City Transformation Framework". *Technological Forecasting & Social Change*, vol. 153, 2020.
- [13] Sumayang, Lalu. 2003. Dasar-Dasar Manajemen Produksi dan Operasi. Salemba Empat, Jakarta.
- [14] Supranto J. 2000. Statistik (Teori dan Aplikasi), Edisi Keenam. Jakarta: Erlangga.
- [15] Heizer, Jay dan Render, Barry. 2009. Manajemen Operasi, Buku 1 Edisi 9. Jakarta: Salemba Empat.
- [16] X. Zhang, M. H. Chen, dan Y. Qin, "NLP-QA Framework Based on LSTM-RNN," dalam *Proc. International Conference on Data Science and Business Analytics*, hal. 307-311, 2018.
- [17] S. Abujar, A. K. M. Masum, S. M. M. H. Chowdhury, M. Hasan, dan S. A. Hossain, "Bengali Text generation Using Bi-directional RNN," dalam *Proc. International Conference on Computing, Communication and Networking Technologies*, hal. 1-5, 2019.
- [18] H. Xie, M. A. Bin Ahmadon, S. Yamaguchi, dan I. Toyoshima, "Random Sampling and Inductive Ability Evaluation of Word Embedding in Medical Literature," dalam *Proc. IEEE International Conference on Consumer Electronics*, hal. 1-4, 2019.
- [19] S. Wang, P. Zhou, W. Chen, J. Jia, dan L. Xie, "Exploring RNN-Transducer for Chinese speech recognition," dalam *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, hal. 1364-1369, 2019.
- [20] M. Jiang, Z. Yang, dan C. Zhao, "What to play next? A RNN-based music recommendation system," dalam *Proc. Asilomar Conference on Signals, Systems, dan Computers*, hal. 356-358, 2017.
- [21] Y. Hu and S. Lin, "Deep Reinforcement Learning for Optimizing Finance Portfolio Management," dalam *Proc. Amity International Conference on Artificial Intelligence*, hal. 14-20, 2019.
- [22] M. Rozenwald, E. Khrameeva, G. Sapunov, dan M. Gelfand, "Prediction of 3D Chromatin Structure Using Recurrent Neural Networks," dalam *Proc. IEEE International Conference on Bioinformatics and Biomedicine*, hal. 2488-2488, 2018.
- [23] A. Xiao, J. Liu, Y. Li, Q. Song, dan N. Ge, "Two-phase rate adaptation strategy for improving real-time video QoE in mobile networks," *China Communications*, vol. 15, no. 10, hal. 12-24, 2018.
- [24] M. W. P. Aldi, J. Jondri, dan A. Aditsania, "Analisis Dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin," *eProceedings of Engineering*, vol. 5, no. 2, 2018.
- [25] S. Taneja, N. Sharma, K. Oberoi, dan Y. Navoria, "Predicting trends in air pollution in Delhi using data mining," dalam *Proc. India International Conference on Information Processing*, hal. 1-6, 2016.
- [26] V. Chaudhary, A. Deshbhrtar, V. Kumar, dan D. Paul, "Time Series Based LSTM Model to Predict Air Pollutant's Concentration for Prominent Cities in India," 2018.
- [27] V. Reddy, P. Yedavalli, S. Mohanty, dan U. Nakhat, "Deep Air: Forecasting Air Pollution in Beijing, China," *Environmental Science*, 2018.
- [28] S. Siami-Namini, N. Tavakoli, dan A. S. Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," dalam *Proc. IEEE International Conference on Machine Learning and Applications*, hal. 1394-1401, 2018.
- [29] R. B. Kurdikeri dan A. B. Raju, "Comparative Study of Short-Term Wind Speed Forecasting Techniques Using Artificial Neural Networks," dalam *Proc. International Conference on Current Trends towards Converging Technologies*, hal. 1-5, 2018.
- [30] R. Achkar, F. Elias-Sleiman, H. Ezzidine, dan N. Haidar, "Comparison of BPA-MLP and LSTM-RNN for Stocks Prediction," dalam *Proc. International Symposium on Computational and Business Intelligence*, hal. 48-51, 2018.
- [31] Keputusan Menteri Negara Lingkungan Hidup Nomor : KEP 45 / MENLH / 1997 Tentang Indeks Standar Pencemar Udara
- [32] Keputusan Kepala Badan Pengendalian Dampak Lingkungan No. 107 Tahun 1997 Tanggal 21 November 1997

- [33] Y. Bengio, P. Simard, dan P. Frasconi, "Learning Long-Term Dependencies with Gradient Descent is Difficult" *IEEE Transaction on Neural Networks*, vol. 5, no. 2, 1994.
- [34] S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, dan R. Winkler, "The Accuracy of Extrapolative (Time Series Methods): Results of a Forecasting Competition," *Journal of Forecasting*, vol. 1, no. 2, hal. 111-153, 1982.
- [35] A. Gulli dan S. Pal. 2017. *Deep Learning with Keras*. Mumbai: Packt Publishing.
- [36] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, dan R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15. 1929-1958, 2014.