TESIS – EE185401

# ANALISIS DATA RUNTUN WAKTU PREDIKSI POLUSI UDARA DI KOTA SURABAYA MENGGUNAKAN DEEP LEARNING RNN-LSTM

MUH. ANAS FAISHOL
07111850067005

DOSEN PEMBIMBING
Dr. Ir. Endroyono, DEA
Astria Nur Irfansyah, ST., M.Eng., Ph.D.

PROGRAM MAGISTER
BIDANG KEAHLIAN TELEMATIKA - PETIK
DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI ELEKTRO DAN INFORMATIKA CERDAS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2020

**TESIS – EE185401**

# ANALISIS DATA RUNTUN WAKTU PREDIKSI POLUSI UDARA DI KOTA SURABAYA MENGGUNAKAN DEEP LEARNING RNN-LSTM

MUH. ANAS FAISHOL
07111850067005

DOSEN PEMBIMBING
Dr. Ir. Endroyono, DEA
Astria Nur Irfansyah, ST., M.Eng., Ph.D.

PROGRAM MAGISTER
BIDANG KEAHLIAN TELEMATIKA - PETIK
DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI ELEKTRO DAN INFORMATIKA CERDAS
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2020

# LEMBAR PENGESAHAN TESIS

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
**Magister Teknik (MT)**
di
**Institut Teknologi Sepuluh Nopember**
Oleh:

**MUH. ANAS FAISHOL**
**NRP: 07111850067005**

Tanggal Ujian: 02 Juli 2020
Periode Wisuda: September 2020

Disetujui oleh:
**Pembimbing:**

1. Dr. Ir. Endroyono, DEA .....
   NIP: 196504041991021001

2. Astria Nur Irfansyah, ST., M.Eng., Ph.D. .....
   NIP: 198103252010121002

**Penguji:**

1. Dr.Ir. Wirawan, DEA .....
   NIP: 196311091989031011

2. Dr. Eko Mulyanto Yuniarno, S.T., M.T. .....
   NIP: 196806011995121009

3. Dr. Supeno Mardi Susiki Nugroho, ST., M.T. .....
   NIP: 197003131995121001

Kepala Departemen Teknik Elektro
Fakultas Teknologi Elektro dan Informatika Cerdas

Dedet Candra Riawan, ST., M. Eng., Ph.D
NIP: 197311192000031001

*Halaman ini sengaja dikosongkan*

# PERNYATAAN KEASLIAN TESIS

Dengan ini saya menyatakan bahwa isi keseluruhan Tesis saya dengan judul "**ANALISIS DATA RUNTUN WAKTU PREDIKSI POLUSI UDARA DI KOTA SURABAYA MENGGUNAKAN DEEP LEARNING RNN-LSTM**" adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diijinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri.

Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya,  Juli 2020

**Muh. Anas Faishol**

NRP. 07111850067005

*Halaman ini sengaja dikosongkan*

# ANALISIS DATA RUNTUN WAKTU PREDIKSI POLUSI UDARA DI KOTA SURABAYA MENGGUNAKAN DEEP LEARNING RNN-LSTM


Nama mahasiswa : Muh. Anas Faishol
NRP : 07111850067005
Pembimbing : 1. Dr. Ir. Endroyono, DEA
2. Astria Nur Irfansyah, ST., M.Eng., Ph.D.


## 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, NOx, 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: polusi udara, *recurrent neural network*, *long short term memory*, Surabaya .

*Halaman ini sengaja dikosongkan*

# TIME SERIES ANALYSIS FOR AIR POLLUTION PREDICTION IN SURABAYA CITY USING DEEP LEARNING RNN-LSTM


By : Muh. Anas Faishol
Student Identity Number : 07111850067005
Supervisor(s) : 1. Dr. Ir. Endroyono, DEA
2. Astria Nur Irfansyah, ST., M.Eng., Ph.D.


## 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, NOx, 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, recurrent neural network, long short term memory, Surabaya

*Halaman ini sengaja dikosongkan*

# KATA PENGANTAR

Syukur Alhamdulillah, penulis panjatkan kepada Allah SWT, berkat rahmat dan karunia-Nya tesis ini dapat diselesaikan di tengah-tengah masa pandemi COVID-19. Pada tanggal 12 Februari 2020, WHO resmi menetapkan penyakit *novel corona virus* pada manusia ini dengan sebutan *Coronavirus Disease* (COVID-19). Pada tanggal 2 Maret 2020, Indonesia melaporkan bahwa terkonfirmasi positif COVID-19. Pada tanggal 11 Maret 2020, WHO sudah menetapkan COVID-19 sebagai pandemi.

Penyelesaian tesis di tahun 2020 ini merupakan suatu proses yang istimewa yang dirasakan penulis. Dengan adanya kebijakan pencegahan penularan COVID-19, salah satunya adalah ditutupnya semua kegiatan akademik di kampus menyebabkan perubahan model bimbingan dan sidang tesis ini, yaitu dilakukan secara daring. Sehingga tak salah jika banyak kalangan menyebut bahwa lulusan tahun 2020 ini menjadi lulusan istimewa yang sering disebut sebagai lulusan Corona. Penulis sadar bahwa Negara Indonesia bahkan seluruh dunia mengalami dampak dari pandemi COVID-19. Tahun yang berat untuk kita semua akan tetapi bukan berarti membatasi dan membuat kita menjadi patah semangat dalam berkarya, bangkit dari keterpurukan.

Penulis menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam menyelesaikan tesis ini. Secara khusus penulis menyampaikan ucapan terima kasih kepada :

1. Kedua orang tua penulis, Bapak M. Amin(alm) dan Ibu Mudrikah yang selalu memberikan doa, kasih sayang dan dukungan kepada penulis dari mulai mengikuti pendidikan hingga selesainya tesis ini.
2. Dr. Ir. Endroyono, DEA dan Astria Nur Irfansyah, ST., M.Eng., Ph.D. selaku dosen pembimbing yang telah mengarahkan, memberi koreksi, memberi motivasi dan mendukung sepenuhnya dalam pembuatan tesis ini.
3. Dr. Adhi Dharma Wibawa, S.T., M.T. selaku koordinator bidang keahlian Telematika

4. Drs. Rijanto,M.M selaku Bupati Blitar, Drs. Totok Subihandono,M.Si selaku Sekretaris Daerah Kabupaten Blitar, Drs. Mahadhin C U, MM selaku Asisten Administrasi Umum Sekda Kabupaten Blitar dan Eko Susanto,S.T,M.Si selaku Kepala Dinas Komunikasi dan Informatika yang telah memberikan ijin dan rekomendasi untuk mendapatkan beasiswa PETIK (Pengelola TIK Pemerintahan)
5. Istri tercinta, Nunik Ratnawati dan Ananda Tersayang, Kenzie Aryasatya Faishol, yang telah setia menemani pengerjaan tesis ini.
6. Saudara penulis: kakak Misbahuddin dan Adek Iftahul Ula yang telah memberika motivasi dan dorongan dalam penyelesaian tesis ini.
7. Sahabat Berbagi Kisah Hidup: Aries Findra Setiawan, Made Dwi Novitasari, Ma'lufatul Maisuroh, Haryani Ambarwati yang selalu mengingatkan dan memberikan motivasi selama perkuliahan.
8. Rekan seperjuangan PETIK 2018: Danang Dismantoro, Sunarmo, Angga, Feri, Lucan, Ika, Rahma, Fauzan, Ziko Iskandar, Ibrahim Aji, Jamal, Henry, Firman Mautinu, Suyasmad yang selalu memberikan motivasi, dukungan dan semangat kepada penulis dari awal hingga akhir perkuliahan.
9. Teman-teman Dinas Lingkungan Hidup Kota Surabaya khususnya Bu Ulfi dan Mas Yudi yang telah memberikan support data guna penelitian dalam tesis ini.
10. Semua dosen, staff dan pegawai di Departemen Teknik Elektro yang tidak dapat saya sebutkan satu-persatu.
11. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu persatu yang telah membantu penyelesaian tesis ini.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna. Oleh karenanya masukan, saran dan kritik untuk perbaikan sangat diharapkan. Besar harapan tesis ini dapat bermanfaat bagi penelitian – penelitian selanjutnya.

Surabaya, Juli 2020

Penulis

# DAFTAR ISI

LEMBAR PENGESAHAN TESIS ..... iii

PERNYATAAN KEASLIAN TESIS ..... v

ABSTRAK ..... vii

ABSTRACT ..... ix

KATA PENGANTAR ..... x

DAFTAR ISI ..... xiii

DAFTAR GAMBAR ..... xvi

DAFTAR TABEL ..... xviii

BAB 1 PENDAHULUAN ..... 1

1.1 Latar Belakang ..... 1
1.2 Rumusan Masalah ..... 5
1.3 Tujuan ..... 5
1.4 Batasan Masalah ..... 6
1.5 Kontribusi ..... 6

BAB 2 KAJIAN PUSTAKA ..... 7

2.1 Kajian Penelitian Terkait ..... 7
2.2 Teori Dasar ..... 8
2.2.1 Polusi Udara ..... 8
2.2.1.1 ISPU ..... 9
2.2.1.2 Karbon monoksida (CO) ..... 11
2.2.2 Data Meteorologi ..... 13
2.2.2.1 DD (Wind direction - Arah angin) ..... 13
2.2.2.2 DD_Boe (Wind gust direction - Arah hembusan angin) ..... 13
2.2.2.3 FF (Wind velocity - Kecepatan angin) ..... 13
2.2.2.4 FF_Boe (Wind gust velocity - Kecepatan hembusan angin) ... 13
2.2.2.5 GRAD (Global radiation - Radiasi global) ..... 14
2.2.2.6 HUMair (Humidity air - Kelembaban udara) ..... 14
2.2.2.7 TEMPair (Temperature air - Temperatur udara) ..... 14
2.2.3 Time Series ..... 14

2.2.3.1 Pola Horizontal ..... 14
2.2.3.2 Pola Trend ..... 15
2.2.3.3 Pola Musiman ..... 15
2.2.3.4 Siklis ..... 15
2.2.4 Analisis Korelasi ..... 15
2.2.5 Interpolasi Data ..... 18
2.2.6 Peramalan (*Forcasting*) ..... 20
2.2.7 Neural Network ..... 20
2.2.7.1 Recurrent Neural Network ..... 21
2.2.7.2 Long Short-Term Memory ..... 22
2.2.7.3 RMSE ..... 26
2.2.8 Perangkat Pendukung ..... 26
2.2.8.1 Python ..... 26
2.2.8.2 Keras ..... 27
2.2.8.3 Tensorflow ..... 27
BAB 3 METODE PENELITIAN ..... 29
3.1 Alur Penelitian ..... 29
3.2 Data Pengamatan Udara di Kota Surabaya ..... 31
3.3 Preprocressing ..... 34
3.4 Visualisasi Data ..... 34
3.5 Analisis Korelasi ..... 35
3.6 Persiapan Data LSTM ..... 35
3.6.1 Normalisasi Data ..... 36
3.6.2 Supervised Learning Problem ..... 37
3.7 Pemodelan Algoritma RNN LSTM ..... 40
3.8 Evaluasi ..... 43
BAB 4 HASIL DAN PEMBAHASAN ..... 45
4.1 Eksplorasi Data ..... 45
4.2 Analisis Korelasi ..... 48
4.3 Analisis Meteorologis Kota Surabaya ..... 49
4.4 Analisis Prediksi Menggunakan RNN-LSTM ..... 52
BAB 5 KESIMPULAN ..... 57

DAFTAR PUSTAKA ..... 59

LAMPIRAN ..... 63

BIODATA PENULIS ..... 77

*Halaman ini sengaja dikosongkan*

# DAFTAR GAMBAR

Gambar 2.1 X dan Y mempunyai hubungan positif ..... 16
Gambar 2.2 X dan Y mempunyai hubungan negative ..... 16
Gambar 2.3 X dan Y mempunyai hubungan kuat atau hubungan lemah sekali ... 17
Gambar 2.4 Fungsi aktivasi pada neural network ..... 21
Gambar 2.5 Bentuk konseptual paling sederhana RNN..... 21
Gambar 2.6 Struktur RNN yang dijabarkan..... 22
Gambar 2.7 Detail Struktur LSTM ..... 23
Gambar 2.8 Struktur Forget Layer ..... 23
Gambar 2.9 Struktur *Remember Gate* ..... 24
Gambar 2.10 Struktur Update Layer ..... 25
Gambar 2.11 Struktur Output Layer ..... 25
Gambar 3.1 Alur Penelitian..... 30
Gambar 3.2 Peta lokasi SUF-1 Taman Prestasi ..... 32
Gambar 3.3 Laporan kualitas udara dari stasiun pemantauan kualitas udara di Kota Surabaya ..... 33
Gambar 3.4. Hasil preprocessing ..... 34
Gambar 3.5 Hasil normalisasi data ..... 36
Gambar 3.6. Proses Supervised Learning Problem..... 40
Gambar 3.7 Hasil Supervised Learning Prolem..... 40
Gambar 3.8 Bentuk 3D masukan LSTM..... 41
Gambar 4.1 Grafik Harian Karbon monoksida musim penghujan ..... 45
Gambar 4.2 Grafik Harian Karbon monoksida musim kemarau ..... 46
Gambar 4.3 Grafik Mingguan Karbon Monoksida ..... 46
Gambar 4.4 Grafik Tahunan Karbon Monoksida ..... 47
Gambar 4.5 Korelasi CO dengan data meteorologi ..... 49
Gambar 4.6 Suhu Udara antara tahun 2015 sampai dengan 2019 ..... 51
Gambar 4.6 Validasi Data Uji dengan Data Aktual pada 3 Skenario ..... 54
Gambar 4.7. Grafik Komparasi Nilai Aktual dan Nilai Prediksi dalam Satu Hari 54

*Halaman ini sengaja dikosongkan*

# DAFTAR TABEL

Tabel 2.1 Batas Indeks Standar Pencemar Udara Dalam Satuan S..... 10
Tabel 2.2 Rentang Indeks Standar Pencemar Udara..... 11
Tabel 3.1 Stasiun pemantuan kualitas udara Kota Surabaya ..... 31
Tabel 3.2 Dataset Polutan Co (mg/$m^3$) ..... 38
Tabel 3.3 Model Supervised Learning Problem ..... 39
Tabel 4.1 Analisis Data berdasarkan rata-rata, nilai minimal dan maksimal..... 48
Tabel 4.2 Indeks ISPU pada kandungan CO..... 48
Tabel 4.3 Nilai Minimal, Maksimal dan Rata-rata Suhu Udara Taman Prestasi.. 50
Tabel 4.2 dan pembagian data Training dan testing ..... 52
Tabel 4.3 Kombinasi Pengujian berdasar jumlah neuron dan pembagian data Training dan testing..... 53

*Halaman ini sengaja dikosongkan*

# BAB 1
# PENDAHULUAN

## 1.1 Latar Belakang

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. Hal ini disebabkan oleh adanya polusi udara. Polusi atau pencemaran lingkungan adalah masuknya makhluk hidup, zat energi, dan atau komponen lain ke dalam lingkungan baik sengaja atau tidak, atau berubahnya tatanan lingkungan oleh kegiatan manusia atau oleh proses alam sehingga kualitas lingkungan turun sampai ke tingkat tertentu yang menyebabkan lingkungan menjadi kurang atau tidak dapat berfungsi lagi sesuai dengan peruntukannya[1]. 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. 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]. Salah satu penyebab polusi udara adalah dari sektor transportasi.

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 sebesar153%. Jumlah sepeda motor mendominasi dengan 82% di tahun 2017 [4]. Di Kota Surabaya menurut data statistik sektoral  Polantas Kota Besar Surabaya yang dikeluarkan oleh Badan Pusat Statistik Kota Surabaya pada tanggal 11 Januari 2018, jumlah kendaraan bermotor di kota Surabaya yang terdiri dari jenis sedan, jeep, stwagoon, bus, truk sepeda motor dan alat berat di kota Surabaya sejak tahun 2009 hingga 2015 mengalami kenaikan rata-rata 6,2% tiap tahun. Pada tahun 2015 jumlah kendaraan bermotor berjumlah 2.126.168 dengan rincian jumlah sepeda motor di kota Surabaya mendominasi sebesar 78% , disusul dengan kendaraan roda empat pribadi seperti sedan, jeep, MPV dan sejenisnya sebesar 16%. Sedangkan sisanya seperti bus, truk dan alat berat dan sejenisnya berjumlah 6%[5].

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 (NOx), SO2 (Sulfur Oksida), senyawa hidro karbon (HC) dan partikulat debu termasuk timbel (PB) sebagai akibat proses pembakaran mesin yang tidak sempurna serta partikel lepas [6]. 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 [7]. Kota pintar tercipta dengan tujuan untuk meningkatkan kenyamanan penghuninya melalui fasilitas mobilitas [8], efisiensi dan hemat energi [9], perbaikan kualitas air dan udara [10], identifikasi masalah dan pemberian solusi cepat, tanggap dalam pemulihan setelah bencana [11], dan integrasi data untuk pengambilan keputusan melalui pemanfaatan *cloud computing* dalam mengolah *big data* [12].

*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 pemanfatannya. 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 [13]. 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, NOx, NO2, SO2 dan beberapa parameter data meteorologi. Data hasil pemantauan tersebut disimpan dalam bentuk *file* dan *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 memanfaatkan implementasi 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 [14]. Peramalan dapat dilakukan dengan pendekatan kualitatif maupun kuantitatif [15]. Tujuan peramalan adalah memprediksi prospek ekonomi dan kegiatan usaha serta pengaruh lingkungan terhadap prospek tersebut [16]. *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 input RNN adalah data yang dijadikan masukan, yaitu 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) [17][18][19], pengenalan suara [20], sintesa musik [21], manajemen keuangan [22], analisis deret DNA [23], analisis video [24], dan sebagainya. Seiring berkembangnya implementasi *Recurrent Neural Network*, dikembangkan pula metode *Long Short Term Memory* (LSTM). Perbedaaan 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 metereorologi, dan analisis prediksi menggunakan metode peramalan *Recurrent Neural Network- Long Short Term Memory* (RNN-LSTM) terhadap partikular Karbon Monoksida (CO) sebagai salah satu partikulat utama penyebab polusi udara di bidang transportasi. Dari penelitian ini diharapkan para pemangku kepentingan khususnya di bidang pengendalian polusi lingkungan dapat menggunakan data prediksi sebagai rekomendasi mengambil keputusan dalam mengendalikan polusi udara untuk tahun-tahun yang akan datang.

## 1.2 Rumusan Masalah

Selama ini Pemerintah Kota Surabaya telah melakukan pemantauan polusi udara. Salah satu parameter yang dipantau adalah Karbon monoksida (CO) sebagai salah satu penyebab polusi udara dan data meteorologi yang mempengarahui keadaan polusi udara. Polusi udara yang melampui nilai ambang batas aman yang telah ditentukan dapat menyebabkan resiko yang serius bagi keberlangsungan hidup makhluk di bumi termasuk manusia sehingga perlu diupayakan pencegahan. Untuk menentukan pencegahan polusi udara yang tepat, maka perlu dilakukan prediksi polusi udara pada tahun-tahun yang datang. Untuk itu diperlukan sebuah metode untuk menganalisis data runtun waktu polusi udara di kota Surabaya. Analisis yang dibuat adalah membuat prediksi polusi udara dari parameter Karbon monoksida, analisis trend dan analisis korelasi antara polutan Karbon monoksida dengan data meteorologi.

## 1.3 Tujuan

Perkembangan artificial intelligence saat ini menjadi primadona sebagai *tools* untuk membantu manusia menyelesaikan masalah. Salah satu yang lagi trend adalah penggunaan *Artificial Neural Network*. RNN-LSTM adalah salah satu pengembangan ANN untuk pengolahan data skuensial yaitu data yang memiliki runtun waktu yang dapat digunakan untuk membuat prediksi/peramalan. Hal ini menjadi landasan untuk membuat analisa data runtun waktu untuk memprediksi Karbon Monoksida di Kota Surabaya menggunakan *Deep Learning Recurrent*

*Neural Network* LSTM serta menganalisis trend dan korelasi polutan Karbon Monoksida dengan parameter meteorologi yang dominan terhadap tingkat polusi udara.

## 1.4 Batasan Masalah

Batasan masalah yang diambil dari penelitian ini adalah sebagai berikut:

a. Sumber data berasal dari Dinas Lingkungan Hidup Kota Surabaya tahun 14 Mei 2015 sampai dengan 04 Agustus 2019

b. Titik lokasi pemantauan udara yang digunakan adalah Taman Prestasi

c. Parameter udara yang digunakan yaitu CO (*Carbon monoxide* - Karbon monoksida) dan data meteorologi yang terdiri dari : 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), HUMair (*Humidity air* - Kelembaban udara), TEMPair (*Temperature air* - Temperatur udara)

d. Nilai kualitas udara yang dipergunakan yaitu Indek Standar Pencemar Udara (ISPU) sebagai standar perhitungan kualitas udara di Indonesia, hal ini sesuai dengan Keputusan Menteri Negara Lingkungan Hidup Nomor : KEP 45 / MENLH / 1997 Tentang Indeks Standar Pencemar Udara

## 1.5 Kontribusi

Penelitian ini diharapkan dapat memberikan suatu kontribusi keilmuan pada masyarakat dan atau kepada pemerintah kota Surabaya terkait prediksi polusi udara yang mungkin terjadi pada tahun tahun mendatang agar dapat menjadi acuan untuk menerapkan suatu kebijakan terkait penanganan pencegahan polusi tersebut. Selain itu, hasil prediksi dari penelitian ini diharapkan dapat menghasilkan model matematis untuk polusi udara menggunakan *Articial Intelligence* yaitu *Recurrent Neural Network – Long Short Term Memory* yang bisa diterapkan di semua wilayah Indonesia sebagai sistem prediksi serta sebagai acuan penerapan strategi yang akan dilakukan pada wilayah tersebut.

# BAB 2
# KAJIAN PUSTAKA

Pada bagian ini membahas penelitian-penelitian yang telah dilakukan oleh peneliti lain, teori dasar tentang polusi udara dan ISPU, time series, interpolasi data, peramalan, neural network, algoritma RNN-LSTM.

## 2.1 Kajian Penelitian Terkait

Ada beberapa penelitian yang 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 [25], *Predicting Trends in Air Pollution in Delhi using Data Mining* [26], *Time Series Based LSTM Model to Predict Air Pollutant's Concentration for Prominent Cities in India* [27], *Deep Air: Forecasting Air Pollution in Beijing, China* [28].

Pada artikel yang berjudul Analisis dan Implementasi *Long Short Term Memory Neural Network* untuk Prediksi Harga Bitcoin [25], dijelaskan bahwa dalam 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 testing 93,5%.

Pada penelitian yang berjudul *Predicting Trends in Air Pollution in Delhi using Data Mining* [26], penelitian ini 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 kendaraaan 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. Polutan lain seperti SO2 juga menunjukkan penurunan.

Pada penelitian yang *berjudul Time Series Based LSTM Model to Predict Air Pollutant's Concentration for Prominent Cities in India* [27], 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 6 Mei 2018. Hasil penelitian ini menyebutkan bahwa nilai RMSE untuk 2 lokasi sensor bernilai1<5 untuk prediksi 12 jam berikutnya, nilai <8 untuk prediksi 6 jam berikutnya, dan nilai < 5 untuk jam berikutnya.

Pada penelitian yang berjudul *Deep Air: Forecasting Air Pollution in Beijing, China* [28], penelitian ini mengisvestigasi 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.

## 2.2 Teori Dasar

### 2.2.1 Polusi Udara

Pencemaran udara merupakan terjadinya perubahan komposisi pada udara dibandingkan dengan keadaan normal. Perubahan tersebut akan membahayakan kehidupan dan kesehatan masyarakat. Menurut PP No. 41 tahun 1999, pencemaran udara adalah masuknya atau dimasukkannya zat, energi, dan/atau komponen lain ke dalam udara ambien oleh kegiatan manusia, sehingga mutu udara ambien turun sampai ke tingkat tertentu yang menyebabkan udara ambien tidak dapat memenuhi fungsinya. Dengan berkembangnya tren otomotif, kendaraan bermotor menjadi

salah satu penyumbang terjadinya polusi udara. Sisa hasil proses pembakaran pada kendaraan bermotor yang menggunakan energi fosil merupakan biang keladi dari banyak kota di dunia yang mengalami polusi udara. Karbon monoksida yang terdapat di alam terbentuk dari proses pembakaran tidak lengkap terhadap karbon atau komponen yang mengandung karbon. Hidrokarbon yang diproduksi oleh manusia yang terbanyak berasa dari transportasi, selain itu dari pembakaran gas, minyak, arang dan kayu, proses-proses industri, pembuangan sampah, kebakaran hutan dan lain sebagainya. Karbon monoksida bisa berbahaya bagi manusia karena reaksi antara CO dengan hemoglobin di dalam darah akan membawanya dari paru-paru dan sel-sel tubuh. Sementara itu, Hidrokarbon dapat menyebabkan iritasi, lemas, pusing, kehilangan koordinasi, dan mengakibatkan kematian[29].

*2.2.1.1 ISPU*

Saat ini Indeks standar kualitas udara yang dipergunakan secara resmi di Indonesia adalah Indek Standar Pencemar Udara (ISPU), hal ini sesuai dengan Keputusan Menteri Negara Lingkungan Hidup Nomor : KEP 45 / MENLH / 1997 Tentang Indeks Standar Pencemar Udara. Dalam keputusan tersebut yang dipergunakan sebagai bahan pertimbangan diantaranya : bahwa untuk memberikan kemudahan dari keseragaman informasi kualitas udara ambien kepada masyarakat di lokasi dan waktu tertentu serta sebagai bahan pertimbangan dalam melakukan upaya-upaya pengendalian pencemaran udara perlu disusun Indeks Standar Pencemar Udara. Dalam [30] 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 Partikulat Matter (PM10), Karbon monoksida (CO), Sulfur dioksida (SO2), Nitrogen dioksida (NO2) dan Ozon (O3). Menurut [31], penentuan ISPU dihitung dengan persamaaan (1) dengan mengacu pada Batas Indeks Standar Pencemar Udara Dalam Satuan SI seperti pada Tabel 2.1.

$$I = \frac{Ia - Ib}{Xa - Xb}(Xx - Xb) + Ib \quad \text{.....(2.1)}$$

Dimana:

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

*Tabel 2.1 Batas Indeks Standar Pencemar Udara Dalam Satuan S*I

| **Indeks Standar Pencemar Udara** | **PM10 (µg/m$^3$)** | **SO2 (µg/m$^3$)** | **CO (µg/m$^3$)** | **O3 (µg/m$^3$)** | **NO2 (µg/m$^3$)** |
|---|---|---|---|---|---|
| 0 | 0 | | | 0 | 0 |
| 50 | 50 | 0 | 0 | 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 |

Indeks Standar Pencemar Udara ditetapkan dengan cara mengubah kadar pencemar udara yang terukur menjadi suatu angka yang tidak berdimensi. Rentang Indeks Standar Pencemar Udara dapat dilihat pada tabel 2.2.

*Tabel 2.2 Rentang Indeks Standar Pencemar Udara*

| KATEGORI | RENTANG | PENJELASAN |
|---|---|---|
| Baik | 0 - 50 | Tingkat kualitas udara yang tidak memberikan efek bagi kesehatan manusia atau hewan dan tidak berpengaruh pada tumbuhan, bangunan atau nilai estetika |
| Sedang | 51 – 100 | Tingkat kualitas udara yang tidak berpengaruh pada kesehatan manusia ataupun hewan tetepi berpengaruh pada tumbuhan yang sensitif dan nilai estetika |
| Tidak Sehat | 101 – 199 | Tingkat kualitas udara yang bersifat merugikan pada manusia ataupun kelompok hewat yang sensitive atau bisa menimbulkan kerusakan tumbuhan atau nilai estetika |
| Sangat Tidak Sehat | 200 – 299 | Tingkat kualitas udara yang dapat merugikan kesehatan pada sejumlah segmen populasi yang terpapar |
| Berbahaya | 300 – lebih | Tingkat kualitas udara yang berbahaya secara umum dapat merugikan kesehatan secara serius |

*2.2.1.2 Karbon monoksida (CO)*

Karbon monoksida, dengan rumus kimia CO. adalah gas yang tidak berwarna, tidak berbau, dan tidak berasa. Karbon monoksida terdiri dari satu atom karbon yang berikatan secara kovalen dengan satu atom oksigen. Karbon monoksida dihasilkan dari pembakaran yang tidak sempurna dari senyawa karbon. Karbon monoksida terbentuk apabila terdapat kekurangan oksigen dalam pembakaran. Karbon monoksida mudah terbakar.

Karbon monoksida dihasilkan dari pembakaran tak sempurna dari senyawa karbon, sering terjadi pada mesin pembakaran dalam. Karbon monoksida terbentuk apabila terdapat kekurangan oksigen dalam proses pembakaran. Karbon monoksida mudah terbakar dan menghasilkan lidah api berwarna biru, menghasilkan karbon dioksida. Walaupun ia bersifat racun, CO memainkan peran yang penting dalam teknologi modern, yakni merupakan prekursor banyak senyawa karbon.

Gas produser dibentuk dari pembakaran karbon di oksigen pada temperatur tinggi ketika terdapat karbon yang berlebih. Dalam sebuah oven, udara dialirkan melalui kokas. $CO_2$ yang pertama kali dihasilkan akan mengalami kesetimbangan dengan karbon panas, menghasilkan CO. Reaksi $O_2$ dengan karbon membentuk CO disebut sebagai kesetimbangan Boudouard. Di atas 800 °C, CO adalah produk yang predominan.

Gas karbon monoksida (CO) adalah gas yang dihasilkan dari proses oksidasi bahan bakar yang tidak sempurna. Gas ini bersifat tidak berwarna, tidak berbau, tidak menyebabkan iritasi, dan tidak berasa. CO diproduksi dari pembakaran bakan bakar fosil yang tidak sempurna, seperti bensin, minyak dan kayu bakar. Konsentrasi CO dapat meningkat di sepanjang jalan raya yang padat lalu lintas dan menyebabkan pencemaran. Asap kendaraan merupakan sumber karbon monoksida yang dikeluarkan di banyak daerah perkotaan.

Karbon monoksida (CO) kadang muncul sebagai parameter di lokasi pemantauan di kota-kota besar dengan kepadatan lalu lintas yang tinggi seperti Jakarta, Bandung dan Surabaya, tetapi pada umumnya konsentrasi CO berada di bawah ambang batas Baku Mutu PP41/1999 (10,000µg/m3/24 jam). Gas karbon monoksida memasuki tubuh melalui pernafasan dan diabsorpsi di dalam peredaran darah. Karbon monoksida akan berikatan dengan haemoglobin menjadi carboxyhaemoglobin. Gas CO mempunyai kemampuan berikatan dengan haemoglobin sebesar 240 kali lipat kemampuannya berikatan dengan O2 sehingga pasokan O2 ke seluruh tubuh menurun tajam, sehingga melemahkan kontraksi jantung dan menurunkan volume darah yang didistribusikan.

Karbon monoksida yang dihasilkan dari proses pembakaran yang tidak sempurna dapat menyebabkan pencemaran lingkungan. Gas karbon monoksida dapat berpindah dari satu tempat ke tempat lain denagn angin, sehingga daerah tersebut dapat tercemari gas CO. Dengan meningkatnya gas CO, oksigen di tempat tersebut akan berkurang. Pada proses pembakaran mesin, pencampuran yang tidak rata antara minyak bakar dengan udara dapat menghasilkan beberapa tempat yang kekurangan oksigen. Semakin rendah perbandingan antara udara dan minyak bakar, semakin tinggi jumlah karbon monoksida yang dihasilkan. Sehingga dari emisi

kendaraan dan industri memberikan kontribusi pada efek rumah kaca dan pemanasan global.

### 2.2.2 Data Meteorologi

Meteorologi adalah ilmu yang mempelajari maslah atmosfer seperti udara, suhu, cuaca, angin, dan dan berbagai sifat fisika serta kimia atmosfer lainnya. Dalam kamus besar bahasa Indonesia (KBBI), meteorologi di defenisikan sebagai cabang ilmu geografi yang mempelajari bagaimana ciri-ciri fisik dan kimia atmosfer untuk meramalkan cuaca di suatu tempat secara khusus serta memperkirakan cuaca dunia secara umum[32]. Ilmu meteorologi sangat sangat bergantung pada kegiatan yang disebut sebagai observasi. Pengamatan dilakukan agar mendapatkan data pada parameter yang berpengaruh pada perubahan cuaca yang kemudian akan di analisis sehingga menghasilkan perkiraan cuaca yang bermanfaat untuk di terapkan di segala bidang kehidupan. Parameter meteorologi yang diamati yakni :

*2.2.2.1 DD (Wind direction - Arah angin)*

Arah angin adalah arah darimana angin berhembus atau darimana arus angin datang dan dinyatakan dalam derajat yang ditentukan dengan arah perputaran jarum jam dan dimulai dari titik utara bumi dengan kata lain sesuai dengan titik kompas. Umumnya arus angin diberi nama dengan arah darimana angin tersebut bertiup, misalnya angin yang berhembus dari utara maka angin utara[33].

*2.2.2.2 DD_Boe (Wind gust direction - Arah hembusan angin)*

Arah hembusan angin adalah peningkatan singkat dalam arah angin, biasanya kurang dari 20 detik[34]. Satuan yang digunakan adalah *degrees.*

*2.2.2.3 FF (Wind velocity - Kecepatan angin)*

Kecepatan angin adalah kecepatan dari menjalarnya arus angin dan dinyatakan dalam knots atau kilometer per jam maupun dalam meter per detik[33]

*2.2.2.4 FF_Boe (Wind gust velocity - Kecepatan hembusan angin)*

Kecepatan hembusan angina adalah peningkatan singkat dalam kecepatan angin[34]. Satuan yang digunakan dalam Sistem Pemantauan Kualitas Udara di Kota Surabaya adalah m/s.

*2.2.2.5 GRAD (Global radiation - Radiasi global)*

Radiasi global adalah radiasi surya yang datang dari piringan matahari dengan sudut utuh yang tidak terhambur dan tidak terpantul oleh atmosfer atau benda-benda lain di dalam atmosfer[35]. Satuan yang digunakan dalam laporan harian Sistem Pemantauan Kualitas Udara di Kota Surabaya adalah $W/m^2$, artinya daya per satuan luas.

*2.2.2.6 HUMair (Humidity air - Kelembaban udara)*

Kelembapan udara adalah banyaknya uap air yang terkandung dalam udara atau atmosfer. Besarnya tergantung dari masuknya uap air ke dalam atmosfer karena adanya penguapan dari air yang ada di lautan, danau, dan sungai, maupun dari air tanah. Disamping itu terjadi pula dari proses transpirasi, yaitu penguapan dari tumbuhtumbuhan. Sedangkan banyaknya air di dalam udara bergantung kepada banyak faktor, antara lain adalah ketersediaan air, sumber uap, suhu udara, tekanan udara, dan angin[36]. Satuan yang digunakan sesuai laporan harian dari Sistem Pemantauan Kualitas Udara di Kota Surabaya adalah prosentase.

*2.2.2.7 TEMPair (Temperature air - Temperatur udara)*

Temperatur udara adalah nilai suhu udara pada suatu wilayah. Satuan yang digunakan adalah *degrees Celcius*

### 2.2.3 Time Series

*Time series* atau data runtun waktu adalah data yang disusun berdasarkan urutan waktu atau data yang dikumpulkan dari waktu ke waktu. Waktu yang digunakan bisa berupa minggu, bulan, tahun, dan sebagainya. Dengan demikian data time series berhubungan dengan data statistik yang dicatat dan diselidiki dalam batas – batas (interval) waktu tertentu, seperti penjualan, harga, persediaan produksi, dan tenaga kerja[37]. Berdasarkan bentuk pola yang dibentuk, data *time series* dapat dibagi menjadi empat yaitu: *horizontal*, *trend*, musiman dan siklis[38].

*2.2.3.1 Pola Horizontal*

Pola horizontal merupakan kejadian yang tidak terduga dan bersifat acak, tetapi kemunculannya dapat memepengaruhi fluktuasi data time series

*2.2.3.2 Pola Trend*

Trend adalah plot data menyebar membentuk suatu kecenderungan menaik ataupun menurun secara konstan. Meskipun dalam realitanya data yang konstan menaik ataupun menurun jarang sekali ditemukan, namun data yang telah menaik atau menurun dalam jangka waktu yang cukup panjang sudah dapat dikatakan sebagai data yang memuat pola trend.

*2.2.3.3 Pola Musiman*

Data ini merupakan data berpola trend yang dibentuk oleh faktor musiman dan cenderung berulang namun teratur. Pola data musiman dapat mempunyai pola musim yang berulang dari periode ke periode berikutnya. Misalnya pola yang berulang setiap bulan tertentu, tahun tertentu, atau minggu tertentu.

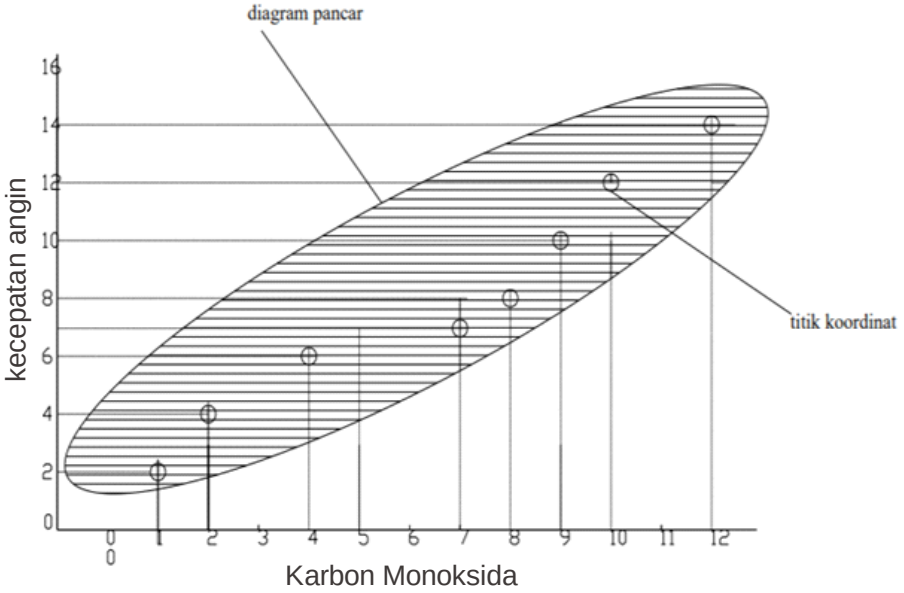
*2.2.3.4 Siklis*

Pola siklus berbentuk trend yang memuat musiman namun periode musimannya lebih panjang.

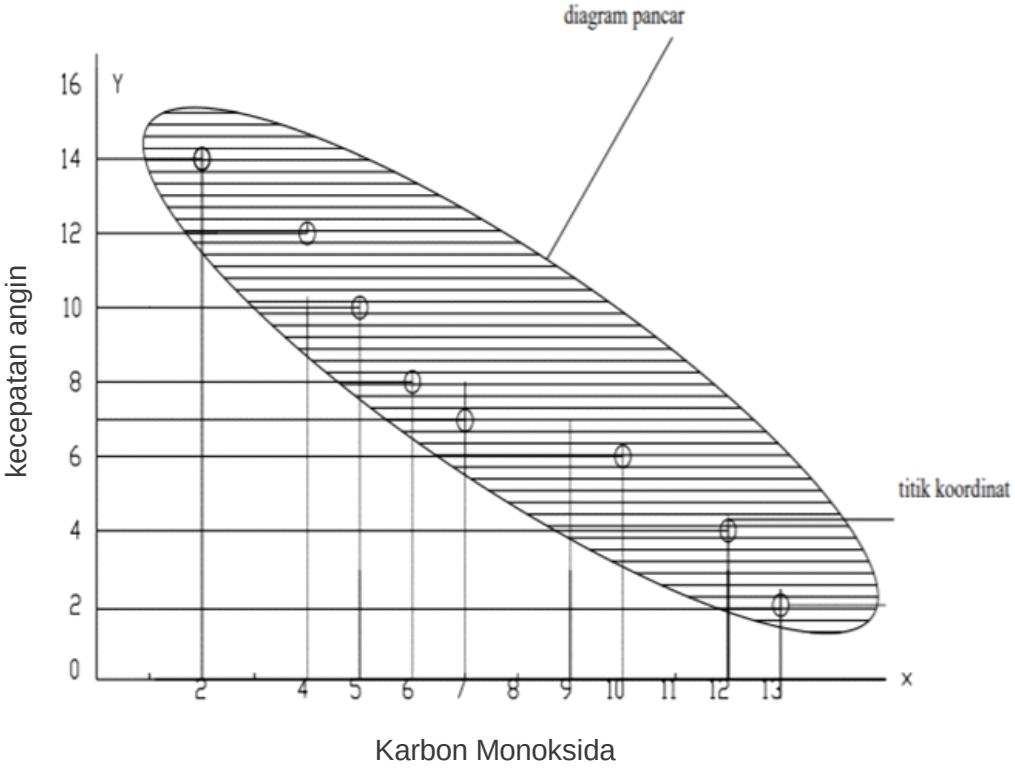
### 2.2.4 Analisis Korelasi

Analisis korelasi adalah metode statistika yang digunakan untuk menentukan kuatnya atau derajat hubungan linier antara dua variabel atau lebih. Semakin nyata hubungan linier (garis lurus), maka semakin kuat tinggi hubungan garis lurus antara kedua variabel atau lebih. Hubungan garis lurus ini dinamakan koefisien korelasi. Korelasi menyatakan hubungan antara dua variabel tanpa memperhatikan variabel mana yang menjadi perubah. Karena itu hubugan korelasi belum dapat dikatakan sebagai hubungan sebab akibat[39].

Hubungan dua variabel ada yang positif dan negatif. Misalkan data polusi udara Karbon Monoksida dinisialisasikan dengan X dan data metereorologi kecepatan angin diinisialisasikan dengan Y. Dari dua variable tersebut akan dicari hubungan korelasinya. Hubungan X dan Y dikatakan positif apabila kenaikan (penurunan) X pada umumnya diikuti oleh kenaikan (penurunan) Y. Sebaliknya dikatakan negatif bila kenaikan (penurunan) X, pada umumnya diikuti oleh penurunan (kenaikan) Y. Hubungan positif dan negatif dapat dilihat pada gambar. 2.1 dan 2.2

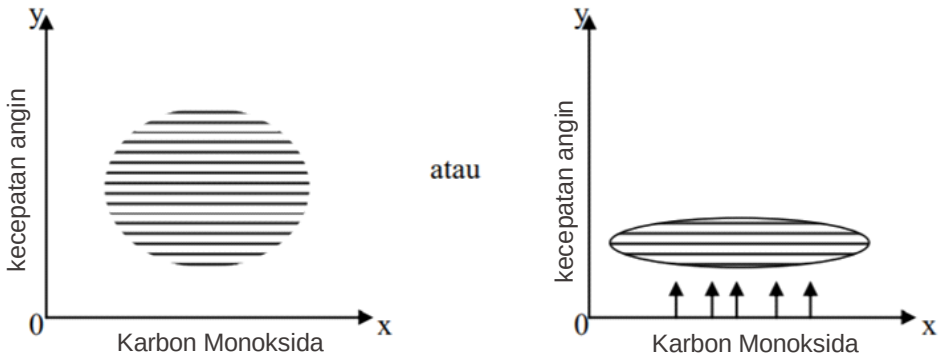



Gambar 2.1 X dan Y mempunyai hubungan positif




Gambar 2.2 X dan Y mempunyai hubungan negative

Seperti yang terlihat pada gambar 2.1 dan 2.2 di atas apabila antara variable X dan Y terdapat hubungan, maka bentuk diagram pancarnya adalah mulus/teratur, dimana menunjukan gerakan diagram pancar dari kiri bawah ke kanan atas ( hubungan positif), sedangkan bila gerakan diagram bergerak dari kiri atas ke kanan bawah (hubungan negatif). Apabila bentuk diagram pancarnya tidak teratur, artinya kenaikan/penuruan X tidak diikuti oleh naik turunnya Y, maka dapat dikatakan X dan Y tidak berkorelasi. Dengan kata lain, jika naik turunnya variabel X tidak mempengaruhi Y dikatakan X dan Y tidak ada hubungan atau hubungan lemah. Hubungan X dan Y tidak mempunyai hubungan dapat dilihat pada gambar 2.3.




Gambar 2.3 X dan Y mempunyai hubungan kuat atau hubungan lemah sekali

Kuat tidaknya hubungan antara X dan Y apabila dapat dinyatakan dengan fungi linier, diukur dengan suatu nilai yang disebut koefisien korelasi. Nilai koefisien korelasi ini paling sedikit -1 dan paling besar 1. Jadi jika r = Koefisiensi korelasi, maka nilai r dinyatakan $-1 \leq r \leq$, dimana jika r = 1 hubungan X dan Y sempurna dan positif (mendekati 1, yaitu hubungan sangat kuat dan positif), sedangkan jika nilai r = -1 hubungan X dan Y sempurna dan negatif (mendekati -1, yaitu hubungan sangat kuat dan negatif). Jika hubungan X dan Y = 0 dapat dinyatakan hubungan lemah sekali dan tidak ada hubungan. Rumus korelasi sederhana yang menggambarkan koefisien korelasi adalah menggunakan persamaan 2.1

$$r_{xy} = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\{n\sum x^2 - (\sum x)^2\}\{n\sum y^2 - (\sum y)^2\}}} \quad \text{.....(2.2)}$$

dimana:

x,y = variable data

r = koefisien korelasi

n = jumlah data

### 2.2.5 Interpolasi Data

Pada saat pengumpulan data, data sering kali tidak tersaji secara lengkap. Seringkali terdapat nilai data yang hilang (*missing value*). Terdapat banyak penyebab dari kondisi tersebut, baik akibat kesalahan manusianya maupun keterbatasan kemampuan alat ukur. Terdapat banyak cara untuk menangani kondisi-kondisi tersebut. Sejumlah peneliti memilih untuk menghapus data tersebut. Namun jika data berupa data time series, hal ini akan mempengaruhi data tersebut. Salah satu cara yang dapat dilakukan untuk mengisi data hilang adalah dengan melakukan interpolasi terhadap data.

Interpolasi adalah perkiraan suatu nilai tengah dari suatu set nilai yang diketahui. Interpolasi dengan pengertian yang lebih luas merupakan upaya mendefinisikan suatu fungsi pendekatan suatu fungsi analitik yang tidak diketahui atau pengganti fungsi rumit yang tak mungkin diperoleh persamaan analitiknya[40]. Dapat pula diartikan bahwa interpolasi merupakan proses pencarian dan perhitungan nilai suatu fungsi yang grafiknya melewati suatu kumpulan titik yang diberikan. Titik-titik tersebut mungkin merupakan hasil eksperimen dari suatu percobaan ataupun dari suatu fungsi yang telah diketahui. Fungsi interpolasi dapat dipakai dari sekelompok fungsi tertentu, diantaranya adalah fungsi polinomial. Jika yang digunakan untuk melakukan interpolasi adalah fungsi polinomial maka polinomial tersebut dinamakan polinomial interpolasi [41]. Polinomial interpolasi bergantung pada nilai-nilai dan banyaknya nilai x dan y yang diberikan.

Interpolasi polinomial sering memberikan hasil yang kurang optimal, karena jika data yang diberikan banyak maka derajat polinomial yang terbentuk akan sangat tinggi, hal tersebut juga menyulitkan dalam perhitungan. Polinomial dengan derajat tinggi akan menghasilkan fluktuasi data yang sangat cepat.

Perubahan data pada interval kecil dapat menyebabkan perubahan besar pada keseluruhan interval. Maka interpolasi biasanya menggunakan polinomial berderajat rendah. Dengan melakukan pembatasan derajat maka dapat ditentukan alternatif lain untuk mendapatkan kurva mulus melalui sejumlah titik, yakni membagi suatu interval yang memuat data titik menjadi beberapa subinterval dan pada setiap subinterval disusun polinomial interpolasi. Hasilnya adalah sebuah kurva yang terdiri atas potongan-potongan kurva polinomial yang berderajat sama, fungsi tersebut disebut dengan fungsi spline. Pada persamaan 2.3 digambarkan sebuah fungsi spline S yang berderajat n.

$$S(x) = \begin{cases} S_1(x) & untuk\ x_1 \leq x \ \leq x_2 \\ S_2(x) & untuk\ x_2 \leq x \ \leq x_2 \\ \vdots & \\ S_{n-1}(x) & untuk\ x_{n-1} \leq x \ \leq x_n \end{cases} \quad \text{.....} (2.3)$$

dengan :

$S_k(x) = a_k x^n + \ a_{k-1}x^{n-1} + \cdots + \ a_2x^2 + a_1x + a_0$ ..... (2.4)

$n = 1, 2\ ,3, \ldots$ ..... (2.5)

$k = 1, 2, 3, \ldots (n-1)$ ..... (2.6)

dimana:

$a_k$ = konstanta

Suatu fungsi bisa dimasukkan dalam fungsi spline adalah jika suatu fungsi yang terdiri atas beberapa bagian fungsi polinomial yang di rangkaikan bersama dengan beberapa syarat kemulusan[42]. Sebuah fungsi *spline* S berderajat n pada interval [*a*,*b*] pada persamaan 2.3, memenuhi syarat-syarat :

a. *S* terdefinisi pada [*a*,*b*]
b. $S, S', S'', S''', \ldots, S^{(n+1)}$ kontinu pada interval [a,b]
c. Terdapat titik-titik $x_k$ simpul-simpul spline *S(x)* sedemikian sehingga $a=x_1 < x_2 < \ldots < x_n = b$ dan *S(x)* merupakan suatu polynomial berderajat $\leq$ n pada setiap subinterval $[x_k , x_{k+1}]$

### 2.2.6 Peramalan (*Forcasting*)

Untuk menyelesaikan masalah di masa datang yang tidak dapat dipastikan, orang senantiasa berupaya menyelesaikannya dengan model pendekatan-pendekatan yang sesuai dengan perilaku aktual data, begitu juga dalam melakukan peramalan. Peramalan (*forecasting*) permintaan akan produk dan jasa di waktu mendatang dan bagian-bagiannya adalah sangat penting dalam perencanaan dan pengawasan produksi. Suatu peramalan banyak mempunyai arti, maka peramalan tersebut perlu direncanakan dan dijadwalkan sehingga akan diperlukan suatu periode waktu paling sedikit dalam periode waktu yang dibutuhkan untuk membuat suatu kebijaksanaan dan menetapkan beberapa hal yang mempengaruhi kebijaksanaan tersebut. Peramalan adalah perhitungan yang objektif dan dengan menggunakan data-data masa lalu, untuk menentukan sesuatu di masa yang akan datang[43].

### 2.2.7 Neural Network

*Neural network* (Jaringan saraf tiruan) adalah pemrosesan dari kumpulan elemen-elemen, unit atau yang saling berhubungan, yang fungsinya secara longgar didasarkan pada fungsi kerja otak. Kemampuan pemrosesan jaringan disimpan dalam kekuatan koneksi antar unit, atau bobot, yang diperoleh dengan proses adaptasi, atau belajar dari, serangkaian pola pelatihan[44].

Jaringan syaraf tiruan terdiri dari beberapa komponen berikut:

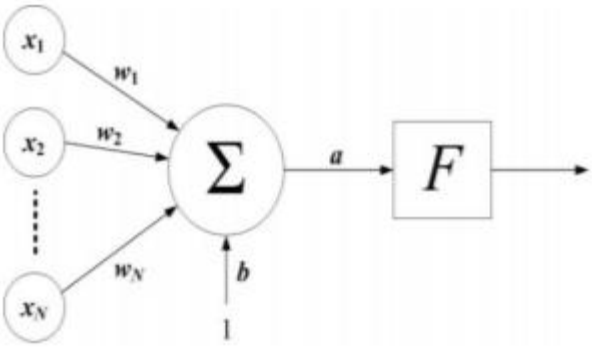
a. Neuron yaitu sel syaraf yang akan mentransformasikan informasi yang diterima melalui sambungan keluarnya menuju neuron - neuron lain;

b. Bobot yaitu hubungan antar neuron;

c. *Neuron Layers* yaitu tempat dimana neuron - neuron akan dikumpulkan dalam lapisan - lapisan (layer).

Pada jaringan syaraf tiruan, neuron - neuron akan dikumpulkan dalam lapisan - lapisan (layer) yang disebut dengan lapisan neuron (neuron layers). Neuron - neuron pada satu lapisan akan dihubungkan dengan lapisan – lapisan sebelum dan sesudahnya. Informasi atau masukan yang diberikan pada jaringan syaraf akan dirambatkan dari lapisan ke lapisan, mulai dari lapisan masukan sampai

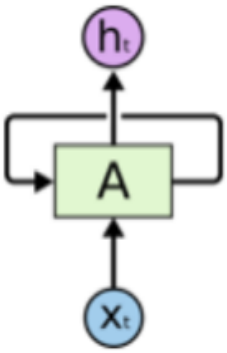
ke lapisan keluaran melalui lapisan yang lainnya yang disebut dengan lapisan tersembunyi (*hidden layer*). Gambar 2.4 merupakan konsepsi sederhana neural network dengna fungsi aktivasi F. *Neural Network* ditentukan oleh 3 hal yang sangat penting yaitu : Pola hubungan antar neuron, metode untuk menentukan bobot penghubung (*training learning*) dan fungsi aktivasi.




Gambar 2.4 Fungsi aktivasi pada neural network

*2.2.7.1 Recurrent Neural Network*

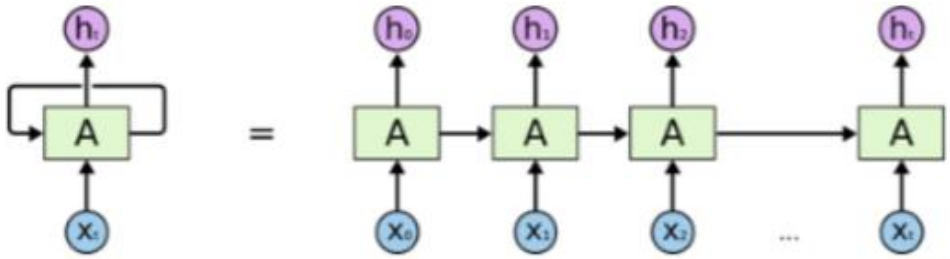
*Recurrent Neural Network* atau RNN merupakan bagian dari jaringan syaraf tiruan untuk pemrosesan data sekuensial. Berbeda dengan *Convolutional Neural Network* (CNN) yang merupakan metode untuk pemrosesan data seperti gambar, RNN sering digunakan untuk memproses nilai yang bersifat sekuensial $x_1$ , ..., ..., ..., $x_n$.




Gambar 2.5 Bentuk konseptual paling sederhana RNN

Pada gambar 2.5, $X_t$ merupakan masukan dan $h_t$ merupakan keluaran. Pengulangan ini membuat informasi dapat dilewatkan dari satu langkah pada

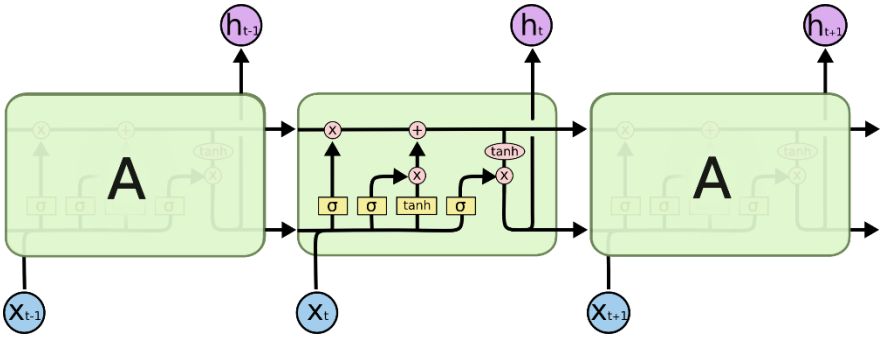
jaringan ke selanjutnya. Pengulangan ini membuat RNN dapat dijabarkan sebagai banyak salinan dari jaringan yang sama seperti digambarkan pada Gambar 2.5. Perulangan RNN sebenarnya akan memproses input dari skala waktu 0 sampai t. Namun jaringan RNN tidak dapat menghubungkan pembelajaran informasi yang lama karena memori lama yang tersimpan akan tergantikan dengan memori baru seiring waktu berjalan. Kekurangan ini telah ditemukan oleh [45], sehingga dikembangkan *RNN- Long Short Term Memory*.




Gambar 2.6 Struktur RNN yang dijabarkan

*2.2.7.2 Long Short-Term Memory*

Long Short Term Memory atau biasa disingkat LSTM adalah bentuk spesial dari RNN yang dapat melakukan pembelajaran pada dependensi jangka panjang (*long-term dependencies*). Model ini diperkenalkan oleh Hochreiter dan Schmidhuber pada tahun 1997. Semua *Recurrent Neural Network* memiliki bentuk rangkaian modul jaringan syaraf yang berulang. LSTM juga memiliki struktur yang sama namun memiliki tambahan fitur berupa gerbang pada sel. Dalam jaringan LSTM dikembangkan sel memori (*memory cells*) dan unit gerbang (*gate units*) untuk mengatur memori pada setiap masukannya. Cara kerja *Long Short Term Memory* (LSTM) adalah membuat perubahan pada RNN dengan menambahkan *memory cell*s 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[45].




Gambar 2.7 Detail Struktur LSTM

LSTM akan menentukan informasi apa yang akan dibuang dari sel. Keputusan ini dibuat oleh *forget gate layer*. Layer ini akan memperhatikan $h_{t-1}$ dan $x_t$ sehingga akan menghasilkan keluaran antara 0 dan 1. Keluaran 0 merepresentasikan bahwa informasi akan dilupakan sedangkan keluaran 1 merepresentasikan bahwa informasi tidak akan dilupakan.

$$ft = \sigma(W_f.[h_{t-1}, x_t] + b_f \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.7)$$

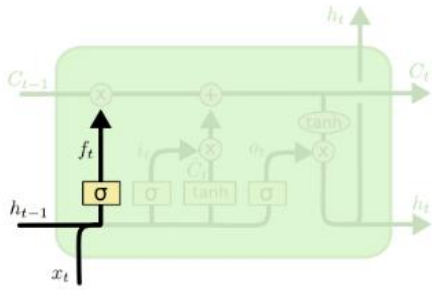
dimana:

$f_t$ = fungsi aktivasi *forget layer* pada *time step t*

$W_f$ = bobot pada fungsi

$h_{t-1}$ = satu hidden state sebelumnya dari *time step t*

$x_t$ = vektor *input*




Gambar 2.8 Struktur Forget Layer

Langkah selanjutnya adalah menentukan apakah informasi akan disimpan pada sel. Pertama, sebuah layer sigmoid bernama "*input gate layer*" menentukan

nilai mana yang akan diperbaharui. Selanjutnya, sebuah layer *tanh* membuat vektor dari nilai kandidat baru, *Ct*, yang dapat ditambah ke *state*. Langkah selanjutnya, kedua layer ini akan dikombinasikan untuk memperharui *state*.

$$i_t = \sigma(W_i \,.\, [h_{t-1} + X_t] + b_i) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.8)$$

$$C_t = tan\, h(W_c . [h_{t-1}, X_t] + b_c) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2.9)$$

dimana:

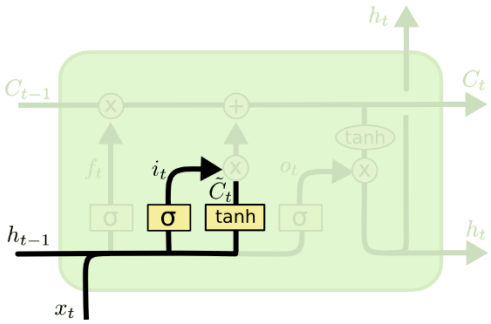
$i_t$ = fungsi *input gate layer*

$W$ = *bobot*

$h_{t-1}$ = satu hidden state sebelumnya dari *time step t*

$x_t$ = vektor *input*

$Ct$ = vector hasil aktivasi *tanh*




Gambar 2.9 Struktur *Remember Gate*

Selanjutnya, state lama akan diperbaharui, $C_{t-1}$ ke state sel baru $C_t$. Kemudian, *ft* akan dikalikan dengan state lama dengan mengabaikan informasi yang sudah dilupakan sebelumnya. Lalu, $i_t$ ditambahkan dengan $C_t$.

$$C_t = f_t * C_{t-1} + i_t * C_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.10)$$

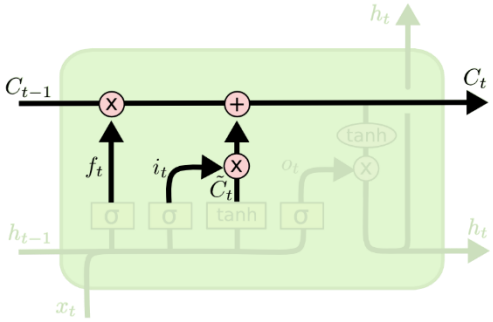
dimana:

$C_t$ = fungsi aktivasi *tanh*

$f_t$ = fungsi pada time step t

$C_{t-1}$ = satu fungsi *tanh state* sebelumnya dari *time step t*

$i_t$ = *input gate layer*




Gambar 2.10 Struktur Update Layer

Langkah terakhir adalah menentukan apa keluarannya. Pertama, *layer sigmoid* akan menentukan bagian dari sel yang akan dikeluarkan. Kemudian, sel tersebut akan dilewatkan pada *layer tanh* (untuk memaksa nilai keluaran diantara -1 dan 1) dan mengalikan dengan keluaran dari gerbang sigmoid.

$$O_t = \sigma(W_0[h_{t-1}, X_t] + b_0) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.11)$$

$$h_t = o_t * \tanh(C_t) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.12)$$

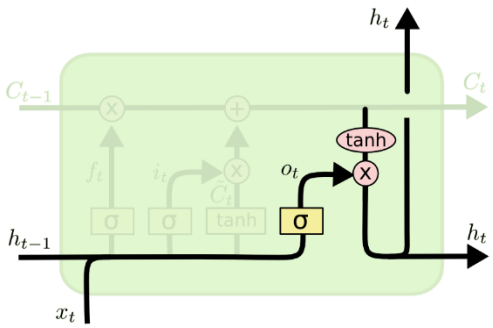
dimana:

$O_t$ = *output layer*

$W_o$ = bobot pada output layer

$X_t$ = vector input

$h_t$ = hidden layer




Gambar 2.11 Struktur Output Layer

*2.2.7.3 RMSE*

Sebuah model peramalan yang dibangun perlu dilakukan evaluasi untuk mengukur tingkat akurasi atau tingkat kesalahan hasil prakiraan yang dibuat. Menurut [46], salah satu metode alternatif untuk mengevaluasi ukuran kesalahan dalam peramalan adalah nilai tengah akar kuadrat atau *Root Mean Square Error* (RMSE). Evalusi 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:

$$RMSE = \sqrt{\frac{1}{m \times n} \sum_{i=0}^{n-1} \sum_{j-0}^{m-1} [f(i,j) - g(i,j)]^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.13)$$

### 2.2.8 Perangkat Pendukung

*2.2.8.1 Python*

Python merupakan salah satu contoh bahasa tingkat tinggi. Contoh lain bahasa tingkat tinggi adalah Pascal, C++, Pert, Java, dan sebagainya. Sedangkan bahasa tingkat rendah merupakan bahasa mesin atau bahasa *assembly*. Secara sederhana, sebuah komputer hanya dapat mengeksekusi program yang ditulis dalam bentuk bahasa mesin. Oleh karena itu, jika suatu program ditulis dalam bentuk bahasa tingkat tinggi, maka program tersebut harus dparaiproses dulu sebelum bisa dijalankan dalam komputer. Hal. ini merupakan salah satu kekurangan bahasa tingkat tinggi yang memerlukan waktu untuk memproses suatu program sebelum program tersebut dijalankan. Akan tetapi, bahasa tingkat tinggi mempunyai banyak sekali keuntungan. Bahasa tingkat tinggi mudah dipelajari, mudah ditulis, mudah dibaca, dan tentu saja mudah dicari kesalahannya. Bahasa tingkat tinggi juga mudah diubah portable untuk disesuaikan dengan mesin yang menjalankannya. Hal ini berbeda dengan bahasa mesin yang hanya dapat digunakan untuk mesin tersebut. Dengan berbagai kelebihan ini, maka banyak aplikasi ditulis menggunakan bahasa tingkat tinggi. Proses mengubah dad bentuk bahasa tingkat tinggi ke tingkat rendah dalam bahasa pemrograman ada dua tipe, yakni interpreter dan compiler.

*2.2.8.2 Keras*

Keras adalah API neural network tingkat tinggi, dibangun dengan Python dan dapat berjalan pada Tensorflow, CNTK, atau Theano yang juga dikenal dengan Keras Backends. Keras dibangun dengan tujuan memungkinkan eksperimen yang cepat. Keras mempunyai motto “Being able to go from idea to result with the least possible delay is key to doing good research.” Keras menyediakan beberapa model untuk CNN, seperti: inception resnet v2, densenet, nasnet, dan densenet

*2.2.8.3 Tensorflow*

TensorFlow adalah perpustakaan perangkat lunak, yang dikembangkan oleh Tim Google Brain dalam organisasi penelitian Mesin Cerdas Google, untuk tujuan melakukan pembelajaran mesin dan penelitian jaringan syaraf dalam. TensorFlow kemudian menggabungkan aljabar komputasi teknik pengoptimalan kompilasi, mempermudah penghitungan banyak ekspresi matematis dimana masalahnya adalah waktu yang dibutuhkan untuk melakukan perhitungan. Fitur utamanya meliputi:

[1] Mendefinisikan, mengoptimalkan, dan menghitung secara efisien ekspresi matematis yang melibatkan array multi dimensi (tensors).

[2] Pemrograman pendukung jaringan syaraf dalam dan teknik pembelajaran mesin.

[3] Penggunaan GPU yang transparan, mengotomatisasi manajemen dan optimalisasi memori yang sama dan data yang digunakan. Tensorflow bisa menulis kode yang sama dan menjalankannya baik di CPU atau GPU. Lebih khusus lagi, TensorFlow akan mengetahui bagian perhitungan mana yang harus dipindahkan ke GPU.

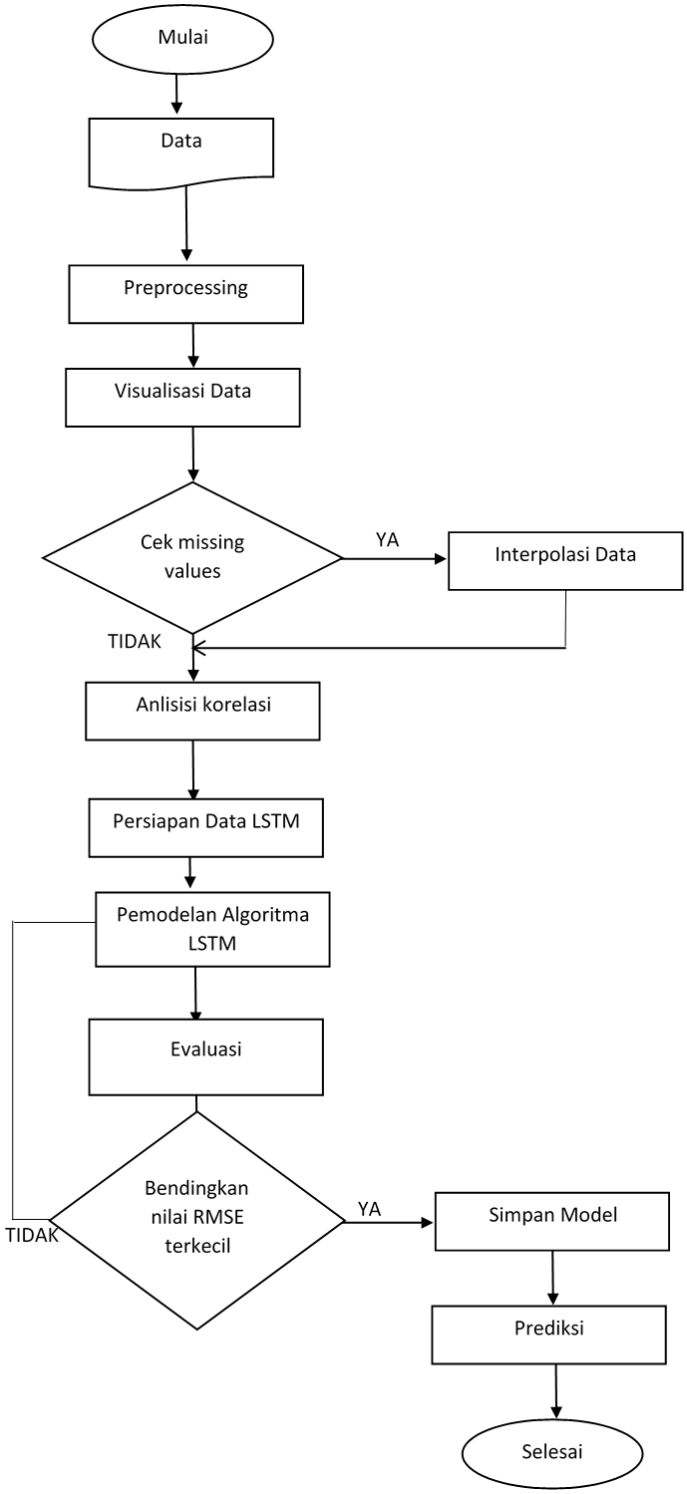
[4] Skalabilitas komputasi yang tinggi di seluruh mesin dan kumpulan data yang besar.

*Halaman ini sengaja dikosongkan*

# BAB 3
# METODE PENELITIAN

## 3.1 Alur Penelitian

Proses prediksi polusi udara dalam penelitian ini terdiri dari 6 langkah utama seperti digambarkan pada Gambar 3.1. Langkah pertama adalah persiapan dataset yang digunakan dalam penelitian ini yang berasal dari pengamatan stasiun pengamatan kualitas udara Dinas Lingkungan Hidup Kota Surabaya yang masuk dalam langkah *Data Preparation*. Kedua, *Data Visualization* yaitu memvisualisasi dataset agar diketahui model data yang diperoleh. Ketiga, sebelum dataset dimasukkan dalam proses algoritma *forcasting*, variabel pendukung dihitung nilai korelasi. Keempat, mempersiapkan data untuk algoritma LSTM. Kelima, proses pemodelan algoritma LSTM dengan membagi data training dan testing serta dilakukan proses iterasi algoritma. Langkah terakhir adalah *Evaluate Model* untuk hasil iterasi dan dihitung error. Jika hasil evaluasi error menggunakna RMSE mendapatkan hasil kecil, maka model disimpan kemudian digunakan untuk membuat prediksi. Namun, jika hasil error masih besar, maka diulangi dengan pemodelan algoritma lagi.




Gambar 3.1 Alur Penelitian

## 3.2 Data Pengamatan Udara di Kota Surabaya

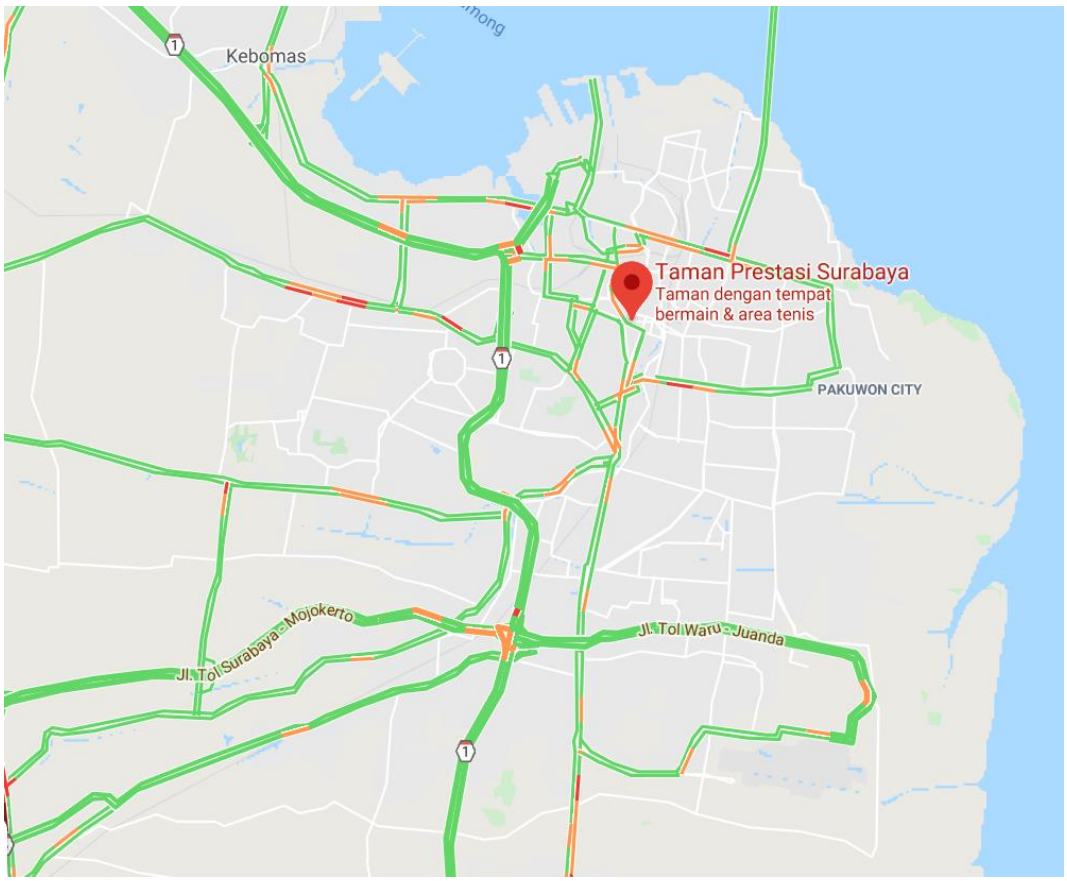
Data yang digunakan dalam penelitian ini dalam bentuk format excel yang didapat dari hasil pengukuran yang dilakukan oleh Dinas Lingkungan Hidup Kota Surabaya. Dinas Lingkungan Hidup Kota Surabaya melakukan pengukuran kualitas udara dalam satuan mg/m$^3$. Dinas Lingkungan Hidup Kota Surabaya memiliki 7 stasiun pemantauan kualitas udara seperti tertera dalam tabel 3.1. Dari 7 titik pemantauan tersebut hanya 3 stasiun yang aktif yaitu Taman Prestasi, Wonorejo dan Kebonsari. Dalam penelitian ini mengambil data dari 1 stasiun pemantauan yaitu Taman Prestasi.

*Tabel 3.1 Stasiun pemantuan kualitas udara Kota Surabaya*

| SUF1 | SUF3 | SUF4 | SUF5 | SUF6 | SUF7 |
|---|---|---|---|---|---|
| Taman Prestasi | Sukomanunggal | Gayungan | Gebang Putih | Wonorejo | Kebonsari |

Jumlah parameter yang diukur ada 14. Nilai parameter yang diukur adalah CO (*Carbon monoxide* - Karbon monoksida), NO (*Nitrogen monoxide* - Nitrogen monoksida), NO2 (*Nitrogen dioxide* - Nitrogen dioksida), NOX (*Nitrogen oxides* - Nitrogen oksida), PM10 (*Floating dust* - Debu), SO2 (*Sulfur dioxide* - Sulfur dioksida), O3 (*Ozone* - Ozon). Sedangkan parameter pendukung dari data meteorologi terdiri dari : 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), HUMair (*Humidity air* - Kelembaban udara), TEMPair (*Temperature air* - Temperatur udara). Kualitas udara dipantau tiap 30 menit sekali, sehingga dalam sehari mendapatkan 48 data pantau dari 14 parameter udara. Data penelitian ini diambil dari 14 Maret 2015 sampai dengan 4 Agustus 2019. Dalam penelitian ini fokus pada analisa parameter Karbon Monoksida (CO) dan dikorelasikan dengan data meteorologi. Laporan dari server stasiun pemantauan dapat dilihat pada Gambar 3.3. Gambar 3.3 merupakan laporan dari stasiun pemantauan dari parameter CO dalam bentuk file excel dan menunjukkan nilai dari 7 stasiun pemantauan. Dalam penelitian ini mengambil dataset dari SUF-1 Taman Prestasi. Lokasi Stasiun Pemantauan SUF-1 berada di halaman taman prestasi Jl. Ketabang Kali, Surabaya. Lokasi ini mewakili daerah pusat kota, pemukiman,

perkantoran, dan perdagangan Surabaya Pusat seperti digambarkan pada Gambar 3.2. Kawasan ini telah berkembang sebagai pusat bisnis di wilayah Jawa Timur dan menjadi salah satu jantung utama kegiatan bisnis dan perdagangan di Surabaya. Beberapa ciri khas bangunan yang ada di kawasan ini di antaranya adalah Wisma BRI Surabaya, Hotel Bumi Surabaya, Wisma Dharmala Surabaya, The Peak Residence, Sheraton Hotel, dan lain sebagainya.




Gambar 3.2 Peta lokasi SUF-1 Taman Prestasi

Sistem Pemantauan Kualitas Udara Kota Surabaya

**Laporan harian kualitas udara untuk parameter**
**Daily Air Quality Report for Parameters**

| | |
|---|---|
| **Tanggal / Date:** | **05/01/2017** |
| **Parameter / Parameter:** | **CO: Karbon monoksida / Carbon monoxide** |
| **Satuan / Unit:** | **mg/m3** |

Laporan disiapkan di / Report prepared at: Data Center Surabaya

Laporan berdasarkan data setengah jam rata-rata / Report based on half hour mean values

| Waktu / Time GMT+7 (WIB) | Surabaya SUF1 Taman Prestasi | Surabaya SUF3 Sukomanunggal | Surabaya SUF4 Gayungan | Surabaya SUF5 Gubeng Pojok | Surabaya SUF6 Wonocolo | Surabaya SUF7 Kebonsari |
|---|---|---|---|---|---|---|
| 0:30 | 0.23 | --- | --- | --- | 0.00 | --- |
| 1:00 | 0.23 | --- | --- | --- | 0.00 | --- |
| 1:30 | 0.23 | --- | --- | --- | 0.01 | --- |
| 2:00 | 0.23 | --- | --- | --- | 0.16 | --- |
| 2:30 | 0.20 | --- | --- | --- | --- | --- |
| 3:00 | 0.16 | --- | --- | --- | 0.00 | --- |
| 3:30 | 0.13 | --- | --- | --- | 0.00 | --- |
| 4:00 | 0.12 | --- | --- | --- | 0.00 | --- |
| 4:30 | 0.13 | --- | --- | --- | 0.00 | --- |
| 5:00 | 0.13 | --- | --- | --- | 0.00 | --- |
| 5:30 | 0.13 | --- | --- | --- | 0.07 | --- |
| 6:00 | 0.20 | --- | --- | --- | 0.18 | --- |
| 6:30 | 0.23 | --- | --- | --- | 0.54 | --- |
| 7:00 | 0.23 | --- | --- | --- | 0.73 | 0.00 |
| 7:30 | --- | --- | --- | --- | 0.55 | --- |
| 8:00 | 0.31 | --- | --- | --- | 0.22 | --- |
| 8:30 | 0.33 | --- | --- | --- | 0.13 | --- |
| 9:00 | 0.33 | --- | --- | --- | 0.00 | --- |
| 9:30 | 0.26 | --- | --- | --- | 0.00 | --- |
| 10:00 | 0.23 | --- | --- | --- | 0.00 | --- |
| 10:30 | 0.23 | --- | --- | --- | 0.33 | --- |
| 11:00 | 0.23 | --- | --- | --- | 0.13 | --- |
| 11:30 | 0.23 | --- | --- | --- | 0.09 | --- |
| 12:00 | 0.20 | --- | --- | --- | 0.00 | --- |
| 12:30 | 0.21 | --- | --- | --- | 0.04 | --- |
| 13:00 | 0.23 | --- | --- | --- | 0.00 | --- |
| 13:30 | 0.21 | --- | --- | --- | 0.07 | --- |
| 14:00 | 0.20 | --- | --- | --- | 0.00 | --- |
| 14:30 | 0.20 | --- | --- | --- | 0.01 | --- |
| 15:00 | 0.23 | --- | --- | --- | 0.03 | --- |
| 15:30 | 0.23 | --- | --- | --- | 0.00 | --- |
| 16:00 | 0.23 | --- | --- | --- | 0.00 | --- |
| 16:30 | 0.23 | --- | --- | --- | 0.01 | --- |
| 17:00 | 0.23 | --- | --- | --- | 0.19 | --- |
| 17:30 | 0.23 | --- | --- | --- | 0.05 | --- |
| 18:00 | 0.27 | --- | --- | --- | 0.08 | --- |
| 18:30 | 0.25 | --- | --- | --- | 0.02 | --- |
| 19:00 | 0.23 | --- | --- | --- | 0.06 | --- |
| 19:30 | 0.23 | --- | --- | --- | 0.18 | --- |
| 20:00 | 0.23 | --- | --- | --- | 0.09 | --- |
| 20:30 | 0.23 | --- | --- | --- | 0.37 | --- |
| 21:00 | 0.23 | --- | --- | --- | 0.30 | --- |
| 21:30 | 0.23 | --- | --- | --- | 0.16 | --- |
| 22:00 | 0.23 | --- | --- | --- | 0.02 | --- |
| 22:30 | 0.28 | --- | --- | --- | 0.23 | --- |
| 23:00 | 0.26 | --- | --- | --- | 0.21 | --- |
| 23:30 | 0.29 | --- | --- | --- | 0.20 | --- |
| 24:00:00 | 0.27 | --- | --- | --- | 0.15 | --- |
| Minimum / Minimum | 0.12 | --- | --- | --- | 0.00 | 0.00 |
| **ai Rata-rata / Me** | 0.22 | --- | --- | --- | 0.12 | 0.00 |
| Maksimum / Maximum | 0.33 | --- | --- | --- | 0.73 | 0.00 |

**Simbolan / Legend:**
1] Nilai tersedia/Value available
2] Nilai antara SBX dan SBXValue between SBX and SBX
3
4] Nilai tidak valid/Value invalid
5] Nilai tak terukur/Value unsure
6] pemeriksaan nilai terakhir/Value finally checked

| Disiapkan oleh / Prepared by | Diperiksa oleh / Checked by |
|---|---|
| Tanggal / Tanda tangan / Date / Signature | Tanggal / Tanda tangan / Date / Signature |

Gambar 3.3 Laporan kualitas udara dari stasiun pemantauan kualitas udara di Kota Surabaya

## 3.3 Preprocressing

Pada tahap ini, data mentah dalam bentuk excel digabungkan menjadi satu file dataset dan disimpan dalam bentuk Csv (*Comma Sparared Value*). Sepert gambar 3.3 informasi data yang ada dala tiap file laporan adalah tanggal, parameter, satuan, waktu (jam) saat pengambilan data, lokasi stasiun pemantauan (SUF-1 sampai dengan SUF-7). Dari laporan tersebut data yang di-ekstrak adalah tanggal, parameter, jam, SUF-1. Data yang dipakai dalam penelitian adalah parameter Karbon Monoksida (CO) dan data meteorologi. Satuan CO yang dikeluarkan dari stasiun pemantauan udara menggunakan satuan mg/m$^3$ kemudian dikonversikan dalam perhitungan standar ISPU menggunakan persamaaan 2.1. Hasil preprocessing dapat dilihat pada Gambar 3.4. Gambar 3.4 merupakan dataset hasil preprocessing yang terdiri dari jumlah baris 77.040 dan 8 kolom (parameter polutan udara: CO yang sudah dikonversi menjadi nilai ISPU dan data meteorologi : DD, DD_Boe, FF, FF_Boe, GRAD, HUMair dan TEMP air).

Out[25]:

| | CO_ISPU | DD | DD_Boe | FF | FF_Boe | GRAD | HUMair | TEMPair |
|---|---|---|---|---|---|---|---|---|
| day time | | | | | | | | |
| 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 × 8 columns

Gambar 3.4. Hasil preprocessing

## 3.4 Visualisasi Data

Visualisasi data ini dimaksudkan untuk melihat karakteristik data. Data digambarkan pada chart line. Dalam visualisasi data ini akan dilakukan analisis trend harian, mingguan dan bulanan. Selain itu akan dilihat jumlah data yang hilang.

### 3.5 Analisis Korelasi

Dalam peneltian ini, analisis korelasi antar parameter polutan CO dengan parameter meteorologi 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 koefiesien. Adapun kriteria angka koefisien memiliki kriteria sebagai berikut:

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

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 angko koefisien bernilai negatif. Dalam hubungan dengan variabel adalah jika variabel X nilainya tinggi, maka variabel Y rendah.

### 3.6 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

memperolah hasil prediksi yang baik karena data tiap waktu memiliki nilai. Dalam penelitian ini interpolasi yang digunakan adalah *spline*.

### 3.6.1 Normalisasi Data

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. Proses normalisasi dapat dilakukan dengan menggunakan persamaan di bawah ini:

$$X_{norm} = (x - x_{min})/(x_{max} - x_{min}) \quad \text{.....(3.1)}$$

dimana $X_{norm}$ adalah nilai normalisasi, $X_{min}$ dan $X_{max}$ adalah nilai minimum dan nilai maksimum dari data *training*. Data yang telah dinormalisasi akan digunakan sebagai input kedalam jaringan LSTM. Keluaran dari jaringan tersebut akan melalui proses denormalisasi untuk mengestimasi nilai prediksi yang aktual. Hasil normalisasi data dapat dilihat pada Gambar 3.4

```
Out[7]: array([[0.00616147, 0.26876032, 0.21398489, ..., 0.44986463, 0.9581206 ,
                0.96044797],
               [0.00616147, 0.26923224, 0.21466568, ..., 0.44986463, 0.9581206 ,
                0.9604207 ],
               [0.00589358, 0.268412  , 0.2187942 , ..., 0.44986463, 0.9581206 ,
                0.9601211 ],
               ...,
               [0.06834674, 0.23154272, 0.20258069, ..., 0.4501698 , 0.6891583 ,
                0.9548843 ],
               [0.06834674, 0.2176578 , 0.20769902, ..., 0.45013618, 0.700973  ,
                0.9546119 ],
               [0.06834674, 0.22810802, 0.20076777, ..., 0.45016035, 0.7205385 ,
                0.9546732 ]], dtype=float32)
```

Gambar 3.5 Hasil normalisasi data

### 3.6.2 Supervised Learning Problem

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. Supervised learning dapat dipahami bahwa ketika ada sebuah data masukan dengan variabel X dan variabel keluaran y dan sebuah algoritma digunakan untuk mempelajari fungsi pemetaan dari masukan menuju keluaran seperti formula berikut:

$$y = f(X).$$

Tujuan dari fungsi pemetaan ini adalah untuk mengaproksimasi pemetaan sedemikian rupa sehingga ketika ada masukan data baru (X), prediksi terhadap variabel keluaran (y) dapat dilakukan dan akurat. Data polusi udara dalam penelitian merupakan data waktu serial, sejumlah angka sekuens yang diurutkan berdasarkan indeks waktu, dalam kasus data dalam penelitian ini, indeks waktu adalah data tanggal dan jam. Gambar 3.4 merupakan data hasil preprocessing nilai polutan CO dan data meteorologi di taman prestasi yang akan dibuat model supervised learning problem. Ini bisa dimodelkan sebagai sebuah list atau kolom dengan nilai yang diurutkan.

Untuk mengubah ke dalam bentuk Supervised Learning Problem adalah dengan cara menggunakan langkah waktu sebelumnya sebagai variabel input dan menggunakan langkah waktu berikutnya sebagai variabel output. Tabel 3.2 merupakan nilai polutan CO pada tanggal 17/07/2017 dari jam 00.30.00 sampai dengan 07.30.00.

*Tabel 3.2 Dataset Polutan Co (mg/m$^3$)*

| 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 |

Jika diubah ke dalam model *Supervised Learning Problem* maka bisa dilihat pada tabel 3.3. Di tabel 3.3 dapat diartikan bahwa *var(t)* adalah nilai yang akan diprediksi yang masih terpengaruh nilai *var(t-1), var(t-2), var(t-3).*

*Tabel 3.3 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 |

Dalam penelitian ini karena sudah dinormalisasi, maka hasil pemodelan *Supervised Learning Problem* dapat dilihat pada Gambar 3.5. Gambar 3.5 dapat dilihat bahwa hasil dari pemodelan mendapatkan 16 kolom. Variabel data sebagai input yang digunakan adalah *var1( t-1), var2(t-1),…, var8(t-1)*. Sedangkat ouput yang akan digunakan dalam prediksi adalah *var1(t).* Untuk *var2(t), var3(t),…., var8(t)* akan dihapus karena tidak digunakan dalam prediksi. Hasil akhir dari *Supervised Learning Problem* dapat dilihat pada Gambar 3.6.

Out[10]:

| var1(t-1) | var2(t-1) | var3(t-1) | var4(t-1) | var5(t-1) | var6(t-1) | var7(t-1) | var8(t-1) | var1(t) | var2(t) | var3(t) | var4(t) | var5(t) | var6(t) | var7(t) | var8( |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| .006161 | 0.268760 | 0.213985 | 0.198214 | 0.796970 | 0.449865 | 0.958121 | 0.960448 | 0.006161 | 0.269232 | 0.214666 | 0.147707 | 0.792401 | 0.449865 | 0.958121 | 0.9604 |
| .006161 | 0.269232 | 0.214666 | 0.147707 | 0.792401 | 0.449865 | 0.958121 | 0.960421 | 0.005894 | 0.268412 | 0.218794 | 0.322870 | 0.796058 | 0.449865 | 0.958121 | 0.9601 |
| .005894 | 0.268412 | 0.218794 | 0.322870 | 0.796058 | 0.449865 | 0.958121 | 0.960121 | 0.006161 | 0.267348 | 0.214280 | 0.292754 | 0.800627 | 0.449865 | 0.958121 | 0.9603 |
| .006161 | 0.267348 | 0.214280 | 0.292754 | 0.800627 | 0.449865 | 0.958121 | 0.960329 | 0.006161 | 0.262210 | 0.214616 | 0.159967 | 0.774107 | 0.449865 | 0.958121 | 0.9597 |
| .006161 | 0.262210 | 0.214616 | 0.159967 | 0.774107 | 0.449865 | 0.958121 | 0.959791 | 0.006161 | 0.267846 | 0.214315 | 0.261263 | 0.780511 | 0.449865 | 0.958121 | 0.9588 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| .113317 | 0.237196 | 0.218781 | 0.017089 | 0.752159 | 0.450124 | 0.631719 | 0.957002 | 0.096896 | 0.226919 | 0.197339 | 0.091508 | 0.750329 | 0.450126 | 0.642266 | 0.9564 |
| .096896 | 0.226919 | 0.197339 | 0.091508 | 0.750329 | 0.450126 | 0.642266 | 0.956451 | 0.070631 | 0.235681 | 0.207988 | 0.038414 | 0.749415 | 0.450096 | 0.675406 | 0.9559 |
| .070631 | 0.235681 | 0.207988 | 0.038414 | 0.749415 | 0.450096 | 0.675406 | 0.955916 | 0.068347 | 0.231543 | 0.202581 | 0.041729 | 0.738441 | 0.450170 | 0.689158 | 0.9548 |
| .068347 | 0.231543 | 0.202581 | 0.041729 | 0.738441 | 0.450170 | 0.689158 | 0.954884 | 0.068347 | 0.217658 | 0.207699 | 0.020122 | 0.751244 | 0.450136 | 0.700973 | 0.9546 |
| .068347 | 0.217658 | 0.207699 | 0.020122 | 0.751244 | 0.450136 | 0.700973 | 0.954612 | 0.068347 | 0.228108 | 0.200768 | 0.016680 | 0.736611 | 0.450160 | 0.720538 | 0.9546 |

vs × 16 columns

Gambar 3.6. Proses Supervised Learning Problem

```
   var1(t-1)  var2(t-1)  var3(t-1)  var4(t-1)  var5(t-1)  var6(t-1)  \
1   0.006161   0.268760   0.213985   0.198214   0.796970   0.449865
2   0.006161   0.269232   0.214666   0.147707   0.792401   0.449865
3   0.005894   0.268412   0.218794   0.322870   0.796058   0.449865
4   0.006161   0.267348   0.214280   0.292754   0.800627   0.449865
5   0.006161   0.262210   0.214616   0.159967   0.774107   0.449865

   var7(t-1)  var8(t-1)   var1(t)
1   0.958121   0.960448  0.006161
2   0.958121   0.960421  0.005894
3   0.958121   0.960121  0.006161
4   0.958121   0.960329  0.006161
5   0.958121   0.959791  0.006161
```

Gambar 3.7 Hasil Supervised Learning Prolem

## 3.7 Pemodelan Algoritma RNN LSTM

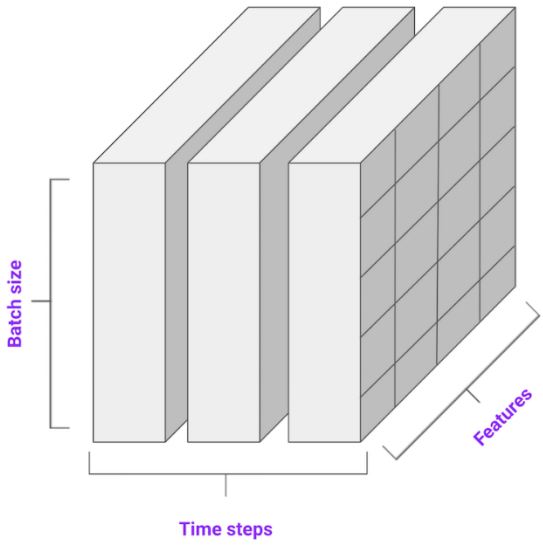
Setelah dataset siap dalam bentuk *Supervised Learning Problem*, selanjutnya menentukan inisialasi parameter LSTM yang dibutuhkan, antara lain membagi dataset menjadi data *train* and *test*, mendefisinisikan neuron *(hidden layer*), dan mendefinisikan jumlah *epoch* untuk iterasi dalam satu jaringan LSTM. Selanjutnya maka proses pembelajaran jaringan LSTM bisa dilakukan.

Pada dasarnya model LSTM memerlukan masukan dengan format *batch size, time steps, features* yang dapat digambarkan seperti sebuah array 3 dimensi seperti pada Gambar 3.7. Setiap dimensi dapat dijelaskan sebagai berikut:

1. *Batch Size* : batch size menunjukkan bahwa berapa banyak sampel dari masukan yang ingin dimasukkan sebelum memperbaharui bobot dari sebuah jaringan saraf tiruan. Misalkan sebuah data memiliki 100 buah sampel dataset masukan dan model jaringan saraf ingin diperbaharui bobotnya tiap kali menerima sebuah masukan. Untuk kasus seperti ini

maka nilai batch size adalah 1. Sebaliknya, jika model jaringan saraf ingin diperbaharui bobotnya setelah menerima semua sampel, maka nilai batch size adalah 100 dan jumlah batch adalah 1. Perlu diketahui bahwa jika menggunakan batch size yang sangat kecil maka akan mengurangi kecepatan untuk mempelajari data. Di sisi lain jika menggunakan terlalu banyak *batch size* akan mengurangi kemampuan data untuk mengeneralisasi terhadap data lainnya dan juga menambah konsumsi memori. Namun nilai *batch size* ini tidak ada standart yang mutlak sehingga untuk mendapatkan nilai *batch size* yang optimal untuk proses training harus melakukan percobaan. Dalam penelitian ini, setelah dilakukan percobaan berulang kali, nilai *batch size* yang digunakan adalah 150.

2. *Time Steps* : mendefinisikan berapa banyak indeks waktu yang ingin dipelajari.
3. *Features* : adalah jumlah atribut yang digunakan untuk merepresentasikan setiap time step.




Gambar 3.8 Bentuk 3D masukan LSTM

*(Sumber: https://www.tensorflow.org/)*

Dalam penelitian prediksi polusi udara di kota Surabaya ini, nilai batch size yang digunakan adalah 150, nilai time steps adalah 1, dan nilai features adalah 8 yang terdiri dari nilai CO, 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), HUMair (*Humidity air* - Kelembaban udara), TEMPair (*Temperature air* - Temperatur udara).

Kemudian dalam mendesain model yang akan digunakan, disini menggunakan model *Sequential*(), yaitu pemodelan yang berisikan lapisan-lapisan linear disusun bertumpuk. Dengan kata lain, setelah model pertama selesai dijalankan, akan dilanjutkan dengan model kedua, dan seterusnya. Dalam kasus prediksi ini, ada 2 model yang disusun secara linear yang terdiri dari poin-poin dibawah ini:

1. *LSTM* : nilai unit dari LSTM diisikan 10 berarti keluaran dari model ini berupa keluaran dengan dimensi 10.
2. *Dense* : Secara matematis, lapisan Dense digunakan untuk mengubah dimensi dari vektor dan mengaplikasikan perhitungan vektor seperti rotasi, skala. translasi, dan transformasi. Dalam kasus ini Dense diisi dengan angka 1 yang berarti keluaran dari lapisan merupakan vektor dengan dimensi 1.

Sebagai bagian dari optimisasi algoritma, kesalahan kondisi model harus diperkirakan berulang kali. Ini membutuhkan pilihan fungsi kesalahan, secara konvensional disebut *fungsi loss*, yang dapat digunakan untuk memperkirakan *loss model* sehingga *weight* dapat diperbarui untuk mengurangi *loss* pada evaluasi selanjutnya. *Neural network model* mempelajari pemetaan dari input ke output dari contoh dan pilihan fungsi loss harus sesuai dengan framing masalah pemodelan prediktif spesifik, seperti klasifikasi atau regresi. Lebih lanjut, konfigurasi lapisan keluaran juga harus sesuai untuk fungsi loss yang dipilih. Dalam penelitian ini menggunakan model regresi sehingga fungsi loss yang digunakan adalah *Mean Squared Error*, atau MSE. MSE adalah default loss yang digunakan untuk masalah regresi. *Mean squared error* dihitung sebagai rata-rata perbedaan kuadrat antara nilai yang diprediksi dan yang sebenarnya. Hasilnya selalu positif terlepas dari

tanda nilai yang diprediksi dan aktual dan nilai sempurna adalah 0,0. Kuadrat berarti bahwa kesalahan yang lebih besar menghasilkan lebih banyak kesalahan daripada kesalahan yang lebih kecil, yang berarti bahwa model akan dihukum karena membuat kesalahan yang lebih besar. Fungsi *mean squared error loss* dapat digunakan dalam Keras dengan menuliskan ‘mse‘ atau ‘mean_squared_error‘ sebagai fungsi loss saat menyusun model.

```
model.compile(loss='mean_squared_error')
//atau bisa disingkat dengan kata ‘mse’ seperti berikut:
model.compile(loss=’mse’)
```

Optimasi algoritma yang digunakan dalam penelitian ini adalah *adaptive moment estimation* (adam)[47]. Algoritma ini diperkenalkan oleh Kingm dan Ba (2015). Secara singkat konsepnya adalah, learning rate bersifat adaptif. Perubahan learning rate dihitung dari estimasi momentum. Momentum adalah nilai yang dihitung berdasarkan arah dari pelatihan sebelumnya. Dalam Adam, kuadrat gradien disimpan dan disimpan juga rata-rata gradien sebelumnya untuk perhitungan momentum. Metode ini dirancang untuk menggabungkan keuntungan dari dua metode yang sudah popular sebelumnya yaitu AdaGrad yang dimunculkan pada tahun 2011 dan RMSProp yang muncul pada tahun 2012.

### 3.8 Evaluasi

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

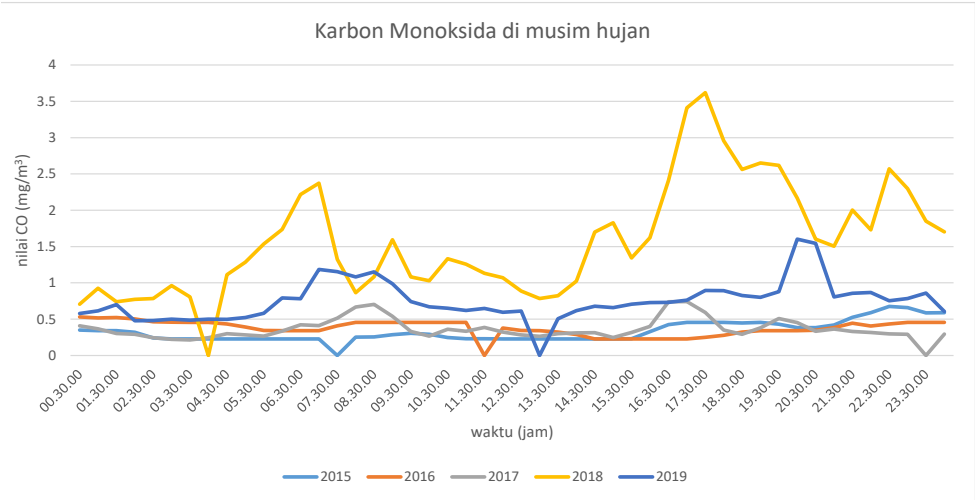
*Halaman ini sengaja dikosongkan*

# BAB 4
# HASIL DAN PEMBAHASAN

## 4.1 Eksplorasi Data

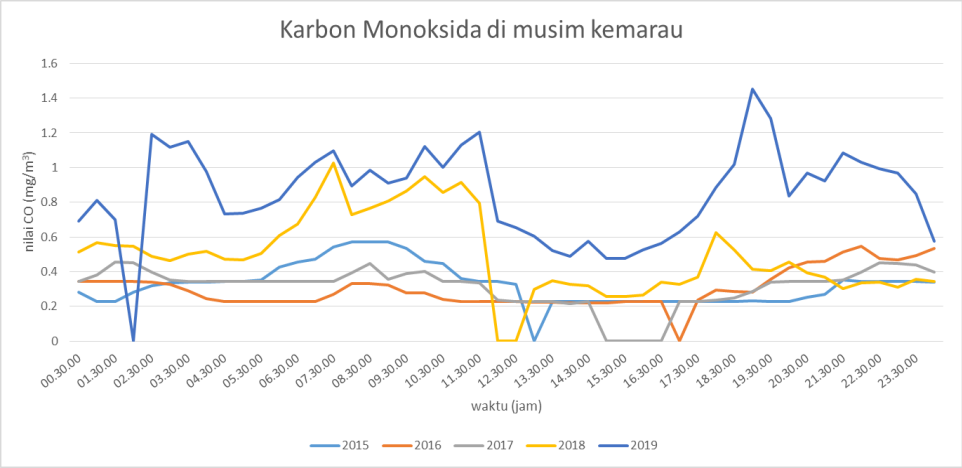
Indonesia merupakan negara yang secara astronomis terletak di garis khatulistiwa sehingga beriklim tropis. Musim yang dimiliki Indonesia ada dua yaitu, musim hujan dan kemarau. Berlangsungnya musim kemarau di Indonesia terjadi antara bulan April-September dan bersamaan dengan bertiupnya angin musim timur. Sedangkan musim hujan di Indonesia terjadi antara bulan Oktober - Februari karena bertiupnya angin musim barat. Bulan Maret merupakan bulan peralihan musim, yang biasanya disebut musim pancaroba. Pada penelitian ini akan digambarkan analisa tren di kedua musim tersebut secara harian dan mingguan, bulanan dan tahunan pada kandungan karbon monoksida di kota Surabaya dengan titik stasiun pemantauan di  Taman Prestasi.

Gambar 4.1 dan Gambar 4.2 menunjukkan bahwa ada perbedaan tren harian pada nilai karbon monoksida antara musim penghujan dan kemarau. Data diambil secara acak di dua musim, tanpa membedakan hari kerja atau hari libur. Di musim penghujan atau pun kemarau tren mulai naik terjadi rata-rata pada jam 05.00 dan 16.00. Perbedaan terjadi pada tren turun. Di musim penghujan nilai Karbon



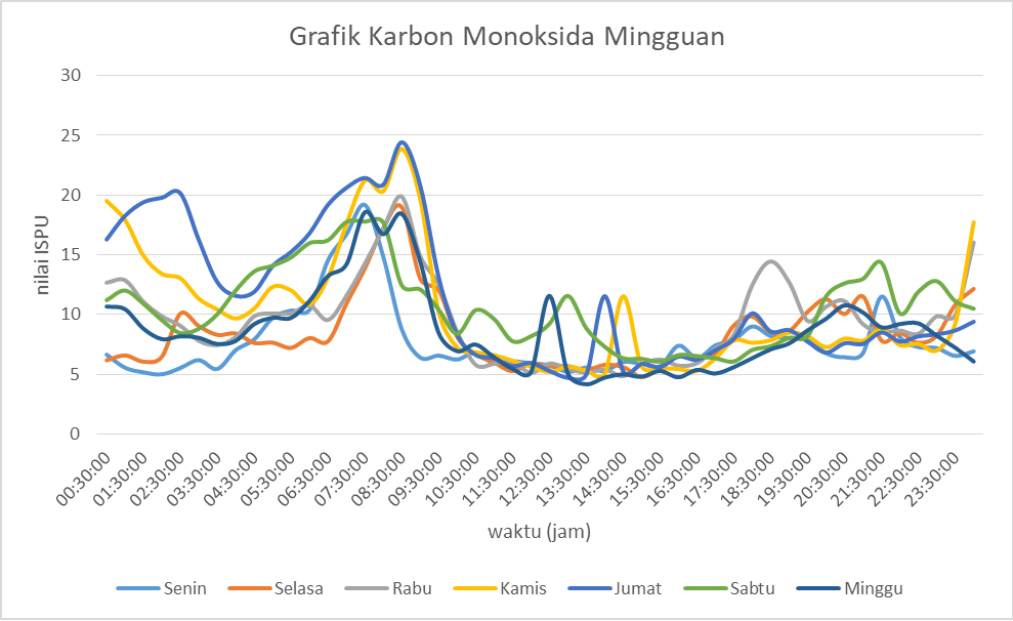

Gambar 4.1 Grafik Harian Karbon monoksida  musim penghujan

monoksida mulai turun pada jam 09.00 sedangkan pada musim kemarau tren turun pada jam 10.00 atau 11.30.



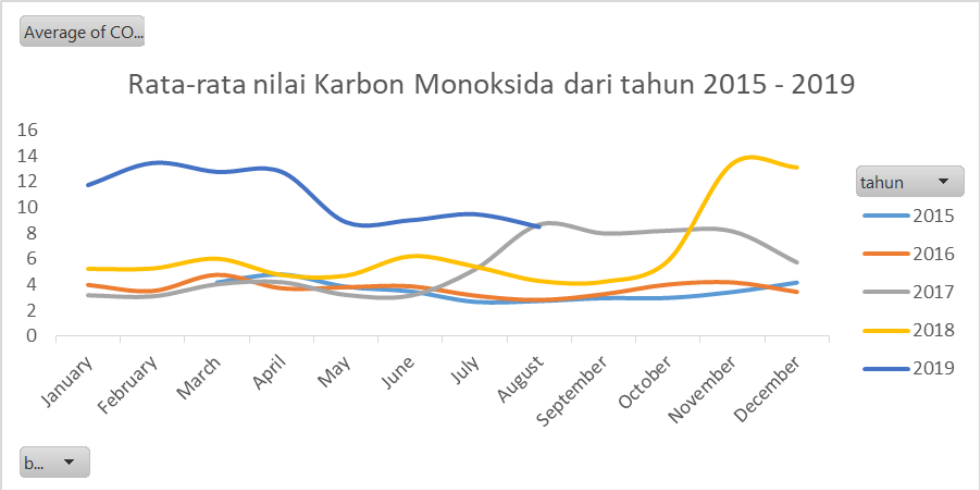

Gambar 4.2 Grafik Harian Karbon monoksida musim kemarau

Gambar 4.3 menunjukkan bahwa rata-rata nilai kandungan karbon monoksida mencapai nilai tinggi pada hari Kamis dan Jumat. Nilai tersebut didapat dari akumulasi dari hari kerja dari Senin sampai dengan Rabu. Untuk *weekend* atau hari Sabtu dan Minggu mengalami penurunan sehingga di hari Senin pagi, kandungan CO tidak sebesar di hari kerja lainnya.




Gambar 4.3 Grafik Mingguan Karbon Monoksida

Kemudian, analisis tahunan dengan mengambil nilai rata-rata karbon monoksida per bulan. Kandungan CO mengalami tren naik tiap tahun. Tren naik mulai kelihatan pada semester kedua tahun 2017 dan mengalami kenaikan cukup signifikan pada bulan Oktober dan Nopember 2018. Pada tahun 2015 dan 2016 kandungan CO mengalami kenaikan dan penurunan secara landai. Rata-rata kenaikan kandungan karbon monoksida pada tahun 2019 sebesar 4,53% dari tahun 2018. Ini menjadi kenaikan yang tinggi selama 4 tahun terakhir. Nilai kandungan monoksida selama 5 tahun terakhir tertinggi terjadi pada tahun 2017 sebesar 106,87 dan terendah terjadi pada tahun 2018 sebesar 0.48. Tabel 4.1 menunjukkan analisis statistic dengan parameter rata-rata, nilai minimal dan nilai maksimal selama 5 tahun terakhir.




Gambar 4.4 Grafik Tahunan Karbon Monoksida

Pada table 4.2 menunjukkan bahwa nilai kandungan karbon monoksida selama 5 tahun terakhir secara umum masih dalam tahap BAIK. Indeks SEDANG dari tahun ke tahun mengalami kenaikan artinya kandungan karbon monoksida mengalami kenaikan. Di tahun 2017 ada satu nilai sangat tinggi yaitu 106,87. Nilai tinggi ini bisa diduga sebagai nilai percilan atau *outlier* pada data seris.

*Tabel 4.1 Analisis Data berdasarkan rata-rata, nilai minimal dan maksimal*

| Tahun | Rata-rata | Min | Max |
|---|---|---|---|
| 2015 | 3,50 | 1,10 | 20,53 |
| 2016 | 3,72 | 1,11 | 46,63 |
| 2017 | 5,42 | 0,99 | 106,87 |
| 2018 | 6,55 | 0,48 | 71,69 |
| 2019 | 11,08 | 2,66 | 78,18 |

*Tabel 4.2 Indeks ISPU pada kandungan CO*

| INDEKS ISPU | JUMLAH KLASIFIKASI NILAI INDEKS (COUNT) | | | | |
|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 |
| BAIK | 14.064 | 17.568 | 17.514 | 17.510 | 10.331 |
| SEDANG | | | 5 | 10 | 37 |
| TIDAK SEHAT | | | 1 | | |
| Grand Total | 14.064 | 17.568 | 17.520 | 17.520 | 10.368 |

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 memiliki 3,60% setara 2.774, DD Boe memiliki 3,58% setara 2761, FF memiliki 3,58% setara 2760, FF Boe memiliki 3,55% setara 2.734, GRAD memiliki 10,38% setara 8000, HUMair memiliki 0,87% setara 671 dan TEMPair memiliki 0,83% setara 639.

## 4.2 Analisis Korelasi

Hasil perhitungan korelasi antar variabel dapat dilihat pada Gambar 4.4 . Dari gambar 4.4 dapat diinterpretasikan hubungan antar variabel bahwa korelasi variabel polutan CO dengan data meteorologi jika dilihat arahnya, maka data meteorologi yang memiliki korelasi searah yakni 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 dari polutan CO. Artinya bahwa jika data meteorologi yang berkorelasi negatif nilainya bertambah tinggi, maka akan mengurangi nilai polutan CO. Korelasi tertinggi dari data meteorologi adalah FF_Boe (Kecepatan hembusan angin) sebesar -0,32 (berkorelasi negatif).

| | CO | DD | DD_Boe | FF | FF_Boe | GRAD | HUMair | TEMPair |
|---|---|---|---|---|---|---|---|---|
| CO | 1 | 0.11 | 0.055 | -0.25 | -0.32 | -0.12 | 0.24 | -0.18 |
| DD | 0.11 | 1 | 0.52 | 0.18 | 0.0016 | -0.14 | 0.32 | -0.28 |
| DD_Boe | 0.055 | 0.52 | 1 | 0.076 | 0.028 | -0.094 | 0.2 | -0.19 |
| FF | -0.25 | 0.18 | 0.076 | 1 | 0.63 | 0.16 | -0.2 | 0.14 |
| FF_Boe | -0.32 | 0.0016 | 0.028 | 0.63 | 1 | 0.4 | -0.55 | 0.39 |
| GRAD | -0.12 | -0.14 | -0.094 | 0.16 | 0.4 | 1 | -0.62 | 0.51 |
| HUMair | 0.24 | 0.32 | 0.2 | -0.2 | -0.55 | -0.62 | 1 | -0.87 |
| TEMPair | -0.18 | -0.28 | -0.19 | 0.14 | 0.39 | 0.51 | -0.87 | 1 |

Gambar 4.5 Korelasi CO dengan data meteorologi

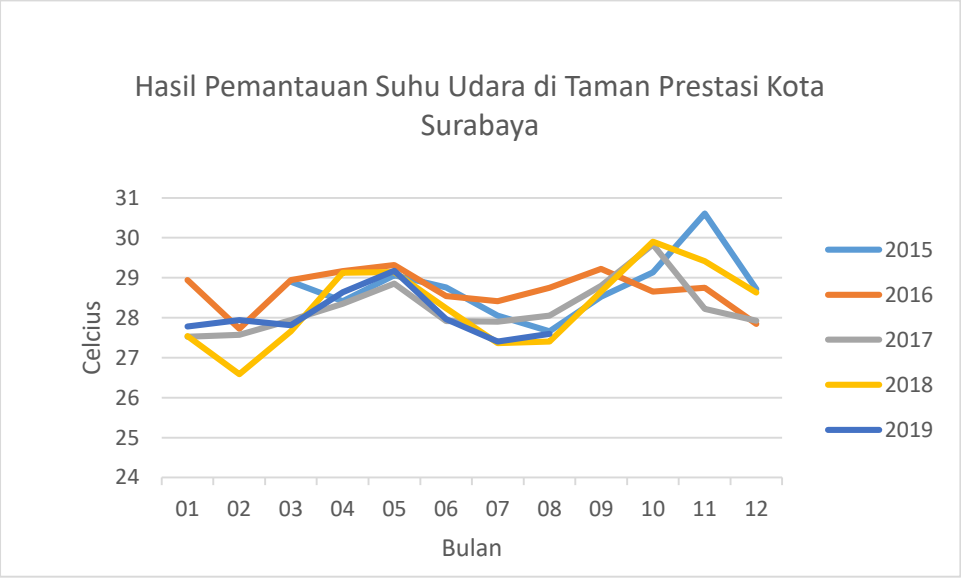
## 4.3 Analisis Meteorologis Kota Surabaya

Suhu udara merupakan salah satu data meteorologi yang mempengaruhi Karbon Monoksida yang berkorelasi tidak searah, artinya ketinggian suhu akan mempengaruhi penurunan kandungan Karbon Monoksida. Tabel 4.3 merupakan nilai minimal, maksimal dan rata-rata suhu yang ada di Stasiun Pemantauan Kualitas Udara di Taman Prestasi dari bulan Maret 2015 sampai dengan bulan Agustus 2019.

*Tabel 4.3 Nilai Minimal, Maksimal dan Rata-rata Suhu Udara Taman Prestasi*

| Tahun / Bulan | 2015 | | | 2016 | | | 2017 | | | 2018 | | | 2019 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Min | Max | Rata-rata | Min | Max | Rata-rata | Min | Max | Rata-rata | Min | Max | Rata-rata | Min | Max | Rata-rata |
| Januari | | | | 23,5 | 37,1 | 28,9 | 23,4 | 34,8 | 27,5 | 23,8 | 34,1 | 27,5 | 23,5 | 35,0 | 27,8 |
| Februari | | | | 23,4 | 35,7 | 27,7 | 23,4 | 36,5 | 27,6 | 16,9 | 34,8 | 26,6 | 23,2 | 35,0 | 27,9 |
| Maret | 23,8 | 38,0 | 28,9 | 24,5 | 36,2 | 28,9 | 23,1 | 36,1 | 27,9 | 23,4 | 34,1 | 27,7 | 23,0 | 35,3 | 27,8 |
| April | 24,2 | 37,4 | 28,4 | 25,1 | 35,4 | 29,2 | 24,0 | 35,5 | 28,3 | 23,9 | 34,4 | 29,1 | 23,5 | 34,1 | 28,6 |
| Mei | 23,7 | 35,1 | 29,1 | 24,0 | 34,8 | 29,3 | 23,7 | 35,3 | 28,9 | 24,4 | 33,3 | 29,1 | 24,5 | 34,4 | 29,2 |
| Juni | 22,6 | 35,4 | 28,8 | 23,8 | 34,0 | 28,5 | 22,9 | 33,1 | 27,9 | 23,4 | 32,6 | 28,2 | 21,6 | 33,7 | 28,0 |
| Juli | 22,3 | 34,9 | 28,1 | 24,0 | 33,4 | 28,4 | 22,3 | 33,8 | 27,9 | 22,1 | 32,3 | 27,4 | 21,9 | 33,3 | 27,4 |
| Agustus | 22,1 | 34,8 | 27,7 | 23,5 | 35,6 | 28,7 | 22,9 | 34,5 | 28,1 | 21,2 | 32,7 | 27,4 | 23,2 | 32,3 | 27,6 |
| September | 23,0 | 36,6 | 28,5 | 23,6 | 36,8 | 29,2 | 23,3 | 36,4 | 28,8 | 22,9 | 35,6 | 28,7 | | | |
| Oktober | 23,3 | 37,1 | 29,1 | 23,6 | 37,6 | 28,7 | 23,9 | 37,1 | 29,8 | 24,1 | 37,4 | 29,9 | | | |
| Nopember | 23,9 | 38,4 | 30,6 | 23,9 | 38,1 | 28,7 | 24,0 | 37,0 | 28,2 | 23,7 | 37,5 | 29,4 | | | |
| Desember | 23,1 | 38,2 | 28,7 | 23,4 | 34,6 | 27,8 | 23,7 | 34,7 | 27,9 | 23,8 | 36,1 | 28,6 | | | |

Dari tabel 4.3 diketahui bahwa rata-rata suhu di Surabaya Pusat dalam rentang tahun 2015 sampai dengan 2019 dilihat dari pengukuran suhu di Taman Prestasi berkisar rata-rata antara 27 derajat *celcius* sampai dengan 30 derajat *celcius*. Suhu terendah terjadi pada bulan Februari 2018 yaitu 16,9 derajat *celcius*. Sedangkan suhu tertinggi yaitu bulan Nopember 2015 sebesar 38,4 derajat *celcius*. Gambar 4.6 menunjukkan pola kenaikan suhu tertinggi dihitung secara rata-rata per bulan terjadi pada bulan Maret, April, Mei, Oktober dan Nopember.




Gambar 4.6 Suhu Udara antara tahun 2015 sampai dengan 2019

Untuk melihat kondisi suhu udara di kota Surabaya secara umum, merujuk pada *website* resmi dari Pemerintah Kota Surabaya (www.surabaya.go.id), laporan Statistik Sektoral Kota Surabaya Tahun 2019[48] yang dirilis pada 13 Nopember 2019 menggunakan data dari Badan Meteorologi, Klimatologi dan Geofisika Kota Surabaya yang dipantau dari 3 lokasi stasiun pemantauan yaitu Perak I, Perak II dan Juanda. Dilaporkan bahwa dari keseluruhan data kelembaban udara di Kota Surabaya dalam tahun 2018, antara Perak I dan Perak II tidak jauh berbeda. Rata-rata kelembaban udara wilayah Perak I dan Perak II juga memiliki kesamaan yaitu 72% ini dikarenakan lokasi yang tidak terlalu jauh letaknya. Perak I memuliki rata-rata batas terbawah 51% dan Perak II memiliki rata-rata batas terbawah 49%. Lebih lanjut lagi, jika memperhatikan tingkat kelembaban udara di Perak dan Juanda, interval kelembaban udara di Juanda sedikit lebih besar dibandingkan di kawasan

Perak. Batas terbawah dan teratas kelembaban udara di kawasan Juanda lebih tinggi bila dibanding Perak. Rata-rata batas terbawah hasil pantauan stasiun Juada adalah 47,2% dan rata-rata batas atasnya adalah 93,1%. Namun demikian, rerata kelembaban udara setiap bulan di wilayah pantauan Juanda lebih tinggi jika dibandingkan dengan Perak. Hal Ini menunjukkan bahwa suhu udara di kawasan Juanda cenderung lebih rendah dibandingkan wilayah pantauan Perak. Hal ini juga tampak bahwa rata-rata suhu udara di kawasan pantauan Juanda adalah 28,1°C, sementara suhu udara di kawasan pantauan Perak I dan Perak II adalah sebesar 29°C dan 28,9° C.

Dari arah anginnya, tiga wilayah pantauan menunjukkan bahwa rerata angin di Surabaya mengarah ke Timur. Sementara kecepatan anginnya, di dua wilayah pantauan Perak berkisar di 3 knot dan 4 knot, dan di Juanda 7,8 knot. Sementara untuk arah angin dan kecepatan maksimalnya, wilayah pantauan Perak I dan Perak II ratarata kecepatan maksimal anginnya adalah 15 knot dan 27 knot, dan di juanda rata-rata sebesar 19,9 knot. Hal ini menunjukkan bahwa wilayah pantauan Perak I memiliki pergerakan angin yang lebih stabil dibandingkan wilayah pantauan Perak II dan Juanda.

## 4.4 Analisis Prediksi Menggunakan RNN-LSTM

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

*Tabel 4.2 dan pembagian data Training dan testing*

| Skenario | training : testing | Jumlah training | Jumlah testing | Data trining *(urutan data ke-)* | Data Testing *(urutan data ke-)* |
|---|---|---|---|---|---|
| I | 75% : 25% | 57.779 | 19.260 | 1 - 57.779 | 57.780 – 77.040 |
| II | 85% : 15% | 65.483 | 11.556 | 1 - 65.483 | 65.484 – 77.040 |
| III | 95%:5% | 73.187 | 3.852 | 1 – 73.187 | 73.188 – 77.040 |

Pembagian data *training* dan *testing* dibagi berdasarkan urutan *time series* yaitu dimulai dari data ke -1. Tabel 4.2 merupakan skenario pembagian data

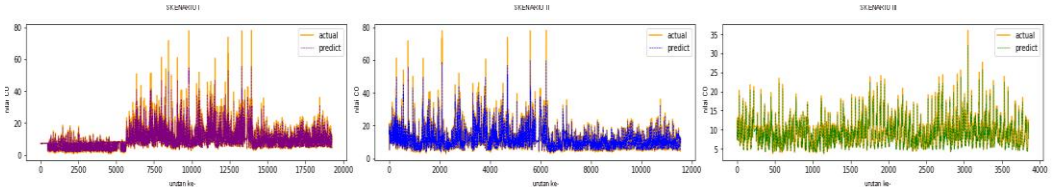
*training* dan *testing*. Dari hasil pengujian didapatkan hasil yang berbeda dilihat dari pembagian dataset *training* dan *testing*. Tabel 4.3 merupakan hasil pengujian dengan menggunakan beberapa komposisi jumlah *neuron hidden* dan pembagian data *training – testing*. Pertama, dengan komposisi data *training* 75% dan data *testing* 25% dengan pembagian data *training* 57.779 sampel dan 19.260 sampel untuk data *testing*, mendapatkan hasil terbaik pada jumlah *neuron hidden* 10 dengan nilai RMSE 2,74. Komposisi kedua adalah dengan komposisi data *training* 85% dan data *testing* 15% dengan pembagian data *training* 65.483 sampel dan data *testing* 11.556 sampel, mendapatkan hasil terbaik pada jumlah *neuron hidden* 100 dengan nilai RMSE 2,994. Komposisi ketiga dengan komposisi data *training* 95% dan data *testing* 5% dengan pembagian data *training* 73187 sampel dan 3852 sampel untuk data *testing*, mendapatkan hasil terbaik pada jumlah neuron 2 dengan nilai RMSE 1,880.

*Tabel 4.3 Kombinasi Pengujian berdasar jumlah neuron dan pembagian data Training dan testing*

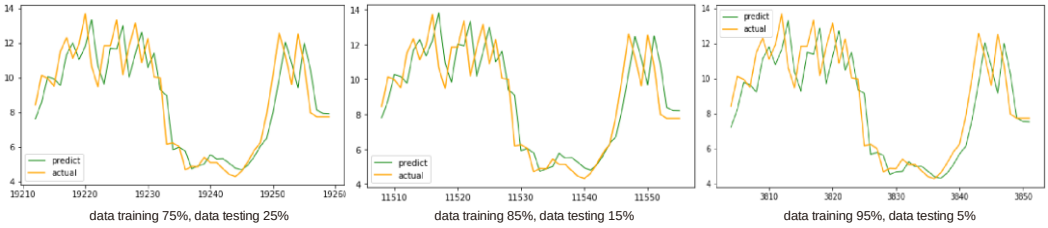
| Max Epoch | Neuron Hidden | 75% : 25% | 85% : 15% | 95%:5% |
|---|---|---|---|---|
| 50 | 2 | 2,884 | 3,108 | 1,880 |
| 50 | 5 | 2,827 | 3,114 | 1,911 |
| 50 | 10 | 2,740 | 3,097 | 1,900 |
| 50 | 15 | 2,796 | 3,042 | 1,904 |
| 50 | 20 | 2,804 | 3,056 | 1,915 |
| 50 | 50 | 2,791 | 3,013 | 1,895 |
|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

*Hidden neuron* (lapisan tersembunyi) 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 4.2. Skenario dengan data validasi tersebut kemudian

divisualisasikan dalam bentuk grafik seperti terlihat pada Gambar 4.6. Gambar 4.7 menunjukkan hasil komparasi antara nilai aktual dan prediksi untuk 48 data (dalam 1 hari). Gambar 4.7 menunjukkan bahwa model yang dibangun menggunakan algoritma RNN-LSTM dapat mempelajari tren harian dengan baik.




Gambar 4.6 Validasi Data Uji dengan Data Aktual pada 3 Skenario




Gambar 4.7. Grafik Komparasi Nilai Aktual dan Nilai Prediksi dalam Satu Hari

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

a. 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.
b. 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 Moniksida di Taman Prestasi.

c. Analisis Algoritma RNN-LSTM menunjukkan bahwa algoritma ini sangat bagus untuk membuat peramalan data polusi udara di kota Surabaya yang bersifat musimam dan secara umum model yang dibuat memberikan hasil yang baik dengan dilihat dari perhitungan nilai RMSE yang relatif kecil.

*Halaman ini sengaja dikosongkan*

# BAB 5
# 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 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.

*Halaman ini sengaja dikosongkan*

# DAFTAR PUSTAKA


[1] SK Menteri Kependudukan Lingkungan Hidup No 02/MENKLH/1988.

[2] Health Effects Institute. 2018. State of Global Air 2018. Special Report. Boston, MA:Health Effects Institut

[3] World health statistics 2018: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO

[4] Badan Pusat Statistik.Jumlah Kendaraan Bermotor in https://www.bps.go.id/linkTableDinamis/view/id/1133 diakses pada 15 April 2019

[5] Badan Pusat Statistik Kota Surabaya. Banyaknya Kendaraan Bermotor menurut Jenisnya 2009 – 2015 in https://surabayakota.bps.go.id/subject/17/transportasi.html#subjekViewTab4 diakses pada 15 April 2019

[6] Awal Syahrani (Nopember 2006).” Analisis Kinerja Mesin Bensin Berdasarkan Hasil Uji Emisi”SMARTek Vol 4, No 4 [Online] tersedia di http://jurnal.untad.ac.id/

[7] Caragliu, A. & Del Bo, C. & Nijkamp, P. 2009. Smart cities in Europe. Serie Research Memoranda 0048, VU University Amsterdam

[8] Benevolo C., Dameri R.P., D’Auria B. (2016) Smart Mobility in Smart City. In: Torre T., Braccini A., Spinelli R. (eds) Empowering Organizations. Lecture Notes in Information Systems and Organisation, vol 11. Springer, Cham

[9] Bhati, Abhishek & Hansen, Michael & Chan, Ching. (2017). Energy conservation through smart homes in a smart city: A lesson for Singapore households. Energy Policy. 104. 10.1016/j.enpol.2017.01.032.

[10] S. Suakanto, S. H. Supangkat, Suhardi and R. Saragih, "Smart city dashboard for integrating various data of sensor networks," International Conference on ICT for Smart Society, Jakarta, 2013, pp. 1-5, doi: 10.1109/ICTSS.2013.6588063

[11] N. Suri et al., "Exploiting smart city IoT for disaster recovery operations," 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), Singapore, 2018, pp. 458-463, doi: 10.1109/WF-IoT.2018.8355117

[12] Al Nuaimi, E., Al Neyadi, H., Mohamed, N. et al. Applications of big data to smart cities. J Internet Serv Appl 6, 25 (2015). https://doi.org/10.1186/s13174-015-0041-5

[13] Harish Kumar, Manoj Kumar Singh, M.P. Gupta, Jitendra Madaan. “Moving towards smart cities: Solutions that lead to the Smart City Transformation Framework”. Technological Forecasting & Social Change.2018

[14] Sumayang, Lalu. 2003. Dasar-Dasar Manajemen Produksi dan Operasi. Salemba Empat,Jakarta

[15] Supranto J. 2000. Statistik (Teori dan Aplikasi), Edisi Keenam. Jakarta: Erlangga

[16] Heizer, Jay dan Render, Barry. 2009. Manajemen Operasi, Buku 1 Edisi 9. Jakarta: Salemba Empat

[17] X. Zhang, M. H. Chen and Y. Qin, "NLP-QA Framework Based on LSTM-RNN," 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA), Changsha, 2018, pp. 307-311.[Online] diakses pada https://ieeexplore.ieee.org/

[18] S. Abujar, A. K. M. Masum, S. M. M. H. Chowdhury, M. Hasan and S. A. Hossain, "Bengali Text generation Using Bi-directional RNN," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-5.

[19] H. Xie, M. A. Bin Ahmadon, S. Yamaguchi and I. Toyoshima, "Random Sampling and Inductive Ability Evaluation of Word Embedding in Medical Literature," 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2019, pp. 1-4.

[20] S. Wang, P. Zhou, W. Chen, J. Jia and L. Xie, "Exploring RNN-Transducer for Chinese speech recognition," 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China, 2019, pp. 1364-1369.

[21] M. Jiang, Z. Yang and C. Zhao, "What to play next? A RNN-based music recommendation system," 2017 51st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, 2017, pp. 356-358.

[22] Y. Hu and S. Lin, "Deep Reinforcement Learning for Optimizing Finance Portfolio Management," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 14-20.

[23] M. Rozenwald, E. Khrameeva, G. Sapunov and M. Gelfand, "Prediction of 3D Chromatin Structure Using Recurrent Neural Networks," 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 2018, pp. 2488-2488.

[24] A. Xiao, J. Liu, Y. Li, Q. Song and N. Ge, "Two-phase rate adaptation strategy for improving real-time video QoE in mobile networks," in China Communications, vol. 15, no. 10, pp. 12-24, Oct. 2018.

[25] MWP Aldi, J Jondri, A Aditsania(Agustus 2018). “Analisis Dan Implementasi Long Short Term Memory Neural Network Untuk Prediksi Harga Bitcoin”. eProceedings of Engineering Vol 5, No 2.[Online] Tersedia di https://libraryeproceeding.telkomuniversity.ac.id/

[26] S. Taneja, N. Sharma, K. Oberoi and Y. Navoria, "Predicting trends in air pollution in Delhi using data mining," 2016 1st India International Conference on Information Processing (IICIP), Delhi, 2016, pp. 1-6.

[27] V Chaudhary, A Deshbhratar, V Kumar, D Paul (2018). “Time Series Based LSTM Model to Predict Air Pollutant's Concentration for Prominent Cities in India”.[Onlilne] Tersedia di https://www.semanticscholar.org/

[28] V Reddy, P Yedavalli, S Mohanty, U Nakhat (2018). “Deep Air: Forecasting Air Pollution in Beijing, China”.[Online] tersedia di https://www.ischool.berkeley.edu/

[29] Fardiaz, Srikandi. 1992. Pencemaran Air dan Udara. Yogyakarta: Penerbit Kanisius.

[30] Keputusan Menteri Negara Lingkungan Hidup Nomor : KEP 45 / MENLH / 1997 Tentang Indeks Standar Pencemar Udara

[31] Keputusan Kepala Badan Pengendalian Dampak Lingkungan No. 107 Tahun 1997 Tanggal 21 November 1997

[32] KBBI,2020. Kamus Besar Bahasa Indonesia (KBBI). [Online] tersedia di https://kbbi.kemdikbud.go.id/entri/meteorologi

[33] Soepangkat. 1994. Pengantar Meteorologi. Jakarta: BPLG.

[34] Gust.2020.American Meteorological Society[Online] tersedia di http://glossary.ametsoc.org/wiki/Gust

[35] KBBI,2020. Kamus Besar Bahasa Indonesia (KBBI). [Online] tersedia di https://kbbi.kemdikbud.go.id/entri/ radiasi global

[36] Wirjohamidjojo, S. & Y.S. Swarinoto. 2007.Praktek Meteorologi Pertanian. Jakarta:Badan Meteorologi dan Geofisika.

[37] Arsyad, L. 2001. Peramalan Bisnis Edisis Pertama. BPFE, Yogyakarta

[38] Hanke, J.E & Wichers,D.W.(20015). *Business Forecasting Eight Edition.* New Jersey: Pearson Prentice Hall

[39] Riyanto,Slamet & Hatmawan,A. Andhita.2020. Metode Riset Penelitian Kuantitatif Penelitian Di Bidang Manajemen, Teknik Pendidikan dan Eksperimen. Penerbit Deepublish:Yogyakarta [Online] tersedia di https://books.google.co.id/books?id=W2vXDwAAQBAJ

[40] Nasution, A. dan Zakaria, H. 2001. Metode Numerik dalam Ilmu Rekayasa Sipil. ITB, Bandung

[41] Munir, Rinaldi. 2006. Metode Numerik. Informatika Bandung : Bandung

[42] Sahid. 2005. Pengantar Komputasi Numerik dengan MATLAB. C.V Andi Offset :Yogyakarta

[43] Sumayang, Lalu. 2003. Dasar-Dasar Manajemen Produksi dan Operasi. Salemba Empat,Jakarta

[44] Kevin Gurney(1997). An introduction to neural networks.[Online] Tersedia pada : https://www.inf.ed.ac.uk/

[45] Yoshua Bengio, Patrice Simard, Pulo Fransconi dan Studen Member IEEE. “Learning Long-Term Dependencies with Gradient Descent is Difficult” dalam IEE Transaction on Neural Networks Vol 5 No.2 March 1994

[46] Makridakis, S. et al., (1982) "The Accuracy of Extrapolative (Time Series Methods): Results of a Forecasting Competition", Journal of Forecasting, Vol. 1, No. 2, pp. 111-153

[47] D.P. Kingma dan J. Ba, “Adam: A Method for Stochastic Optimization,” arXiv:1412.6980 [cs.LG], hal. 1–15, 2014.

[48] Pemeritah Kota Surabaya. Statistik Sektoral Kota Surabaya Tahun 2019. [Online] Tersedia pada : https://surabaya.go.id/id/berita/52759/statsekt2019

# LAMPIRAN

***Surat Permohonan Rekomendasi Penelitian***

KEMENTERIAN RISET, TEKNOLOGI, DAN PENDIDIKAN TINGGI
**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**
FAKULTAS TEKNOLOGI ELEKTRO
DEPARTEMEN TEKNIK ELEKTRO
Gedung B, C & AJ Kampus ITS Sukolilo, Surabaya 60111
Telp. (031) 5947302, 5994251-55 (Ext.1206, 1239) Fax. (031) 5931237
Email: elits@ee.its.ac.id ; www.ee.its.ac.id

Nomor : 8/828/76/IT2.VI.3.1/PN.01.00.00/2019 07 OCT 2019
Lampiran : -
Perihal : **Permohonan Ijin Penelitian Dan Bantuan Data**

Yth. : Kepala Badan Kesatuan **Bangsa** dan Perlindungan Masyarakat Pemkot Surabaya
**Jalan Jaksa Agung Suprapto No. 2-4 Surabaya**

Dengan hormat,

Dengan ini kami hadapkan mahasiswa Program Pascasarjana Departemen Teknik Elektro FTE-ITS:

| | |
|---|---|
| Nama | : Muh. Anas Faishol |
| NRP | : 07111850067005 |
| Program Studi | : Magister Teknik Elektro |
| Bidang Keahlian | : Telematika (PETIK) |

Mahasiswa tersebut sedang mengerjakan Tesis dengan judul:

***"Analisa Data Runtun Waktu untuk Prediksi Polusi Udara di Kota Surabaya Menggunakan Algoritma Recurrent Neural Network".***

Sehubungan dengan hal tersebut diatas, kami mohon bantuan agar mahasiswa di atas dapat di berikan kesempatan untuk melakukan penelitian/pencarian data dan informasi di instansi yang Bapak/Ibu Pimpin.
Untuk data yang diambil adalah:

- Data Series Pulusi Udara Kota Surabaya di Dinas Lingkungan Hidup

Demikian permohonan ini kami sampaikan, atas perhatian dan kerjasamanya kami mengucapkan terima kasih.

Kepala Departemen,

Dr.Eng. Ardyono Priyadi,ST., M.Eng.
NIP. 19730927 199803 1 004

*Halaman ini sengaja dikosongkan*

Surat Rekomendasi Penelitian dari Badan Kesbanglinmas Pemkot Surabaya

PEMERINTAH KOTA SURABAYA

**BADAN KESATUAN BANGSA, POLITIK DAN PERLINDUNGAN MASYARAKAT**

Jalan Jaksa Agung Suprapto Nomor 2 Surabaya 60272
Telepon (031) 5343000, (031) 5312144 Pesawat 112

Surabaya, 8 Oktober 2019

Kepada
Yth. Kepala Dinas Lingkungan Hidup Kota Surabaya
di –
SURABAYA

Nomor : 070/1167/436.8.5/2019
Lampiran : -
Hal : Penelitian.

**REKOMENDASI PENELITIAN**

Dasar :
1. Peraturan Menteri Dalam Negeri Nomor 64 Tahun 2011 Tentang Pedoman. Penerbitan Rekomendasi Penelitian, Sebagaimana Telah Diubah dengan Peraturan Menteri Dalam Negeri Nomor 7 Tahun 2014 tentang Perubahan Atas Peraturan Menteri Dalam Negeri Nomor 64 Tahun 2011 ;
2. Peraturan Walikota Surabaya Nomor 37 Tahun 2011 Tentang Rincian Tugas dan Fungsi Lembaga Teknis Daerah Kota Surabaya, Bagian Kedua Badan Kesatuan Bangsa, Politik dan Perlindungan Masyarakat.

Memperhatikan : Surat Kepala Departemen Teknik Elektro Fakultas Teknologi Elektro Institut Teknologi Sepuluh Nopember Surabaya Tanggal 07 Oktober 2019 Nomor : B/82876/IT2.VI.3.1/PN.00.00.00/2019 Perihal : Permohonan ijin Penelitian Dan bantuan Data

Plt. Kepala Badan Kesatuan Bangsa, Politik Dan Perlindungan Masyarakat Kota Surabaya memberikan rekomendasi kepada :

a. Nama : Muh Anas Faishol.
b. Alamat : Dsn Bangsri RT 04 RW 03 Ds Bangsri Kec Wonodadi Kab Blitar.
c. Pekerjaan/Jabatan : Mahasiswa.
d. Instansi/Organisasi : Institut Teknologi Sepuluh Nopember Surabaya.
e. Kewarganegaraan : Indonesia.

Untuk melakukan penelitian/survey/kegiatan dengan :

a. Judul / Thema : Analisis Data Rruntun Waktu Untuk Prediksi Polusi Udara Di Kota Surabaya Menggunakan Algoritma RNN.
b. Tujuan : Penelitian.
c. Bidang Penelitian : Lingkungan.
d. Penanggung Jawab : Dr. Ir. Endroyono, DEA.
e. Anggota Peserta : -.
f. Waktu : 3 (Tiga) Bulan, TMT Surat Dikeluarkan.
g. Lokasi : Dinas Lingkungan Hidup Kota Surabaya.

Dengan persyaratan :
1. Penelitian/survey/kegiatan yang dilakukan harus sesuai dengan surat permohonan dan wajib mentaati persyaratan/peraturan yang berlaku di Lokasi/Tempat dilakukan Penelitian/survey/kegiatan;
2. Saudara yang bersangkutan agar setelah melakukan Penelitian/survey/kegiatan wajib melaporkan pelaksanaan dan hasilnya kepada Kepala Bakesbang, Politik dan Linmas Kota Surabaya;
3. Penelitian/survey/kegiatan yang dilaksanakan tidak boleh menimbulkan keresahan dimasyarakat, disintegrasi bangsa atau mengganggu keutuhan NKRI;
4. Rekomendasi ini akan dicabut/tidak berlaku apabila yang bersangkutan tidak memenuhi persyaratan seperti tersebut diatas.

Demikian atas bantuannya disampaikan terima kasih.

a.n. Plt. KEPALA BADAN
Plt. Sekretaris,

Ir. Yusuf Masruh, M.M.
Pembina
NIP 19671224 199412 1 001

Tembusan :
Yth. 1. Kepala Departemen Teknik Elektro Fakultas Teknologi Elektro Institut Teknologi Sepuluh Nopember Surabaya
2. Saudara yang bersangkutan.

*Halaman ini sengaja dikosongkan*

Source Code Interpolasi Data

```
#!/usr/bin/env python
# coding: utf-8

# In[40]:


from pandas import read_csv
from pandas import datetime
from pandas import DataFrame
from matplotlib import pyplot
import numpy as np
import pandas as pd


# In[41]:


mini_tamanprestasi=read_csv('data2019tamanprestasi_notime.csv', header=0,
sep=',')


# In[42]:


#mini taman prestasi
mini_tamanprestasi.replace('---',np.nan, inplace=True)
mini_tamanprestasi.replace(0,np.nan, inplace=True)
mini_tamanprestasi=mini_tamanprestasi.astype(float)


# In[43]:


print('-----------------------------------------------------------------------------------------
---------------------')
print('mini taman prestasi')
print(mini_tamanprestasi.isnull().sum())


# mini_tamanprestasi=mini_tamanprestasi.reset_index()

# In[38]:


mini_tamanprestasi.info()
```

```
# In[39]:


mini_tamanprestasi


# In[11]:


mini_tamanprestasi.info()


# In[13]:


data_interpolated=DataFrame()
print(data_interpolated)


# In[26]:


mini_tamanprestasi['CO_ISPU']
taman=mini_tamanprestasi[['DD','DD_Boe','FF','FF_Boe','GRAD','HUMair','T
EMPair']]
taman.replace(0,np.nan, inplace=True)


# In[27]:


header=taman.head()

i=0
for kolom in header:
    print(kolom)
    data = mini_tamanprestasi[kolom].interpolate(method='spline', order=2)
    data_interpolated[kolom]=data.values
    print(i)
    i=i+1
```

```
#
data_interpolated['CO_ISPU']=data_interpolated['CO_ISPU'].replace(data_inter
polated['CO_ISPU']<0,0, inplace=True)

#
data_interpolated=data_interpolated[['CO_ISPU','DD','DD_Boe','FF','FF_Boe','
GRAD','HUMair','TEMPair']]

# In[28]:


data_interpolated['CO']=mini_tamanprestasi['CO_ISPU']
data_interpolated=data_interpolated[['CO','DD','DD_Boe','FF','FF_Boe','GRAD'
,'HUMair','TEMPair']]


# In[29]:


data_interpolated


# In[30]:


data_interpolated.replace(0,np.nan, inplace=True)
print(data_interpolated.isnull().sum())


# In[31]:

data_interpolated.replace(np.nan,0, inplace=True)


# In[32]:


data_interpolated.to_csv('dataset_paper_interpolated2019.csv')

interpolated = mini_tamanprestasi['CO'].interpolate(method='linear', order=2)
interpolated2 = mini_tamanprestasi['CO'].interpolate(method='spline', order=2)
#print(interpolated)


interpolated2.plot(figsize=(20,5), label='spline')
pyplot.legend()
pyplot.show()interpolated.isnull().sum()print(interpolated2.values)
```

*Halaman ini sengaja dikosongkan*

Source Code RNN-LSTM

```
#!/usr/bin/env python
# coding: utf-8

# In[15]:


from math import sqrt
from numpy import concatenate
from matplotlib import pyplot
from pandas import read_csv
from pandas import DataFrame
from pandas import concat
from pandas import timedelta_range
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import mean_squared_error
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras . layers import Dropout
from keras.callbacks import EarlyStopping
from sklearn.metrics import accuracy_score


# In[16]:


# convert series to supervised learning
def series_to_supervised(data, n_in=1, n_out=1, dropnan=True):
	n_vars = 1 if type(data) is list else data.shape[1]
	df = DataFrame(data)
	cols, names = list(), list()
	# input sequence (t-n, ... t-1)
	for i in range(n_in, 0, -1):
		cols.append(df.shift(i))
		names += [('var%d(t-%d)' % (j+1, i)) for j in range(n_vars)]
	# forecast sequence (t, t+1, ... t+n)
	for i in range(0, n_out):
		cols.append(df.shift(-i))
		if i == 0:
			names += [('var%d(t)' % (j+1)) for j in range(n_vars)]
		else:
			names += [('var%d(t+%d)' % (j+1, i)) for j in 
range(n_vars)]
	# put it all together
```

```
        agg = concat(cols, axis=1)
        agg.columns = names
        # drop rows with NaN values
        if dropnan:
                agg.dropna(inplace=True)
        return agg



# In[17]:


# load dataset
dataset = read_csv('dataset_ispu_interpolated2-Copy1.csv', header=0)
values = dataset.values


# In[18]:


dataset

# integer encode direction
encoder = LabelEncoder()
values[:,4] = encoder.fit_transform(values[:,4])
# In[19]:


# ensure all data is float
values = values.astype('float32')


# In[20]:


# normalize features
scaler = MinMaxScaler(feature_range=(0, 1))
scaled = scaler.fit_transform(values)


# In[21]:


# frame as supervised learning
reframed = series_to_supervised(scaled, 1, 1)
```

```
# In[22]:


reframed.head()


# In[23]:


# drop columns we don't want to predict
reframed.drop(reframed.columns[[9,10,11,12,13,14,15]], axis=1, inplace=True)
print(reframed.head())


# In[24]:


# split into train and test sets
values = reframed.values
n_train_hours = (int)(0.95 * (len(values)))  #365 * 24
train = values[:n_train_hours, :]
test = values[n_train_hours:, :]


# In[25]:


print('train',len(train),'test',len(test))


# In[26]:


# split into input and outputs
train_X, train_y = train[:, :-1], train[:, -1]
test_X, test_y = test[:, :-1], test[:, -1]


# In[27]:


# reshape input to be 3D [samples, timesteps, features]
train_X = train_X.reshape((train_X.shape[0], 1, train_X.shape[1]))
test_X = test_X.reshape((test_X.shape[0], 1, test_X.shape[1]))
print(train_X.shape, train_y.shape, test_X.shape, test_y.shape)
```

```
# In[28]:


# design network
model = Sequential()
model.add(LSTM(2, input_shape=(train_X.shape[1], train_X.shape[2])))
model.add(Dense(1))
model.compile(loss='mae', optimizer='adam')
#model.add(Dropout(1))
# simple early stopping
es = EarlyStopping(monitor='val_loss', mode='min', verbose=1)


# In[29]:


# fit network
history = model.fit(train_X, train_y, epochs=50, batch_size=150,
validation_data=(test_X, test_y), verbose=2, shuffle=False, callbacks=[es])
# plot history
pyplot.plot(history.history['loss'], label='train')
pyplot.plot(history.history['val_loss'], label='test')
pyplot.legend()
pyplot.show()

# plot history
pyplot.plot(history.history['loss'], label='train')
pyplot.plot(history.history['val_loss'], label='test')
pyplot.rcParams['figure.figsize'] = [7, 4]

pyplot.legend()
pyplot.show()
# In[114]:


# make a prediction
yhat = model.predict(test_X)
test_X = test_X.reshape((test_X.shape[0], test_X.shape[2]))


# In[99]:


# invert scaling for forecast
inv_yhat = concatenate((yhat, test_X[:, 1:]), axis=1)
inv_yhat = scaler.inverse_transform(inv_yhat)
inv_yhat = inv_yhat[:,0]
```

```
# In[112]:


inv_yhat=inv_yhat[0:3852]
print(len(inv_yhat))


# In[113]:


print(inv_y)


# In[109]:


# invert scaling for actual
test_y = test_y.reshape((len(test_y), 1))
inv_y = concatenate((test_y, test_X[:, 1:]), axis=1)
inv_y = scaler.inverse_transform(inv_y)
inv_y = inv_y[:,0]


# In[56]:


# calculate RMSE
rmse = sqrt(mean_squared_error(inv_y, inv_yhat))
print('Test RMSE: %.3f' % rmse)


# In[57]:


model.save('model95.h5')


# In[188]:


inv_yhat=DataFrame(inv_yhat)
print(inv_yhat)


# In[189]:
```

```
inv_y=DataFrame(inv_y)
print(inv_y)


# In[192]:


x=timedelta_range(0, periods=48, freq='30T').astype(str)
pyplot.plot(inv_yhat.tail(48), label='predict', color='green', linewidth=1,
linestyle="-" )
pyplot.plot(inv_y.tail(48), label='actual', color='orange')
pyplot.rcParams['figure.figsize'] = [5, 4]
pyplot.legend()
pyplot.show()



# In[ ]:
```

# BIODATA PENULIS

Penulis dilahirkan di Blitar pada tanggal 02 Agustus 1987, merupakan anak kedua dari tiga bersaudara dari pasangan Bapak Muh. Amin dan Ibu Mudrikah dengan nama lengkap Muh. Anas Faishol. Pendidikan formal yang pernah ditempuh penulis adalah MI Darul Ulum (1994 – 2000), MTsN Kunir (2000 – 2003), MAN Kota Blitar (2003 – 2006). Kemudian penulis melanjutkan di perguruan tinggi pada jurusan Teknik Informatika di Universitas Islam Negeri Maulana Malik Ibrahim Malang masuk pada tahun 2006 dan lulus pada tahun 2011. Setelah lulus S1, penulis bekerja pada perusahaaan swasta yaitu Bank Danamon dan Telkomsel sampai dengan tahun 2015. Sejaka tahun 2015 penulis diangkat menjadi ASN pada Pemerintah Kabupaten Blitar dan aktif di Dinas Komunikasi dan Informatika Kabupaten Blitar. Pada tahun 2018 melalui beasiswa PETIK dari Kementrian Komunikasi dan Informatika RI, penulis mendapatkan kesempatan melanjutkan jenjang pendidikan pasca sarjana di bidang telematika Departemen Teknik Elektro, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember Surabaya.