

PREDIKSI JUMLAH PENDUDUK TINGKAT KECAMATAN DI WILAYAH BOGOR MENGUNAKAN METODE LONG SHORT TERM MEMORY

Owen¹, Dyah Erny Herwindiati², Teny Handayani³

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

E-mail: ¹owen.535180016@stu.untar.ac.id, ²dyahh@fti.untar.ac.id, ³tenyh@fti.untar.ac.id

Abstrak

Pertumbuhan penduduk merupakan penambahan atau pengurangan jumlah penduduk yang dipengaruhi beberapa faktor. Di Indonesia hal ini merupakan sesuatu yang sangat diperhatikan dan menjadi pantauan pemerintah khususnya di pulau Jawa. Kekhawatiran terjadinya lonjakan penduduk menjadi salah satu alasan pemantauan ini yang di mana dapat menyebabkan permasalahan pada daya dukung dan daya tampung lingkungan hidup. Tujuan dari perancangan ini adalah membuat prediksi jumlah penduduk dan perhitungan pertumbuhan penduduk pada tingkat kecamatan di wilayah Bogor untuk tahun 2021 dan tahun 2022 menggunakan data kependudukan dengan interval tahunan yang berbeda pada setiap wilayah. Prediksi dilakukan menggunakan metode Long Short Term Memory. Konfigurasi parameter model yang digunakan untuk melakukan pelatihan dan pengujian berbeda-beda untuk setiap wilayah yang diperoleh dari hasil eksperimen parameter yang diulang sebanyak 5 kali untuk setiap konfigurasi sehingga memperoleh rata-rata Mean Absolute Percentage Error (MAPE) terbaik. Semua model untuk metode LSTM menghasilkan rata-rata MAPE di bawah 10% pada setiap wilayah sehingga model prediksi dinyatakan sangat baik.

Kata kunci—Penduduk, Long Short Term Memory, Mean Absolute Percentage Error, Prediksi

Abstract

Population growth is addition or reduction of the population which is influenced by several factors. In Indonesia, this is something that pays great attention and is monitored by the government, especially on Java Island. Worries of population increase is one of the reasons for this monitoring which can cause problems with the support power and capacity power of the environment. The purpose of this design is to predict the population and calculate population growth rate at sub-district level in the Bogor area for 2021 and 2022 using population data at different annual intervals in each areas. Prediction is done using Long Short Term Memory. The configuration parameters of the model used for training and testing is different for each areas which obtained from the results of the parameter experiment which was repeated 5 times for each configuration to obtain the best Mean Absolute Percentage Error (MAPE) average. All models for LSTM method gain an average MAPE below 10% in each areas so that the models for prediction were stated to be very good.

Keywords—Population, Long Short Term Memory, Mean Absolute Percentage Error, Predict

1. PENDAHULUAN

Penduduk adalah sejumlah atau sekelompok orang yang mendiami suatu daerah tertentu. Indonesia yang merupakan negara keempat dengan jumlah penduduk terbanyak di dunia, memiliki beberapa pulau besar salah satunya adalah pulau Jawa yang dihuni oleh 151,59 juta penduduk atau 56,10 persen penduduk Indonesia [1]. Padatnya jumlah penduduk di pulau Jawa menjadi salah satu alasan menjadi pantauan pemerintah yang dikhawatirkan akan terjadinya lonjakan jumlah penduduk. Pada dasarnya perubahan pertumbuhan penduduk di suatu wilayah dapat disebabkan oleh beberapa faktor seperti kelahiran, kematian, dan migrasi.

Selain lonjakan jumlah penduduk yang bisa terjadi karena kepadatan penduduk, akibat lain dari permasalahan ini adalah daya dukung dan daya tampung lingkungan hidup yang semakin menurun. Sensus penduduk yang dilakukan setiap 10 tahun sekali oleh pihak instansi pemerintahan seperti Badan Pusat Statistik memiliki tujuan untuk pendataan beserta dengan membantu dalam melakukan proyeksi jumlah penduduk untuk beberapa tahun kedepan. Oleh karena itu kebutuhan akan informasi prediksi jumlah penduduk sangat dibutuhkan sebagai sebuah informasi yang bisa membantu perencanaan dan pengambilan keputusan dalam tujuan menjadi suatu tindakan antisipasi dari pemerintahan, walaupun hasil prediksi yang dihasilkan masih bersifat bias.

Oleh sebab itu, melalui jurnal ini, dibuatlah aplikasi yang dapat memprediksi jumlah penduduk dan menghitung pertumbuhan penduduk (*growth rate*) pada tingkat kecamatan di wilayah Bogor dengan menggunakan metode *Long Short Term Memory*. Data yang digunakan dalam melakukan prediksi pada awalnya berasal dari file Microsoft Excel berekstensi *Comma Separated Value* (CSV) yang diunggah ke dalam aplikasi dengan format yang sudah ditetapkan. Aplikasi yang dibuat akan terdiri dari beberapa modul yang memiliki kegunaannya masing-masing dalam membantu proses prediksi.

2. METODE PENELITIAN

2.1. Metode Perancangan

Metode SDLC (*Software Development Life Cycle*) merupakan proses pembuatan dan pengubahan sistem serta model dan sebuah metode yang digunakan untuk mengembangkan suatu aplikasi [2]. Metode ini terdiri dari beberapa langkah penerapan yaitu *Planning* (Perencanaan), *Analysis* (Analisis), *Design* (Rancangan/Desain), dan *Implementation* (Pembuatan/Implementasi). Metode inilah yang akan digunakan dalam proses perancangan aplikasi prediksi jumlah penduduk.

2.1.1. *Planning* (Perencanaan)

Melakukan perencanaan rancangan aplikasi yang akan dibuat dengan maksud untuk menentukan masalah dan ruang lingkup sistem yang ada, serta menentukan tujuan pembuatan aplikasi. Perencanaan sangat penting dilakukan Ketika melakukan pembuatan atau perancangan suatu proyek atau aplikasi agar tujuan perancangan dapat tercapai.

2.1.2. *Analysis* (Analisis)

Pengumpulan informasi secara mendetail mengenai spesifikasi yang diperlukan untuk sistem, baik yang berhubungan dengan perangkat lunak (*software*) maupun perangkat keras (*hardware*), serta mengumpulkan data-data yang diperlukan di dalam aplikasi yang dirancang

2.1.3. Design (Desain/Rancangan)

Pada tahap ini, akan dilakukan penggambaran mengenai aplikasi yang dirancang baik dari sisi diagram, relasi antar struktur maupun rancangan antar muka aplikasi yang akan dibuat. Dengan penggambaran ini diharapkan pengguna maupun pembaca akan memahami alur sebenarnya dari aplikasi prediksi ini.

2.1.4. Implementation (Implementasi/Pembuatan)

Tahapan selanjutnya adalah mengimplementasikan apa yang ditentukan pada tahap desain, di antaranya adalah membuat tampilan antar muka dan membuat sistem aplikasi yang nantinya kedua hal tersebut akan digabungkan menjadi sebuah aplikasi yang bisa digunakan oleh pengguna untuk memprediksi jumlah penduduk.

2.2. Polinom Interpolasi Lagrange

Metode ini adalah metode yang memiliki bentuk hampir sama dengan polinom interpolasi Newton, tetapi tidak menggunakan bentuk pembagian beda hingga. Interpolasi Lagrange diterapkan untuk mendapatkan fungsi polinomial berderajat tertentu yang dapat digunakan untuk mencari dan menggantikan data yang hilang (*missing value*) pada sebuah kumpulan data salah satu contohnya adalah data deret waktu (*time series*).

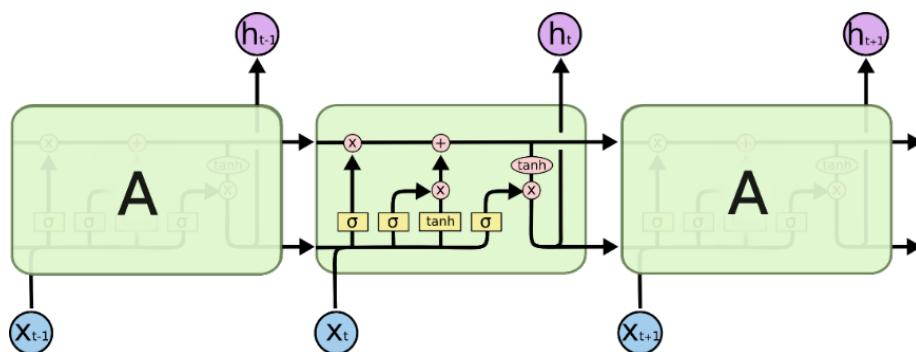
Pada kumpulan data mentah jumlah penduduk ditemukan banyaknya *missing value* yang tidak beraturan sehingga digunakanlah Interpolasi Lagrange orde tiga karena semakin tinggi orde interpolasi yang digunakan dapat meningkatkan kemungkinan mencapai akurasi terbaik [3]. Rumus Interpolasi Lagrange orde tiga adalah sebagai berikut:

$$f_3(x) = \sum_{i=0}^{n=3} L_i(x)f(x_i) = L_0(x)f(x_0) + L_1(x)f(x_1) + L_2(x)f(x_2) + L_3(x)f(x_3) \quad (1)$$

$$L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j} \quad (2)$$

2.3. Long Short-Term Memory

Long Short-Term Memory (LSTM) adalah jenis *Recurrent Neural Network* (RNN) yang implementasinya untuk mengatasi permasalahan pada data urutan (*sequence data*) serta menangani permasalahan *long-term dependencies* yang tidak bisa diselesaikan menggunakan RNN [4]. Variasi model LSTM yang digunakan untuk permodelan prediksi jumlah penduduk adalah Vanilla LSTM dengan memiliki 3 lapisan, yaitu sebuah *input layer*, *hidden layer*, dan *output layer*. Bentuk jaringan LSTM dapat dilihat pada Gambar 1.



Gambar 1 Bentuk Jaringan Long Short Term Memory (LSTM)

Sumber: Christopher Colah, [Understanding LSTM Networks](https://colah.github.io/posts/2015-08-Understanding-LSTMs/), <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>, 09 September 2021.

Di dalam metode *Deep Learning Long Short Term Memory* (LSTM) memiliki beberapa tahapan diantaranya adalah *forget gate*, *input gate*, *cell state*, dan *output gate*.

2.4. Preprocessing

Preprocessing adalah proses transformasi data mentah sebelum diolah atau digunakan di dalam algoritma *Machine Learning* atau *Deep Learning* [5]. Pada pembuatan model prediksi LSTM tahapan-tahapan yang dilakukan pada *preprocessing*, yaitu:

1. Mengumpulkan data-data yang diperlukan dalam permodelan dan prediksi. Data yang dibutuhkan dalam perancangan ini berekstensi .csv dengan ketentuan satu file untuk setiap kecamatan di setiap wilayah dan memiliki format tabel “Tahun” dan “Populasi”.
2. Mengunduh pustaka (*library*) bahasa pemrograman *Python* yang membantu *data preprocessing* dan permodelan metode LSTM.
3. Melakukan normalisasi untuk mengubah kisaran nilai (*value range*) pada data agar bisa sejajar. Berdasarkan data yang digunakan yaitu jumlah penduduk bisa diketahui bahwa data tersebut berjenis univariat sehingga tidak perlu dilakukan normalisasi.
4. Membagi kumpulan data menjadi 70% data pelatihan (*training data*) dan 30% data pengujian (*testing data*).
5. Melakukan pembagian masing-masing data pelatihan dan data pengujian menjadi model urutan (*sequence*) menggunakan variabel *timesteps* yang diperoleh dari input manual pengguna aplikasi prediksi.
6. Melakukan proses *reshaping* untuk membentuk kembali data array yang terdiri dari [*samples, features*] menjadi [*samples, timesteps, features*] sehingga data array dapat digunakan dalam permodelan metode LSTM.

2.5. Mean Absolute Percentage Error

Mean Absolute Percentage Error (MAPE) adalah ukuran kesalahan relatif yang dimana menyatakan persentase kesalahan hasil prediksi terhadap hasil aktual selama periode tertentu yang akan memberikan informasi akhir dengan ditentukan oleh tinggi atau rendahnya persentase kesalahan [6]. Rumus MAPE adalah sebagai berikut:

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \times 100\%}{n} \quad (3)$$

Keterangan:

- MAPE = *Mean Absolute Percentage Error*
- n = Jumlah Data
- A_t = Data Aktual ke t
- F_t = Hasil Peramalan ke t

Pada saat evaluasi model prediksi, nilai MAPE yang akan dijadikan sebagai penentu keakuratan model adalah hasil MAPE dari prediksi data pengujian (*testing data*) yang dimana pengelompokkan persentase penilaian MAPE dapat dilihat pada Tabel 1.

Tabel 1 Penilaian MAPE

Range MAPE	Arti Nilai
< 10%	Kemampuan model peramalan sangat baik.
10 – 20%	Kemampuan model peramalan baik.
20 – 50%	Kemampuan model peramalan layak.
>50%	Kemampuan model peramalan buruk.

2.6. Pembuatan Model

Sebelum data jumlah penduduk digunakan untuk prediksi, perlu dilakukan eksperimen yang bertujuan memperoleh konfigurasi parameter paling baik untuk permodelan di Kota Bogor dan Kabupaten Bogor. Parameter-parameter yang diperlukan yaitu *learning rate*, neuron, epoch, dan *timesteps* atau *window size*.

Pada eksperimen pertama dilakukan penentuan nilai *learning rate* terbaik menggunakan 3 nilai konfigurasi yaitu 0.001, 0.005, dan 0.01 serta digabungkan dengan nilai konfigurasi parameter lain yaitu 64 neuron, 3 *timesteps*, dan 100 epoch. Rancangan konfigurasi penentuan nilai *learning rate* dapat dilihat pada Tabel 2.

Tabel 2 Rancangan Konfigurasi Eksperimen Penentuan *Learning Rate*

No	Konfigurasi	Learning Rate	Neuron	Epochs	Timesteps
1	Konfigurasi 1	0.001	64	100	3
2	Konfigurasi 2	0.005	64	100	3
3	Konfigurasi 3	0.01	64	100	3

Pada eksperimen kedua dilakukan penentuan nilai neuron terbaik menggunakan 3 nilai konfigurasi yaitu 16, 32, dan 100 serta digabungkan dengan nilai konfigurasi parameter lain yaitu X *learning rate*, 3 *timesteps*, dan 100 epoch. Rancangan konfigurasi penentuan nilai neuron dapat dilihat pada Tabel 3.

Tabel 3 Rancangan Konfigurasi Eksperimen Penentuan Neuron

No	Konfigurasi	Learning Rate	Neuron	Epochs	Timesteps
1	Konfigurasi 1	X	16	100	3
2	Konfigurasi 2	X	32	100	3
3	Konfigurasi 3	X	64	100	3

Pada eksperimen ketiga dilakukan penentuan nilai epoch terbaik menggunakan 3 nilai konfigurasi yaitu 50, 75, dan 100 serta digabungkan dengan nilai konfigurasi parameter lain yaitu *learning rate* eksperimen pertama, 3 *timesteps*, dan X neuron. Rancangan konfigurasi penentuan nilai epoch dapat dilihat pada Tabel 4.

Tabel 4 Rancangan Konfigurasi Eksperimen Penentuan Epoch

No	Konfigurasi	Learning Rate	Neuron	Epochs	Timesteps
1	Konfigurasi 1	X	X	50	3
2	Konfigurasi 2	X	X	75	3
3	Konfigurasi 3	X	X	100	3

Pada eksperimen keempat dilakukan penentuan nilai *timesteps* atau *window size* terbaik menggunakan 2 nilai konfigurasi yaitu 2 dan 3 serta digabungkan dengan nilai konfigurasi parameter *learning rate*, neuron, dan epoch yang diperoleh dari hasil eksperimen pertama sampai eksperimen ketiga. Rancangan konfigurasi penentuan nilai epoch dapat dilihat pada Tabel 5.

Tabel 5 Rancangan Konfigurasi Eksperimen Penentuan *Timesteps*

No	Konfigurasi	Learning Rate	Neuron	Epoch	Timesteps
1	Konfigurasi 1	X	X	X	2
2	Konfigurasi 2	X	X	X	3

Semua rancangan konfigurasi eksperimen di atas akan digunakan pada 1 kecamatan untuk masing-masing wilayah sebagai penentu konfigurasi parameter terbaik pada data wilayah Bogor, Depok, dan Tangerang Selatan serta untuk setiap konfigurasi akan dilakukan sebanyak 5 kali pengulangan untuk diambil rata-rata nilai MAPE pengujian terbaik.

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Hasil dari penelitian ini akan disajikan ke dalam beberapa bagian uraian yang terdiri dari data yang diperoleh dan digunakan, nilai konfigurasi parameter terbaik, dan rata-rata MAPE pengujian hasil prediksi.

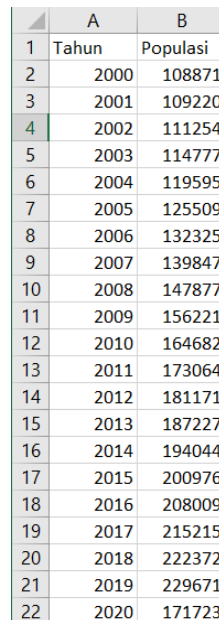
3.1.1. Data

Di dalam penelitian ini wilayah yang digunakan untuk permodelan dan dilakukan prediksi adalah Kota Bogor dan Kabupaten Bogor. Untuk setiap wilayah jumlah data kependudukan yang diperoleh berbeda-beda dan memiliki interval tahunan. Kelengkapan data kependudukan setiap kecamatan pada wilayah Bogor akan ditampilkan pada Tabel 6.

Tabel 6 Perolehan Data Kependudukan Setiap Wilayah

Wilayah	Data Peroleh (Tahun)
Kota Bogor	2004 sampai 2020
Kabupaten Bogor	2000 sampai 2020

Data yang diperoleh akan disimpan dalam file berekstensi .csv dengan format seperti pada Gambar 2.

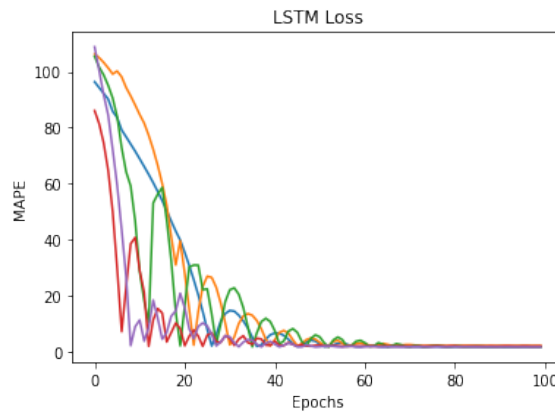


	A	B
1	Tahun	Populasi
2	2000	108871
3	2001	109220
4	2002	111254
5	2003	114777
6	2004	119595
7	2005	125509
8	2006	132325
9	2007	139847
10	2008	147877
11	2009	156221
12	2010	164682
13	2011	173064
14	2012	181171
15	2013	187227
16	2014	194044
17	2015	200976
18	2016	208009
19	2017	215215
20	2018	222372
21	2019	229671
22	2020	171723

Gambar 2 Format Data Kependudukan pada File Microsoft Excel

3.1.2. Konfigurasi Parameter Terbaik Setiap Wilayah

Pada eksperimen pertama diperoleh konfigurasi nilai *learning rate* terbaik yaitu 0.005 setelah melalui 5 kali pengulangan untuk metode LSTM. Grafik laju pembelajaran dan perubahan loss metode LSTM dapat dilihat pada Gambar 3.



Gambar 3 Grafik Kurva Perubahan Loss LSTM *Learning Rate* 0.005

Setelah memperoleh nilai *learning rate* yang akan digunakan pada semua permodelan dan prediksi di Kota Bogor dan Kabupaten Bogor, dilanjutkan dengan hasil akhir eksperimen permodelan untuk neuron, epoch, dan *timesteps* yang dapat dilihat pada Tabel 7.

Tabel 7 Hasil Akhir Eksperimen Penentuan Konfigurasi Parameter LSTM Terbaik

Wilayah	Neuron	Epoch	Timesteps
	LSTM	LSTM	LSTM
Kota Bogor	64	100	3
Kabupaten Bogor	64	75	3

Menggunakan nilai-nilai konfigurasi parameter di atas, prediksi jumlah penduduk dapat dilakukan pada tingkat kecamatan di wilayah Bogor untuk tahun 2021 dan tahun 2022 menggunakan metode LSTM. Setelah keseluruhan hasil prediksi jumlah penduduk dan perhitungan pertumbuhan penduduk (*growth rate*) diperoleh hal selanjutnya yang dilakukan adalah melakukan evaluasi dengan merata-ratakan nilai MAPE pengujian dari setiap hasil prediksi jumlah penduduk. Rata-rata MAPE pengujian metode LSTM secara keseluruhan dapat dilihat pada Tabel 8.

Tabel 8 Rata-Rata MAPE LSTM Hasil Prediksi Jumlah Penduduk

No	Wilayah	Rata-Rata MAPE LSTM Pengujian
1	Kota Bogor	3.246
2	Kabupaten Bogor	2.53765

3.2. Pembahasan

Hasil sistem prediksi jumlah penduduk pada tingkat kecamatan di wilayah Bogor yang dibuat berhasil memperoleh rata-rata nilai MAPE pengujian LSTM sebesar 3.25% untuk Kota Bogor, serta 2.54% untuk Kabupaten Bogor.

4. KESIMPULAN

Beberapa kesimpulan yang diperoleh dari sistem prediksi jumlah penduduk tingkat kecamatan di wilayah Bogor menggunakan metode LSTM yaitu nilai konfigurasi parameter yang diperoleh untuk prediksi di Kota Bogor dan Kabupaten Bogor berbeda-beda. Kelebihan dari hasil

penelitian ini adalah rata-rata nilai MAPE pengujian yang diperoleh dari hasil prediksi secara keseluruhan memiliki persentase dibawah 10% sehingga model yang digunakan untuk pengujian sangat baik. Kekurangan dari hasil penelitian ini adalah masih ditemukannya *overfitting* pada saat permodelan LSTM dikarenakan jumlah data yang digunakan sangat sedikit.

Kemungkinan pengembangan selanjutnya yang dapat dilakukan adalah mencoba salah satu metode yaitu *k-fold cross validation* untuk membantu mengatasi permasalahan data terlalu sedikit yang menyebabkan *overfitting* dan melakukan eksperimen kembali menggunakan berbagai berbagai nilai konfigurasi parameter untuk menghasilkan model prediksi yang lebih baik.

UCAPAN TERIMA KASIH

Terima kasih kepada Prof. Dr. Ir. Dyah Erny Herwindiati, M.Si dan Teny Handayani, PhD yang telah memberikan saran kepada penulis dapat menyelesaikan jurnal penelitian ini. Jurnal Computatio berterima kasih kepada IJCCS sebagai landasan bentuk format makalah ini.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, 2021, Hasil Sensus Penduduk 2020, <https://www.bps.go.id/pressrelease/2021/01/21/1854/hasil-sensus-penduduk-2020.html>, diakses tanggal 28 Desember 2021.
- [2] Setiawan, Rony, 2021, Metode SDLC Dalam Pengembangan Software, <https://www.dicoding.com/blog/metode-sdlc/>, diakses tanggal 28 Desember 2021.
- [3] Chapra, Steven, 2015, *Numerical Methods for Engineers*, Seventh Edition, McGraw-Hill Education, New York.
- [4] Suganda, Abba Girsang, 2019, *Long Short Term Memory (LSTM)*, <https://mti.binus.ac.id/2019/12/02/long-short-term-memory-lstm/>, diakses tanggal 28 Desember 2021.
- [5] Hadrienj, 2018, *Preprocessing for deep learning: from covariance matrix to image whitening*, <https://www.freecodecamp.org/news/https-medium-com-hadrienj-preprocessing-for-deep-learning-9e2b9c75165c/>, diakses tanggal 28 Desember 2021.
- [6] Aindhae, 2019, Cara Menghitung *Mean Absolute Percentage Error (MAPE)* dengan Excel, <https://www.aindhae.com/2019/12/cara-menghitung-mean-absolute.html>, diakses tanggal 28 Desember 2021.