



Prediksi Perubahan Iklim di Indonesia pada Tahun 2013-2014 Menggunakan LSTM

Dhifan Diandra H^{1*}, Fatbina Atsila F¹, Raihan Hanif F¹, Sulthon Akhdan G¹, Novanto Yudistira¹

¹Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya.

* Correspondence: E-mail: dhifand@student.ub.ac.id

ABSTRACTS

The existence of deep learning today is very beneficial and helps humans in everyday life. One of them is prediction, any prediction can be done by deep learning as long as the data obtained allows it to be processed. In this study, climate change predictions were carried out. Because over time and global warming, weather changes are increasingly erratic. Therefore, mistakes often occur in determining the time of the rainy season and summer. Deep learning is an artificial neural network consisting of many layers, where the layer is composed of many nodes. The deep learning architecture used for prediction in this study is Long Short Term Memory (LSTM). In this study, the use of LSTM is because based on previous research, predictions using this architecture can produce quite good output. Based on the test results from 12 data with 1,859 training data, it has been found that the r2 score as a similarity value is -3.7490698103542037, with a mean squared error value of 0.001512831193394959. This figure is a fairly good result with the level of similarity between the actual results and the predicted results which are not much different.

ARTICLE INFO

Article History:

Received 17 Aug 2021

Revised 3 March 2022

Accepted 18 Okt 2022

Available online 20 Nov 2022

Keyword:

deep learning, LSTM, climate, prediction

ABSTRAK

Keberadaan *deep learning* saat ini sangat menguntungkan dan membantu manusia di kehidupan sehari-hari. Salah satunya adalah prediksi, prediksi apapun dapat dilakukan dengan *deep learning* selama data yang didapatkan memungkinkan untuk diolah. Dalam penelitian ini, dilakukan prediksi perubahan iklim. Karena seiring perkembangannya zaman dan terjadinya *global warming*, perubahan cuaca semakin tidak menentu. Maka dari itu sering terjadi kekeliruan dalam menentukan kapan waktu musim hujan dan musim panas. *Deep learning* merupakan sebuah jaringan syaraf tiruan yang terdiri dari banyak *layer*, yang dimana *layer* tersebut tersusun dari banyak *node*. Arsitektur *deep learning* yang digunakan untuk prediksi di penelitian ini adalah *Long Short Term Memory (LSTM)*. Pada penelitian ini penggunaan LSTM dikarenakan berdasarkan penelitian sebelumnya, prediksi dengan yang menggunakan arsitektur ini dapat menghasilkan output yang cukup baik. Berdasarkan hasil pengujiannya dari 12 data dengan 1.859 *data training*, telah didapatkan bahwa nilai *r2 score* sebagai nilai kemiripannya sebesar -3.7490698103542037, dengan nilai *mean squared error* sebesar 0.001512831193394959. Angka ini merupakan hasil yang cukup baik dengan tingkat kemiripan antara hasil sebenarnya dengan hasil prediksi yang terpaut tidak jauh berbeda.

Keyword:

deep learning, LSTM, cuaca, prediksi

1. INTRODUCTION

Seiring berkembangnya zaman, berkembang pula data yang berjalan secara eksponensial. Informasi pun datang dari berbagai sumber, dengan kecepatan yang tinggi. hal tersebut tentunya menjadi suatu tantangan tersendiri bagi pemangku kepentingan yang berurusan dengan data. Data akan sangat bermanfaat jika diolah dengan baik. Terdapat banyak informasi dan manfaat yang didapat dari suatu data, diantaranya dapat bermanfaat bagi bidang kesehatan, prediksi, klasifikasi, pengelompokan, dan bidang lainnya.

Ilmu pengetahuan telah berkembang, hingga ditemukan suatu pendekatan yang dapat melakukan suatu komputasi dengan otomatis. *Artificial Intelligence* (AI) merupakan simulasi dari kecerdasan yang dimiliki oleh manusia yang dimodelkan didalam mesin dan diprogram agar bisa berpikir seperti halnya manusia. Dengan kata lain, AI merupakan sistem komputer yang bisa melakukan pekerjaan-pekerjaan yang umumnya memerlukan tenaga manusia atau kecerdasan manusia untuk menyelesaikan pekerjaan tersebut.

Dalam perkembangan AI, terdapat beberapa inovasi yang terjadi. Diantaranya penerapan *machine learning* dan *deep learning* yang marak dilakukan saat ini. *Machine learning* merupakan suatu mesin yang dikembangkan untuk bisa belajar dengan sendirinya tanpa arahan dari penggunanya. Dalam penerapan *machine learning* dibagi menjadi beberapa bagian, yaitu *supervised*, *unsupervised*, dan *reinforcement*. Lalu terdapat *deep learning* yang merupakan perkembangan dari AI dan merupakan bagian dari *machine learning*, namun memiliki struktur yang lebih kompleks. *Deep learning* merupakan suatu penerapan jaringan syaraf tiruan yang menerapkan *deep neural network* sebagai penyelesaian permasalahan.

Ada banyak permasalahan yang dapat diselesaikan dengan penerapan *deep learning*, dan ada banyak pula cara atau metode yang dapat digunakan pada *deep learning* itu sendiri. *Deep*

learning biasa diterapkan pada dataset yang berjumlah besar dan terdapat proses komputasi yang kompleks.

Salah satu contoh penerapan dari *deep learning* yaitu prediksi cuaca. Dikarenakan berkembangnya zaman, maka cuaca semakin tak menentu yang disebabkan oleh *global warming*. Jika biasanya kita dapat dengan mudah menentukan waktu musim hujan dan musim panas, kini hal tersebut sudah sulit untuk dilakukan. Oleh karena itu, diperlukan adanya suatu metode penyelesaian yang dapat memberikan hasil prediksi yang akurat. *Deep learning* cocok untuk dijadikan sebagai penyelesaian permasalahan dari hal tersebut. Dikarenakan adanya banyak data dengan jumlah yang cukup besar dan memerlukan komputasi yang cukup kompleks untuk penyelesaiannya, tentu saja dengan hasil akurasi yang baik. Untuk mengetahui prediksi cuaca, kita dapat menggunakan beberapa metode yang ada pada *deep learning*, diantaranya *backpropagation*, *Extreme Machine Learning* (EML), *Long Short Term Memory* (LSTM), dsb.

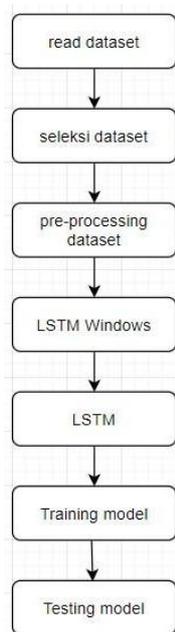
Long Short Term Memory (LSTM) merupakan salah satu jenis dari *Recurrent Neural Network* (RNN) dimana dilakukan modifikasi pada RNN dengan menambahkan *memory cell* yang dapat menyimpan informasi untuk jangka waktu yang lama. Menurut Hochreiter (1997), LSTM merupakan suatu metode yang memanfaatkan hubungan berbasis gradien. Penerapannya pun menggunakan konstanta error melalui *internal states* dari *special units*. Dalam penelitian sebelumnya yang berjudul “*Long Short Term Memory for Monthly Rainfall Prediction in Camau, Vietnam*” didapatkan bahwa metode LSTM merupakan suatu solusi terbaik diantara metode *Artificial Neural Network* (ANN) dan *Seasonal Artificial Neural Network* (SANN). LSTM diusulkan sebagai solusi untuk mengatasi terjadinya *vanishing gradient* pada RNN saat memproses data *sequential* yang panjang. Dan pada penelitian kali ini akan dilakukan suatu implementasi prediksi cuaca dengan menggunakan metode LSTM.

2. METHODS

Pada penelitian ini, dataset *time series* perubahan iklim di Indonesia diperoleh dari website Kaggle.com yang dibuat oleh Berkeley Earth yang berafiliasi dengan Lawrence Berkeley National Laboratory. Data yang diteliti berasal dari studi yang menggabungkan 1,6 juta laporan perubahan temperatur dari tiap negara yang sudah ada dari 16 arsip yang berbeda, data tersebut dapat diunduh di alamat website Kaggle

<https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalTemperatures.csv>

Dalam penerapan penelitian, diterapkan beberapa tahapan sebagai implementasinya. Dimulai dari membaca dataset yang digunakan, seleksi dataset mana saja yang akan digunakan, *preprocessing* dataset, menentukan jumlah data *training* dan data *testing*, penerapan metode LSTM, melakukan *training* model terhadap data yang telah disiapkan, dan melakukan *testing* terhadap data yang telah ditentukan. Untuk lebih lanjut, berikut gambar dari tahapan alur penelitian yang digunakan.



Gambar 3.1. Diagram Alur Penelitian

2.1 DATASET

Dataset perubahan temperatur yang digunakan penelitian ini didapatkan dari *website*

Kaggle. Dimana data tersebut diteliti dan dikumpulkan dengan menggabungkan 1,6 juta laporan perubahan temperatur dari tiap negara yang sudah ada dari 16 arsip yang berbeda. Repositori data *time series* yang digunakan terdiri dari 577462 data *time series* dan memiliki beberapa fitur diantaranya: tanggal, dataset *time series* dimulai dari tahun 1750 untuk fitur *average temperature* dan tahun 1850 untuk max dan min *land temperatures* dan *global ocean* dan *land temperatures*. Yang kedua adalah *Land Average Temperatures*, yang menjelaskan temperatur secara global dalam celcius. Yang ketiga adalah *Land Average Temperature Uncertainty* yang menjelaskan 95% nilai *confidence* interval sekitar rata rata temperatur.

2.2 SELEKSI DATASET

Dataset yang digunakan dalam implementasi mengalami seleksi. Karena dataset terdiri dari berbagai macam negara, maka hanya diambil data yang berasal dari negara Indonesia saja. Setelah dilakukan proses seleksi dataset, maka didapatkan data dengan jumlah 2.265 data.

2.3 PREPROCESSING

Preprocessing data adalah sebuah proses persiapan data yang kemudian akan digunakan. Pada penelitian ini *preprocessing* data dilakukan dengan menghilangkan data yg kosong sehingga terjadi pengurangan data hingga tersisa 1.871 data, dengan pembagian 1.859 data sebagai *data training* dan 12 data sebagai *data testing*.

2.4 LSTM WINDOWS

Pada *LSTM Windows*, dilakukan pembagian data untuk *training* dan *testing*. Pada penelitian ini ditentukan bahwa *data testing* berjumlah 12 data. Hal ini dikarenakan data prediksi yang akan dicari merupakan data tahunan, dimana pada satu tahun terdiri dari 12 bulan.

2.5 LSTM

Pada model LSTM, digunakan model *one-to-one* karena model menerima input berupa data *time series* berjumlah 1 fitur dan menghasilkan prediksi *time series* berjumlah 1 fitur. model juga menggunakan 100 *hidden size* atau banyaknya sel memori.

2.6 TRAINING MODEL

Training model yang digunakan pada penelitian ini menggunakan model LSTM yang sebelumnya telah dibuat. *Training*-nya pun dilakukan penggunaan *optimizer zero gradien*. *Training model* yang dilakukan pada penelitian ini dilakukan sebanyak 100 *epoch*.

2.7 TESTING MODEL

Testing model yang dilakukan pada penelitian ini menggunakan hasil yang sebelumnya telah di *training*. Pada *testing model* diterapkan untuk memprediksi iklim 1 tahun kedepan. Input yang digunakan sebagai *data testing* merupakan hasil pembagian data pada lstm *windows* dengan penyesuaian *preprocessing* terlebih dahulu seperti *data training*.

3. RESULTS AND DISCUSSION

3.1. Results

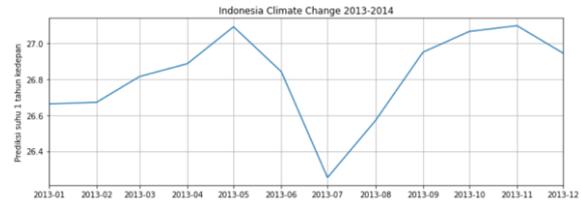
Pada penelitian ini telah didapatkan hasil prediksi perubahan iklim di Indonesia dengan menggunakan 577.462 data *time series*. Namun setelah dilakukan proses seleksi hanya negara Indonesia saja didapatkan data berjumlah 2.265 data. Setelah dilakukan *preprocessing* pun terjadi pengurangan data hingga menjadi 1.871 data, dengan pembagian 1.859 data sebagai *data training* dan 12 data sebagai *data testing*.



Gambar 4.1. Grafik prediksi perubahan iklim di Indonesia tahun 2012-2013



Gambar 4.2. Grafik nilai perubahan iklim sebenarnya di Indonesia tahun 2012-2013



Gambar 4.3. Grafik prediksi perubahan iklim di Indonesia tahun 2013-2014

Untuk mendapatkan nilai ke-validan-nya digunakan perhitungan *Mean Squared Error*, didapati hasil dari nilai prediksi dibandingkan dengan nilai sebenarnya sebesar 0.001512831193394959. Pada penelitian ini pun dilakukan perhitungan *r2 score* untuk melihat nilai kemiripan pada model linear regresi. Nilai yang didapatkan pada perhitungan *r2 score* sebesar -3.7490698103542037.

Tanggal	Suhu Sebenarnya	Suhu Hasil Prediksi
2012-01-01	26.145	26.36730834
2012-02-01	26.152	26.72661442
2012-03-01	26.285	26.72798698
2012-04-01	26.359	26.84287666
2012-05-01	26.640	27.18908446
2012-06-01	26.312	26.81916249
2012-07-01	25.802	26.6221208

2012-08-01	26.065	26.78174865
2012-09-01	26.433	26.8671109
2012-10-01	26.595	26.98781161
2012-11-01	26.651	27.14711866
2012-12-01	26.426	26.88593756

Tabel 4.1. Tabel perbandingan prediksi dengan nilai perubahan iklim sebenarnya di Indonesia tahun 2012-2013

Nilai MSE	0.001512831193394959
-----------	----------------------

Tabel 4.2. Tabel nilai Mean Squared Error

4. CONCLUSION

Berdasarkan hasil implementasi yang dilakukan dapat diambil kesimpulan bahwa prediksi cuaca di Indonesia dapat menggunakan metode LSTM. Sistem dapat melakukan prediksi cuaca dengan dilakukan *preprocessing*, pembagian dataset, inialisasi model LSTM, hingga *training model* terlebih dahulu.

6. REFERENCES

- Anh, D. T., Rutschmann, P., & Bui, M. D. (2018). Long Short Term Memory for Monthly Rainfall Prediction in Camau, Vietnam.
- Dan. (2020, February 18). Predicting weather using LSTM. Predicting weather using LSTM (rs-online.com) (diakses tanggal 29 Desember 2020)
- Elsworth, S., & Guttel, S. (2020, March 12). Time Series Forecasting Using LSTM Networks: A Symbolic Approach.
- Faishol, M. A., Endroyono, & Irfansyah, A. N. (2020, July). Prediksi Polusi Udara Perkotaan di Surabaya Menggunakan Recurrent Neural Network - Long Short Term Memory. *JUTI: Jurnal Ilmiah Teknologi Informasi, Volume 18*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2017). Deep Learning. *The MIT Press*.
- Miftahuddin. (2016). Analisis Unsur-unsur Cuaca dan Iklim Melalui Uji Mann-Kendall Multivariat. *Jurnal Matematika, Statistika, & Komputasi, 13*.
- Qi, M., & Zhang, G. P. (2008). Trend Time-Series Modeling and Forecasting With Neural Networks. *IEEE TRANSACTIONS ON NEURAL NETWORKS, 19*.

Berdasarkan hasil pengujiannya dari 12 data dengan 1.859 *data training*, telah didapatkan bahwa nilai *r2 score* sebagai nilai kemiripannya sebesar -3.7490698103542037, dengan nilai *mean squared error* sebesar 0.001512831193394959. Angka ini merupakan hasil yang cukup baik dengan tingkat kemiripan antara hasil sebenarnya dengan hasil prediksi yang terpaut tidak jauh berbeda.

Dari hasil analisis, dapat disimpulkan bahwa metode LSTM tepat digunakan sebagai prediksi cuaca. Hal ini pun dikarenakan metode LSTM memiliki *memory cell* yang dapat menyimpan informasi untuk jangka waktu yang lama dan dapat mengatasi mengatasi terjadinya *vanishing gradient* pada RNN saat memproses data *sequential* yang panjang. Namun dari hal tersebut pun dapat dilakukan upaya agar hasilnya lebih maksimal dengan penggunaan dataset yang lebih banyak dengan waktu *training* yang lebih lama lagi. Hal ini disebabkan semakin banyak data yang digunakan dan semakin lama data di *training*, maka semakin besar pula kemungkinan agar model lebih memahami dan mempelajari pola-pola yang tersedia.

5. ACKNOWLEDGEMENTS

Saxena, P., Bahad, P., & Kamal, R. (2020). Long Short-Term Memory-RNN based model for Multivariate Car Sales Forecasting. *International Journal of Advanced Science and Technology*, Vol.29.

Varsamopoulos, S., Bertels, K., & Almudever1, C. G. (n.d.). Designing neural network based decoders for surface codes.

Yudistira, N., Widodo, A. W., & Rahayudi, B. (n.d.). Deteksi Covid-19 pada Citra Sinar-X Dada menggunakan Deep Learning yang Efisien. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*.

Zhao, Z., Chen, W., Wu, X., Chen, P., C., Y., & Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11.

Hochreiter., Sepp. & Schmidhuber., Jürgen. (1997). Long Short-Term Memory. *The MIT Press*.

Sarle., W., S. (1994). Neural Networks and Statistical Models. *Proceedings of the Nineteenth Annual SAS Users Group International Conference*.