

Pendekatan Deep Learning Untuk Prediksi Durasi Perjalanan

Nur Ghaniaviyanto Ramadhan^{1*}, Yohani Setiya Rafika Nur², Faisal Dharma Adhinata³

^{1,3}Program Studi Rekayasa Perangkat Lunak, Institut Teknologi Telkom Purwokerto, Jawa Tengah

²Program Studi Informatika, Institut Teknologi Telkom Purwokerto, Jawa Tengah

Email: ^{1*}ghani@ittelkom-pwt.ac.id, ²yohani@ittelkom-pwt.ac.id, ³faisal@ittelkom-pwt.ac.id

(Naskah masuk: 3 Mar 2022, direvisi: 28 Apr 2022, 2 Jun 2022, diterima: 6 Jun 2022)

Abstrak

Setiap orang dalam kehidupan memiliki kecenderungan untuk berpindah dari satu tempat ke tempat lainnya. Perpindahan tersebut dapat dilakukan dengan menggunakan berbagai macam cara seperti menggunakan transportasi pribadi atau umum (bus, taksi, pesawat, dan kereta api). Pada perkembangan teknologi saat ini mode transportasi sudah semakin canggih. Akan tetapi masih ada mode transportasi yang belum modern misalnya seperti taksi, dimana salah satunya tidak dapat memprediksi lama waktu perjalanan. Meskipun sudah ada taksi yang berbasis *online* seperti *Uber*, akan tetapi masih banyak taksi yang belum berbasis *online* sehingga tidak bisa dilakukan estimasi waktu dan jarak. Permasalahan di atas dapat diselesaikan dengan cara melakukan pendekatan berbasis pembelajaran mesin. Salah satu keuntungan yang didapatkan jika kita dapat mengetahui lama waktu estimasi perjalanan yaitu dapat mengatur waktu perjalanan sesuai dengan rutinitas yang sedang dikerjakan ataupun juga dapat menghemat biaya yang dikeluarkan dengan mengetahui jarak yang akan dijalankan. Pada penelitian ini bertujuan untuk memprediksi durasi perjalanan pada *dataset New York taxi trip duration* menggunakan pendekatan *deep learning* yaitu *Long Short Term Memory Recurrent Neural Network (LSTM-RNN)*. Eksperimen dilakukan dengan melakukan *tuning* parameter terkait seperti *epoch*, nilai *dropout*, dan *neurons*. Pengukuran hasil menggunakan nilai *Root Mean Square Error (RMSE)* dan nilai *loss*. Hasil yang didapatkan menggunakan model LSTM-RNN sebesar 0,0012 untuk nilai *loss* dan RMSE 0,4.

Kata Kunci: *Deep Learning*, Durasi Perjalanan, RMSE, Fungsi *Loss*

Deep Learning Approach for Trip Duration Prediction

Abstract

Everyone in life tends to move from one place to another. The transfer can be done using various ways, such as using private or public transportation (bus, taxi, plane, and train). However, there are modes of transportation that are not yet modern, such as taxi, which cannot predict the length of the trip. Although there are already online-based taxi such as Uber, there are still many taxi that are not online-based so it is impossible to estimate time and distance. The above problems can be solved by using a machine learning-based approach. One of the advantages if we know the estimated time of travel is that we can adjust the travel time according to the routine that is being worked on or can also save costs incurred by knowing the distance to be run. This study aims to predict the duration of the trip in New York taxi trip duration dataset using a deep learning approach, namely Long Short Term Memory Recurrent Neural Network (LSTM-RNN). Experiments are carried out by tuning related parameters such as epochs, dropout values, and neurons. Measurement of results using the value of Root Mean Square Error (RMSE) and loss value. The results obtained using the LSTM-RNN model are 0.0012 for the loss value and RMSE 0.4.

Keywords: *Deep Learning*, Trip Duration, RMSE, Loss Function.

I. PENDAHULUAN

Setiap orang dalam kehidupan memiliki kecenderungan untuk berpindah dari satu tempat ke tempat lainnya.

Perpindahan tersebut dapat dilakukan dengan menggunakan berbagai macam cara seperti menggunakan transportasi pribadi atau umum (bus, taksi, pesawat, dan kereta api). Pada perkembangan teknologi saat ini mode transportasi sudah

semakin canggih. Akan tetapi masih ada mode transportasi yang belum modern misalnya seperti taksi, yang dimana salah satunya tidak dapat memprediksi lama waktu perjalanan. Meskipun sudah ada taksi yang berbasis *online* seperti *Uber*, akan tetapi masih banyak taksi yang belum berbasis *online* sehingga tidak bisa dilakukan estimasi waktu dan jarak.

Permasalahan di atas dapat diselesaikan dengan cara melakukan pendekatan berbasis pembelajaran mesin. Salah satu keuntungan yang didapatkan jika kita dapat mengetahui lama waktu estimasi perjalanan yaitu dapat mengatur waktu perjalanan sesuai dengan rutinitas yang sedang dikerjakan ataupun juga dapat menghemat biaya yang dikeluarkan dengan mengetahui jarak yang akan dijalanakan.

Beberapa penelitian sebelumnya yang sudah dilakukan untuk memprediksi waktu tempuh perjalanan, seperti pada *paper* berikut. Penulis melakukan prediksi waktu perjalanan menggunakan model *deep learning* yaitu *XGBoost* dan *MLP* [1]. Pada penulis lain melakukan studi tentang memperkirakan tujuan perjalanan berdasarkan lintasan awal dengan menggunakan model klasifikasi *long short-term memory* dan *embedding technology* [2]. Penulis lain melakukan prediksi waktu perjalanan secara statis dari tempat satu ke tempat lainnya dengan menerapkan model regresi *XGBoost* [3].

Penelitian lainnya membahas tentang prediksi waktu perjalanan taksi *Brunswick* berdasarkan lintasan data mobil dengan menggunakan model *XGBoost* dan *random forest* [4]. Penelitian ini menyajikan sistem model untuk prediksi tujuan perjalanan dengan data *GPS* multi-hari dengan membangun model rantai *markov* dan *multinomial logit* [5]. Pada penelitian ini melakukan prediksi waktu pengiriman untuk layanan pos yang mana dilakukan percobaan beberapa metode seperti regresi linier dan *tree ensembles*, *bagging*, dan *boosting* [6].

Penelitian lainnya melakukan prediksi jangka panjang dengan pandangan beberapa hari dan prediksi jangka pendek dengan pandangan satu jam. Kedua perspektif tersebut relevan untuk tugas-tugas perencanaan dalam konteks mobilitas perkotaan dan layanan transportasi [7]. Pada penelitian ini akan dilakukan pengembangan model prediksi tujuan perjalanan berdasarkan layanan pencarian dan penemuan berbasis lokasi *online* (khususnya, *Google Places API*) dan kumpulan data perjalanan terbatas yang biasanya tersedia setelah perjalanan selesai [8].

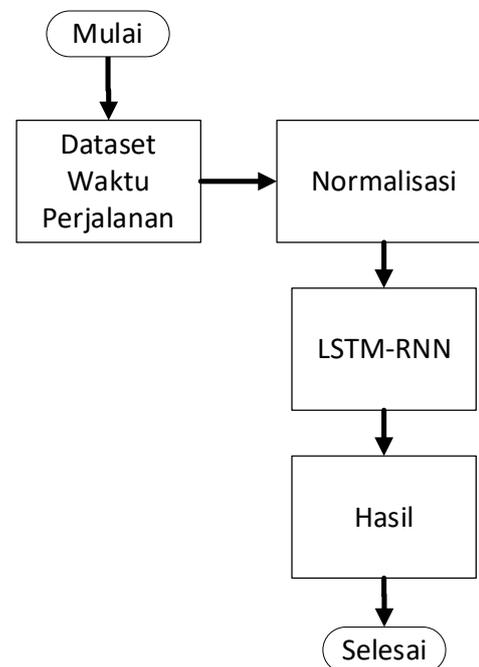
Pada makalah ini dilakukan prediksi *round trip time* berbasis pembelajaran mendalam menggunakan *Recurrent Neural Networks (RNN)* [9]. Dalam *paper* ini, penulis melakukan eksplorasi model pembelajaran mesin yang mendalam yaitu model jaringan saraf *LSTM* yang dilakukan untuk prediksi waktu perjalanan [10]. Studi ini mengevaluasi metode penyediaan informasi pra-perjalanan yang memprediksi waktu perjalanan dengan mencocokkan akumulasi masa lalu dan data lalu lintas pada hari perjalanan dengan tetap memperhitungkan waktu yang dibutuhkan untuk mencapai tujuan [11].

Paper berikut ini melakukan penelitian tentang prediksi waktu kedatangan bus dan desain ulang jadwal dengan menggunakan penggabungan model *Support Vector Regression (SVR)* dan *Kalman Filter* [12]. Penelitian ini memiliki tujuan untuk menganalisis variasi waktu perjalanan

bus dan memprediksi waktu perjalanan secara akurat dengan penerapan model *SVR* [13]. Penelitian ini mengusulkan suatu metode berdasarkan setiap zona bangunan, situasi penggunaan lahan, dan aksesibilitas bus untuk meramalkan arus perjalanan penumpang bus pada periode mendatang [14]. Metode yang digunakan yaitu *Artificial Neural Network (ANN)* [14]. Penelitian ini untuk memprediksi durasi perjalanan sewa sepeda di *Seoul Bike sharing system* [15]. Prediksi dilakukan dengan kombinasi data *Seoul Bike* dan data cuaca dengan beberapa teknik pada data mining yaitu *regression*, *gradient*, *KNN*, dan *Random Forest* [15].

Berdasarkan pemaparan permasalahan yang ada pada penelitian sebelumnya beserta metode-metode yang pernah digunakan, sehingga pada penelitian ini akan melakukan prediksi waktu perjalanan dari satu titik ke titik lainnya pada moda transportasi dengan menggunakan model *deep learning Long-Short Term Memory Recurrent Neural Network (LSTM-RNN)*.

II. METODE PENELITIAN



Gambar 1. Desain Proses

A. Dataset

Penelitian ini menggunakan *dataset New York taxi trip duration* yang berasal dari rekam waktu perjalanan menggunakan transportasi taksi [16]. *Dataset* ini terdiri dari 11 kolom dan 10.485.576 baris data.

Tabel 1. Karakteristik Dataset

No	Nama Kolom	Type Data
1	ID	Integer
2	Vendor ID	Integer
3	Pickup Datetime	Date
4	Dropoff Datetime	Date
5	Passenger Count	Integer
6	Pickup Longitude	Integer
7	Pickup Latitude	Integer
8	Dropoff Longitude	Integer
9	Dropoff Latitude	Integer
10	Store	String
11	Trip Duration	Integer

Pada dataset Tabel 1 memiliki fitur-fitur yang terkait dengan durasi perjalanan seperti *pickup datetime* dan *dropoff datetime*. Dataset tersebut berisikan informasi-informasi terkait perjalanan, jumlah penumpang, ketinggian lokasi penjemputan, dan ketinggian lokasi *dropoff* penumpang. Hal tersebut berguna untuk menentukan proses prediksi waktu yang ditempuh.

B. Normalisasi

Proses ini dilakukan normalisasi menggunakan formula *min-max* (1) [17]. Normalisasi bertujuan untuk menyamakan nilai pada setiap fitur antara nilai 0-1.

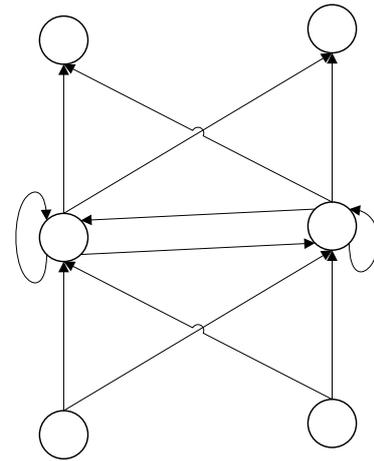
$$N^* = \frac{N - \min(n)}{\max(n) - \min(n)} \tag{1}$$

Dimana *N* yaitu data yang belum dinormalisasi. *min(n)* merupakan nilai minimum dari semua data dan *max(n)* adalah nilai maksimum dari semua data.

C. LSTM-RNN

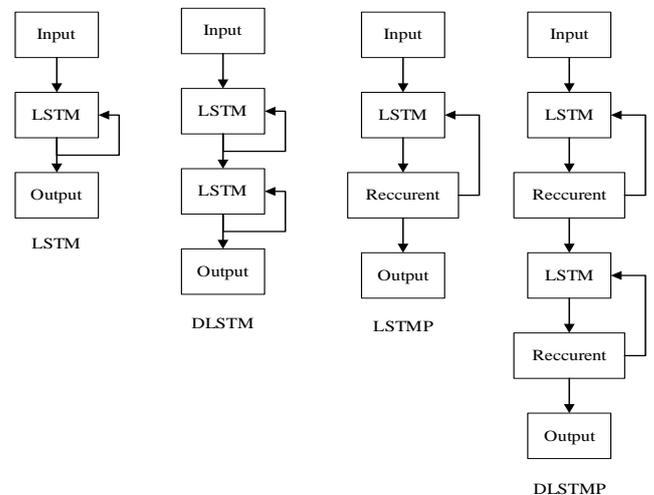
Model *LSTM-RNN* merupakan model yang termasuk ke dalam pendekatan *deep learning*. *LSTM-RNN* dapat diaplikasikan ke berbagai permasalahan seperti *early learning tasks* dan *cognitive learning tasks* (*speech recognition*, *machine translation*, dan *image processing*). *LSTM* dapat mempelajari cara menjembatani jeda waktu minimal lebih dari 1.000 langkah waktu diskrit [17]. Solusinya menggunakan korsel kesalahan konstan (CEC) yang menegakkan aliran kesalahan konstan dalam sel khusus. Akses ke sel ditangani oleh unit gerbang multiplikasi yang mempelajari kapan harus memberikan akses.

Reccurent Neural Networks (RNN) adalah sistem dinamis yang memiliki keadaan internal pada setiap langkah waktu klasifikasi [17]. Hal ini disebabkan lingkaran koneksi antara neuron lapisan yang lebih tinggi dan lebih rendah dan umpan balik mandiri opsional koneksi. Koneksi umpan balik ini memungkinkan *RNN* untuk menyebarkan data dari peristiwa sebelumnya ke langkah pemrosesan saat ini. Jadi, *RNN* membangun memori waktu acara seri. Untuk arsitektur *basic* dari *RNN* dapat dilihat pada Gambar 2.



Gambar 2. RNN Arsitektur

Sedangkan untuk arsitektur model *LSTM-RNN* penelitian ini dapat dilihat pada Gambar 3 yang mana menggabungkan kedua arsitektur dari *LSTM* dan *RNN* [18].



Gambar 3. Arsitektur LSTM-RNN

Untuk formula dari *LSTM* dapat dilihat pada formula (2) (3) (4) dan (5).

$$cf_t = \sigma_1 (W_{cf} \cdot [O_{t-1}, x_t] + b_{cf}) \tag{2}$$

$$I_t = \sigma_2 (W_t \cdot [O_{t-1}, x_t] + b_t) \tag{3}$$

$$S_t = \tanh (W_s \cdot [O_{t-1}, x_t] + b_s) \tag{4}$$

$$S_t = cf_t \times S_{t-1} + I_t \times S_{t-1} \tag{5}$$

Dimana informasi baru yang akan disimpan dalam keadaan sel dihitung menggunakan dua lapisan jaringan. Lapisan *sigmoid* (σ_2) memutuskan nilai yang akan diperbarui (I_t) (3) dan lapisan *tanh* 1 yang mengembangkan vektor dari nilai kandidat baru (S_t) seperti yang ditunjukkan pada (4). Kombinasi untuk menambahkan status. Akhirnya, status sel diperbarui menggunakan (5).

Untuk menghitung hasil dari model *LSTM-RNN* menggunakan nilai *Root Mean Square Error (RMSE)* dan fungsi *loss*. Formula *RMSE* dapat dilihat pada (6) [17] dan fungsi *loss* (7) [17].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{ii} - y_i)^2}{n}} \quad (6)$$

Dimana y_{ii} dan y_i masing-masing adalah nilai prediksi dan aktual dari beban kerja pada saat sampel i dan N adalah jumlah data sampel.

$$Loss(p, q) = \frac{1}{|M|} \sum_{i=1}^{i=M} -qi \log pi \quad (7)$$

Di mana p dan q mewakili label yang diprediksi dan label yang sebenarnya, dan M menunjukkan ukuran batch.

III. HASIL DAN ANALISIS

Eksperimen yang dilakukan pada penelitian ini yaitu dengan melakukan *tuning* beberapa parameter seperti *epoch* dan nilai *dropout*. Hasil eksperimen dapat dilihat pada Tabel 2.

Tabel 2. Hasil Eksperimen

Epoch	Neurons	Dropout	RMSE	Loss
100	25	0,1	0,4	0,0013
200	15	0,2	0,4	0,0013
500	35	0,3	0,4	0,0012
1000	55	0,7	0,4	0,0012

Berdasarkan hasil eksperimen Tabel 2 memperlihatkan bahwa nilai *RMSE* dan *loss* pada penelitian ini sangat stabil di angka 0,4 untuk nilai *RMSE* dan 0,12-0,13 untuk *loss function*. Proses *tuning* parameter dilakukan seperti pada nilai *epoch* awalnya bernilai 100 sampai dengan *epoch* 1000. Untuk nilai *neurons* dan nilai *dropout* juga dilakukan perubahan nilai semakin naik, guna melihat hasil *RMSE* dan *loss* apakah ada perubahan. Namun, berdasarkan hasil *tuning* parameter untuk model *LSTM-RNN* pada permasalahan prediksi waktu perjalanan sangat tepat digunakan.

Pada Tabel 3 merupakan contoh hasil dari prediksi durasi perjalanan yang dihasilkan.

Tabel 3. Hasil Prediksi

Trip Duration Test (Minutes)	Trip Duration Predict (Minutes)
455	436
663	647
2124	2009
429	399
435	404

Trip Duration Test (Minutes)	Trip Duration Predict (Minutes)
443	431
341	337
1551	1479

Berdasarkan hasil prediksi Tabel 3 terlihat bahwa nilai prediksi durasi yang dihasilkan mendekati waktu tempuh aslinya. Misalnya contoh pada *trip duration test* dengan waktu 435, hasil prediksinya yaitu yaitu 404. Maka hal tersebut juga berdampak pada nilai *RMSE* dan *loss* yang dihasilkan. Jika hasil prediksi sangat jauh dari nilai *testing* maka untuk nilai *RMSE* dan *loss* akan bernilai cukup besar, Nilai *RMSE* dan *loss* dikatakan baik jika yang dihasilkan nilainya semakin kecil.

IV. KESIMPULAN

Berdasarkan hasil eksperimen yang telah dilakukan dan nilai yang dihasilkan. Penelitian ini mencoba melakukan pendekatan metode *deep learning* untuk mendapatkan pencarian rute terbaik pada moda transportasi umum berupa taksi. Metode *LSTM-RNN* ternyata memberikan hasil yang cukup memuaskan. Hal tersebut dapat dilihat dari nilai *RMSE* yang dihasilkan sebesar 0,4 dan nilai *loss* sebesar 0,12. Metode *LSTM-RNN* pada penelitian ini menggunakan *dataset New York taxi trip duration* dengan jumlah rekord sebanyak 10.485.576, akan tetapi dari proses eksperimen untuk waktunya jika nilai *epoch* semakin besar maka waktu yang dibutuhkan semakin lama. Perbedaan waktunya dapat mencapai 30 detik untuk kelipatan nilai *epoch* 200. Untuk parameter yang tidak berpengaruh signifikan yaitu *neurons*.

REFERENSI

- [1] M. Poongodi, M. Malviya, C. Kumar, M. Hamdi, V. Vijayakumar, J. Nebhen, and H. Alyamani, "New York City taxi trip duration prediction using MLP and XGBoost," *International Journal of System Assurance Engineering and Management*, vol. 13, no. S1, pp. 16–27, 2021.
- [2] J. Tang, J. Liang, T. Yu, Y. Xiong, and G. Zeng, "Trip destination prediction based on a deep integration network by fusing multiple features from taxi trajectories," *IET Intelligent Transport Systems*, 2021.
- [3] K. D. Kankanamge, Y. R. Witharanage, C. S. Withanage, M. Hansini, D. Lakmal, and U. Thayasivam, "Taxi trip travel time prediction with isolated XGBoost regression," *2019 Moratuwa Engineering Research Conference (MERCOn)*, 2019.
- [4] J. Fiosina, "Explainable federated learning for taxi travel time prediction," *Proceedings of the 7th International Conference on Vehicle Technology and Intelligent Transport Systems*, 2021.
- [5] F. Zong, Y. Tian, Y. He, J. Tang, and J. Lv, "Trip destination prediction based on multi-day GPS data,"

- Physica A: Statistical Mechanics and its Applications*, vol. 515, pp. 258–269, 2019.
- [6] J. Khiari and C. Olaverri-Monreal, “Boosting algorithms for delivery time prediction in Transportation Logistics,” *2020 International Conference on Data Mining Workshops (ICDMW)*, 2020.
- [7] H. Huang, M. Pouls, A. Meyer, and M. Pauly, “Travel time prediction using tree-based ensembles,” *Lecture Notes in Computer Science*, pp. 412–427, 2020.
- [8] A. Ermagun, Y. Fan, J. Wolfson, G. Adomavicius, and K. Das, “Real-time trip purpose prediction using online location-based search and Discovery Services,” *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 96–112, 2017.
- [9] A. Dong, Z. Du, and Z. Yan, “Round trip time prediction using recurrent neural networks with minimal gated unit,” *IEEE Communications Letters*, vol. 23, no. 4, pp. 584–587, 2019.
- [10] Y. Duan, Y. L.V., and F.-Y. Wang, “Travel time prediction with LSTM neural network,” *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016.
- [11] H. Warita, H. Morita, M. Kuwahara, E. Chung, and A. Tanaka, “Travel time prediction for pre-trip information using latest traffic conditions on expressway,” *International Journal of Intelligent Transportation Systems Research*, vol. 2, no. 1, pp. 11–19, 2004.
- [12] X. Zhang, M. Yan, B. Xie, H. Yang, and H. Ma, “An automatic real-time bus schedule redesign method based on bus arrival time prediction,” *Advanced Engineering Informatics*, vol. 48, p. 101295, 2021.
- [13] A. K. Bachu, K. K. Reddy, and L. Vanajakshi, “Bus travel time prediction using support vector machines for high variance conditions,” *Transport*, vol. 36, no. 3, pp. 221–234, 2021.
- [14] S. Yu, C. Shang, Y. Yu, S. Zhang, and W. Yu, “Prediction of bus passenger trip flow based on Artificial Neural Network,” *Advances in Mechanical Engineering*, vol. 8, no. 10, p. 168781401667599, 2016.
- [15] V. E. Sathishkumar, J. Park, and Y. Cho, “Seoul bike trip duration prediction using data mining techniques,” *IET Intelligent Transport Systems*, vol. 14, no. 11, pp. 1465–1474, 2020.
- [16] “New York City taxi trip duration,” *Kaggle*. [Online]. Available: <https://www.kaggle.com/c/nyc-taxi-trip-duration>. [Accessed: 28-Dec-2021].
- [17] R. C. Staudemeyer and E. R. Morris, “Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks,” *ArXiv*.
- [18] N. G. Ramadhan , N. A. F. Tanjung , and F. D. Adhinata , “Implementation of LSTM-RNN for Bitcoin Prediction ,” *Indonesia Journal on Computing (Indo-JC)*, vol. 6, no. 3, pp. 17–24, 2021.